PROCESO GESTIÓN DEL TALENTO HUMANO

FORMATO INFORME MENSUAL EJECUCIÓN CONTRACTUAL

Medellín, 31 de julio del 2022

Señor (a)
John Jairo Castro Maldonado
SUPERVISOR(A) CONTRATO No. CO1.PCCNTR.3367636
Instructor Grado 16
Centro de Servicios y Gestión Empresarial
Medellín

Asunto: Informe mensual de ejecución contractual de julio del año 2022.

Referencia: No 3367636 del año 2022

Jesus Arbey Benavides Guevara, identificado con la cédula de ciudadanía No. 1.085.260.713 de Pasto, en mi calidad de Contratista del SENA, en Centro de Servicios y Gestión Empresarial, en cumplimiento del Contrato de Prestación de Servicios de la referencia, a continuación presento el Informe de actividades realizadas en el mes objeto de cobro.

Valor y forma de Pago: Se fija como valor total para el contrato la suma de TREINTA Y CUATRO MILLONES TRESCIENTOS MIL PESOS MCTE ($ 34.300.000) Incluido IVA. Esta suma será pagada por el SENA al contratista de la siguiente manera: Valor por mes: CUATRO MILLONES NOVECIENTOS MIL PESOS MCTE ($ 4.900.000), y el equivalente por fracción de mes.

Plazo: Será hasta el treinta (30) de agosto (08) de 2022.

OBJETO: Prestar servicios profesionales para el desarrollo de las actividades de levantamiento de indicadores, bases de datos y el modelo matemático con base en analítica de datos y Big Data en el prototipo de software para modelo de Innovación en el desarrollo del proyecto de I+D+i SGPS 9748

Obligaciones Específicas:

No	Obligaciones	Acciones realizadas	Evidencias
1	Apoyar el desarrollo de aplicaciones o algoritmos usados en la actualidad para solucionar de forma eficiente y eficaz problemas o	Se finalizó modelos de Kmeans y PCA para la determinación de conglomerados y en la clasificación de centros de formación en torno a la dimensión su productividad y gestión de	ANEXOS 2. Avances de redacción de artículo científico orientado en la visualización de mapas para actividades de ciencia tecnología e innovación. ANEXO 3. Modelo finalizado de clusterización- K means y clasificación PCA no supervisados para las

	retos identificados en Big Data y analítica de datos con aplicación en gestión de la innovación	procesos con datos aleatorios, para su implementación en la Interfaz de programación de aplicaciones (API). Se realizó mapas geoespaciales de Antioquia, donde se visualizó los centros de formación donde, posterior a la etapa de clusterización se realizó una clasificación de alta, media y baja productividad y gestión de manera respectiva, y que se visualiza como una alerta para cada centro, para su implementación en la Interfaz de programación de aplicaciones (API).	dimensiones de productividad y procesos de centros de formación en la regional Antioquia para aplicación en API ANEXOS 4. Presentación de avances técnicos julio 2022. ANEXOS 5. Actas de reuniones. ANEXOS 2. Avances de redacción de artículo científico orientado en la visualización de mapas para actividades de ciencia tecnología e innovación. ANEXO 3. Modelo finalizado de clusterización- K means y clasificación PCA no supervisados para las dimensiones de productividad y procesos de centros de formación en la regional Antioquia para aplicación en API ANEXOS 4. Presentación de avances técnicos julio 2022. ANEXOS 5. Actas de reuniones.
2	Revisar las herramientas tecnológicas o digitales de Big Data y analítica de datos necesarias para dar cumplimiento a los objetivos propuestos dentro del proyecto.	Se implementó diferentes tipos de herramientas ofrecidas por la librería scikit-learn de python, de software libre para la implementación de modelos supervisados.	ANEXO 3. Modelo finalizado de clusterización- K means y clasificación PCA no supervisados para las dimensiones de productividad y procesos de centros de formación en la regional Antioquia para aplicación en API
3	Revisar y actualizar indicadores para la gestión de la	Se implementó, por medio de modelos de clusterización y	ANEXO 3. Modelo finalizado de clusterización- K means y clasificación PCA no supervisados para las

	innovación según objetivos del proyecto	caracterización KMeans y Métodos de Componentes Principales, dos dimensiones en términos de indicadores para la gestión de la innovación - proyecto SGPS 9748 prototipo innovación a lo largo del tiempo: procesos y productividad, para implementación en API.	dimensiones de productividad y procesos de centros de formación en la regional Antioquia para aplicación en API ANEXOS 4. Presentación de avances técnicos julio 2022. ANEXOS 5. Actas de reuniones.
4	Realizar pruebas relacionadas con la compatibilidad de la tecnología y la aplicación de los sistemas de Big Data y analítica de datos para la evaluación y valoración de la Gestión de CTeI para los Centros SENA Regional Antioquia	Se organizaron modelos de machine learning, desde la perspectiva geoespacial, para posterior visualización de los centros de formación. Se acopló los algoritmos generados desde python hacia el interfaz de programación de aplicaciones.	ANEXO 3. Modelo finalizado de clusterización- K means y clasificación PCA no supervisados para las dimensiones de productividad y procesos de centros de formación en la regional Antioquia para aplicación en API ANEXOS 4. Presentación de avances técnicos julio 2022. ANEXOS 5. Actas de reuniones.
5	Apoyar la elaboración del manual del sistema y usuario y las actividades de transferencia	Gestión por realizar	Gestión por realizar
6	Presentar los informes de seguimiento del proyecto de acuerdo con las solicitudes del líder del área o supervisor	Se elaboró el informe de pago mensual GTH-F-062 V09 Se elaboró diapositivas en presentación de informes técnicos para el supervisor y coordinador del contrato del mes en curso.	Informe en formato pdf. ANEXOS 4. Presentación de avances técnicos julio 2022.
7	Divulgar en eventos académicos/empresariales los	Se obtuvo resultados participación en evento de RedCOLSI Nodo Antioquia.	ANEXO 1. Medios de divulgación: Resultados de participación REDCOLSI, participación en Workshop on Modeling and Simulation for Science and Engineering, y reunión con

	resultados del proyecto.		representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño.
		Participación en en Workshop on Modeling and Simulation for Science and Engineering, en modalidad de Artículo.	ANEXO 1. Medios de divulgación: Resultados de participación REDCOLSI, participación en Workshop on Modeling and Simulation for Science and Engineering, y reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño.
		Reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño, para la divulgación de los proyectos de analítica de datos en el laboratorio de inteligencia artificial el día 06 de Julio del 2022 en la plataforma de google meet.	ANEXO 1. Medios de divulgación: Resultados de participación REDCOLSI, participación en Workshop on Modeling and Simulation for Science and Engineering, y reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño.
8	Redactar el informe de investigación y de resultados del proyecto.	Se elaboró un (1) informe de pago mensual GTH-F-062 V09.	Informe en formato pdf.
		Se elaboró diapositivas en presentación de informes técnicos para el supervisor y coordinador del contrato del mes en curso.	ANEXOS 4. Presentación de avances técnicos julio 2022.
9	Redactar y someter a publicación artículo de	Se sometió a publicación resumen de artículo de investigación en Workshop	ANEXO 1. Medios de divulgación: Resultados de participación REDCOLSI, participación en Workshop

	investigación, libro de investigación o capítulo de libro de investigación relacionado con el proyecto, en medio reconocido por Minciencias.	on Modeling and Simulation for Science and Engineering	on Modeling and Simulation for Science and Engineering, y reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño.
10	Apoyar actividades de Grupo de Investigación, Semilleros de investigación y SENNOVA que sean propias de su área de conocimiento y en relación con el objeto contractual	Se obtuvo resultados de participación en evento de REDCOLSI nodo Antioquia, obteniendo un puntaje de 91.0	ANEXO 1. Medios de divulgación: Resultados de participación REDCOLSI, participación en Workshop on Modeling and Simulation for Science and Engineering, y reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño.
11	Apoyar la evaluación técnica, administrativa y presupuestal de los proyectos de investigación aplicada, desarrollo tecnológico e innovación que se formulen por el SENA para el desarrollo de actividades misionales, acorde con las indicaciones dadas por la Dirección General.	Gestión por realizar	Gestión por realizar
12	Desarrollar sus actuaciones en el marco de la normatividad	Se elaboró un (1) informe de pago mensual GTH-F-062 V09.	Informe en formato pdf.

	vigente y presentar los informes mensuales de acuerdo con el objeto del presente contrato.		

A continuación relaciono los desplazamientos que realicé previo a la presentación de este informe. Una vez finalizado cada desplazamiento presenté al ordenador del gasto el informe en el Formato Informe Legalización Desplazamiento Contratista GTH-F-087, en el que se describieron las actividades desarrolladas y los resultados de cada desplazamiento. Cada informe cuenta con el visto bueno del Supervisor.

Se lista a continuación el soporte de la legalización de los desplazamientos realizados, los cuales forman parte integral del presente informe de ejecución contractual.

ITEM	No DE LA ORDEN DE VIAJE	LUGAR DE DESPLAZAMIENTO	FECHAS DE DESPLAZAMIENTOS	
			INICIO	FINALIZACIÓN
1.		No he realizado ningún desplazamiento		

Nota 1: *Por cada desplazamiento que haya realizado el contratista, adjuntará el respectivo informe que la soporte. En caso de haber realizado el desplazamiento en fecha posterior a la presentación del informe de ejecución contractual, deberá reportarlo en el siguiente informe de ejecución contractual.*

Para el trámite de la cuenta me permito adjuntar: Documentos electrónicos enunciados como evidencias del cumplimiento de las obligaciones contractuales y los desplazamientos realizados y el No.60015723 de la planilla, miplanilla correspondiente al mes de junio del año 2022. (Decreto Ley 2106 de 2019 – " *Decreto Ley Antitrámites"*)

Declaración bajo la gravedad de juramento.

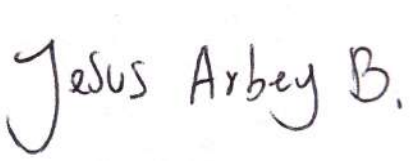
Me permito declarar **bajo la gravedad de juramento** que durante el periodo objeto de cobro SI____NO_X_ he suscrito otro u otros contratos con entidade(s) del Estado el (los) cual(es) se encuentra(n) en ejecución; la anterior declaración de acuerdo con lo establecido en el numeral 5º de la Directiva Presidencial No. 001 de 2022, a fin de identificar potenciales conflictos de interés y hacer la gestión del riesgos para la toma de decisiones dentro del principio de la debida diligencia frente a la ejecución de los contratos de prestación de servicios profesionales o de apoyo a la gestión.

Nota: En el evento de señalar **SI,** el análisis efectuado por el supervisor y ordenador de pago deberá quedar documentado y publicado en el SECOP II en los documentos relacionados en el archivo Gestión Contractual GC_ del periodo objeto de cobro.

Evidencias en 219 folios:

- ANEXOS.
- https://drive.google.com/drive/folders/1j61SX1uD8L8vP3UX7BeYt0NUtED3ZxwB?usp=sharing

Cordialmente,



JESUS ARBEY BENAVIDES GUEVARA
Contratista
C.C. No. 1.085.260.713 DE PASTO (NARIÑO)

Recibí a satisfacción:

JOHN JAIRO CASTRO MALDONADO
SUPERVISOR(A) CONTRATO No. CO1.PCCNTR.3367636 DE 2022
INSTRUCTOR GRADO 16
C.C. No. 88.265.552

ANEXOS

ANEXO 1. Medios de divulgación: Resultados de participación REDCOLSI, participación en *Workshop on Modeling and Simulation for Science and Engineering,* y reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño.

Resultados de participación en evento de REDCOLSI nodo Antioquia, obteniendo un puntaje de 91.0.

Desarrollo de un prototipo de software para la medición de la Gestión de CTeI en los Centros SENA Regional Antioquia aplicando analítica de datos.	91.00

Figura A1. Resultados obtenidos de evento REDCOLSI

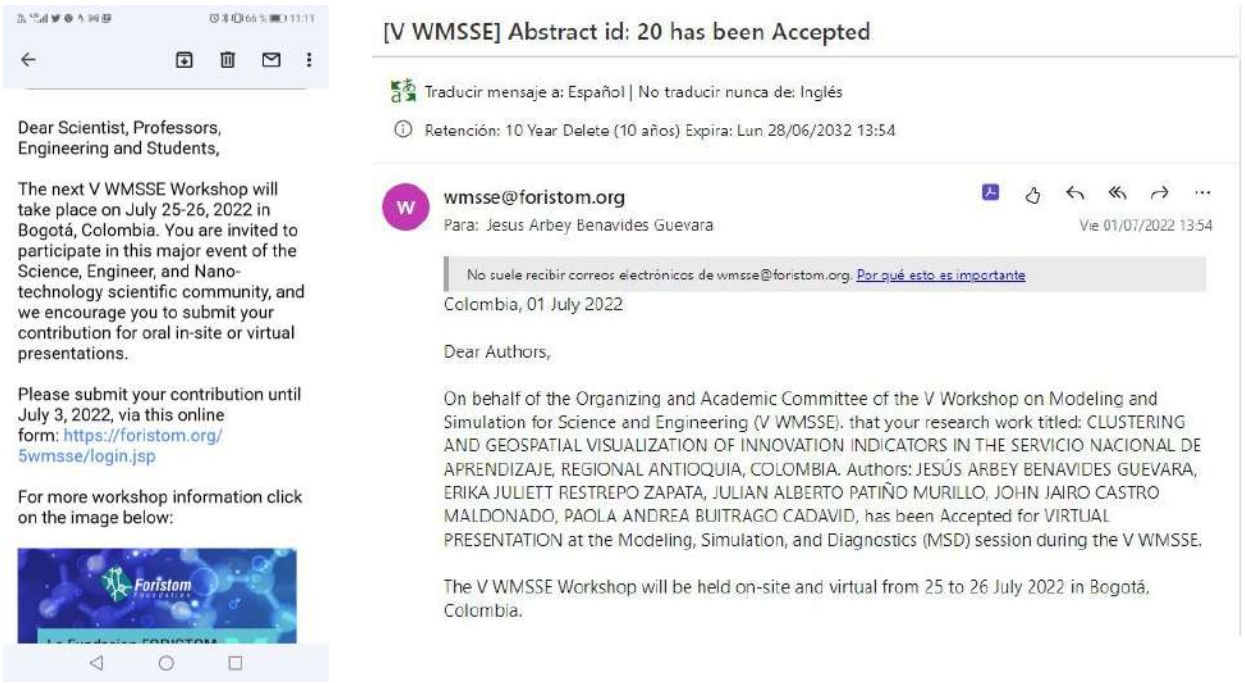


Figura A2. Aplicación y aceptación en el evento *V Workshop on Modeling and Simulation for Science and Engineering*. Proyecto: “Clustering and geospatial visualization of innovation indicators in the Servicio Nacional de Aprendizaje Regional Antioquia, Colombia.”

Reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño, para la divulgación de los proyectos de analítica de datos en el laboratorio de inteligencia artificial el día 06 de Julio del 2022 en la plataforma de google meet.

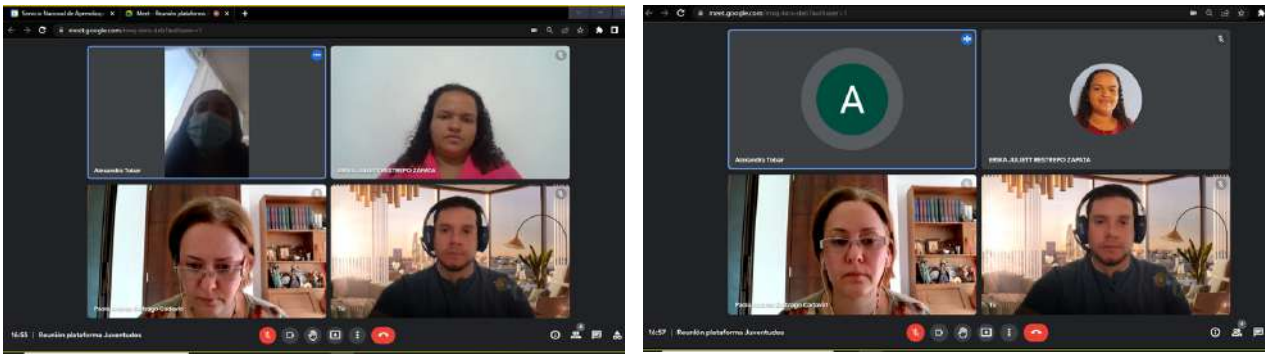


Figura A3. Reunión con representante de Consejeros de juventudes y plataforma juventudes del departamento de Nariño para divulgación del laboratorio de analítica de datos.

XVI SIMPOSIO INTERNACIONAL DE ENERGÍAS

ENERGÍA RENOVABLE

14 y 15 de octubre

INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO®

Apoyan:

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

Memorias

ISSN 2422-3891

2021 EXPO TECNOLÓGICA

XVI SIMPOSIO INTERNACIONAL DE ENERGÍAS

ENERGÍA RENOVABLE

Memorias

Contenido

1. COMPARATIVE ASSESSMENT OF COMPUTATIONAL MODELS FOR THE EFFECTIVE TENSILE STRENGTH OF NANO-REINFORCED COMPOSITES.
2. EFFECT OF ADDITION MWCNT TO A MINERAL LUBRICANT ON DIESEL ENGINE PERFORMANCE
3. XVI SIMPOSIO INTERNACIONAL DE ENERGÍA EXPOTECNOLÓGICA 2021 (XVI SIE – 2021) "HEAT FLOW MODELING IN THE ALKALINE ACTIVATION PROCESS OF FLY ASH".
4. ESTRATEGIAS DE ESTABILIZACIÓN DE LLAMAS DE MATERIAL RUMINAL SECO PARA APROVECHAMIENTO ENERGÉTICO
5. RESPUESTA EN FRECUENCIA DE SISTEMA BESS ACOPLADO A RED DE BAJA INERCIA.
6. RESULTADOS DE LA IMPLEMENTACIÓN DE MÓDULO EMBEBIDO DE CONVERSIÓN DE TENSIONES EN MOTORES TRIFÁSICOS.
7. MODELADO DE LA ADMISIÓN DE LA CARGA EN UNA METODOLOGÍA SECUENCIAL PARA COMBUSTIÓN HCCI
8. SIMULATION OF THE THERMAL PERFORMANCE OF AN IGNITION ENGINE FUELED WITH GASOLINE/ISO-BUTANOL BLENDS
9. ESTUDIO COMPARATIVO DEL RENDIMIENTO DE UNA TURBINA TIPO TORNILLO DE ARQUÍMEDES EN ARREGLO TRADICIONAL Y EN ESPIRAL PARA APLICACIONES HIDROCINÉTICAS
10. NEW APPROACH FOR FILLING SIMULATION OF DUAL-SCALE PREFORMS USED IN THE MANUFACTURING OF COMPOSITES MATERIALS.
11. EVALUACIÓN DE CONFIABILIDAD EN MICROGRIDS: COMPARACIÓN DE MÉTODOS DE MODELAMIENTO.
12. THERMAL FINITE ELEMENT ANALYSIS IMPLEMENTED WITH OPEN SOURCE TOOLS AND HIGH-PERFORMANCE COMPUTING
13. PROTOTIPO DE RED PARA LA IDENTIFICACIÓN Y SEGUIMIENTO DE PERSONAS EN UN ENTORNO EDUCATIVO CON IoT.
14. MODELO DINÁMICO DE UN SISTEMA DE GASIFICACIÓN DE LECHO FIJO ESTRATIFICADO
15. EMOTION DETECTION FROM GALVANIC SKIN RESPONSE USING MACHINE LEARNING IN A HUMAN-ROBOT SCENARIO
16. IDENTIFICACIÓN DE LOS CONSUMOS ENERGÉTICOS EN LA INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO.
17. ANÁLISIS TEÓRICO – EXPERIMENTAL DEL DESEMPEÑO TÉRMICO DE UN HELIÓSTATO DE BAJA POTENCIA
18. MEMBRANA DE SEPARACIÓN CH4/CO2 PARA EL INCREMENTO DE EFICIENCIA ENERGÉTICA EN LA FOTOSÍNTESIS ARTIFICIAL
19. "MODELO PARA LA EVALUACIÓN DE LA GESTIÓN DE CIENCIA, TECNOLOGÍA E INNOVACIÓN EN ORGANIZACIONES"

POSTER

ORGANIZING COMMITTEE

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Apoyan:

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X

REVISTA cintex

INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO®

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

redin
revista facultad de ingeniería
Universidad de Antioquia
Hacia la Acreditación Institucional
ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex
Alcaldía de Medellín

COMPARATIVE ASSESSMENT OF COMPUTATIONAL MODELS FOR THE EFFECTIVE TENSILE STRENGTH OF NANO-REINFORCED COMPOSITES.

2021 EXPO TECNOLÓGICA

Participantes

Mateo Duarte García (1), Iván David Patiño (2), César Isaza Merino (3)
1.Grupo de Investigación e Innovación Ambiental (GIIAM), Colombia, mateo.duarte462@pascualbravo.edu.co
2.Grupo de Investigación e Innovación Ambiental (GIIAM), Colombia, i.patinoar@pascualbravo.edu.co
3.Grupo de Investigación e Innovación en Energía (GIIEN), Colombia, c.isaza2059@pascualbravo.edu.co

Keywords: Nanocomposites, Computational methods, Mechanical properties.

Resumen

Some of the most important industries as aeronautics, aerospace, automotive, military, energy, among others, have been stipulated new requirements for materials behavior [Agarwal, 2020; Benzait, 2018; Ferreira, 2018] that include high mechanical properties, thermal properties, and environmental friendliness [Vartak, 2020]. According to this, composites have emerged in order to satisfy some of these requirements; some of these materials are polymeric, metallic and ceramic matrix composites. Additionally, in the last decade nanocomposites emerged and different researches have shown a significant increase in mechanical properties due to the features of small-sized reinforcements. However, manufacturing these nano-composites implies cost and time-consuming problems that do not allow their use in technological applications, additionally, the lack of knowledge about the prediction of the mechanical properties is an obstacle for its technological implementation [Tekin, 2016].

Therefore, several researches have been focused in the development of computational models to predict the mechanical behavior of composites reinforced with nanofibers, nanoparticles and whiskers [Zare, 2020]. These models emphasize the understanding of nanocomposites phenomena for optimizing and predicting the mechanical behavior.

In the present work, a comparative assessment of the main computational models for predicting the tensile strength of nano-composites is carried out. Firstly, a new taxonomy of these models is proposed, which allows identifying the principal experimental variables, models evolution, and precision characteristics. The developed taxonomy suggests the classification of the existing models into three families: first generation, second generation, and third generation, as can be observed in Figure 1. First generation comprises the models that provide the fundamental basis of nano-composites mechanical behavior and starting points of calculation methods in material science, namely, Rule of Mixtures, Shear Lag model, and Halpin and Tsai model [Armbrister, 2015; Jagannatham, 2020]. Models of second generation includes modifications of the first generation models to improve the prediction accuracy by using fitting constants and phenomenological variables; George, Orowan-Asby and Kelly-Tyson are incorporated in this second family [Chen, 2017; Song, 2021]. Finally, third generation models account for other aspects as manufacturing characteristics, weighted combination methods, and additional phenomenological factors; models like Arithmetic Summation Method, Quadratic Summation Method and Compounding Method have place in this third family [Mohammed, 2020].

Once a new categorization is proposed, computational algorithms are developed for these models for predicting the tensile strength of nano-composites, accomplishing a comparative analysis of accuracy, robustness, and time-cost among them The precision of these models is evaluated deeming benchmark experimental works focused on the tensile strength characterization of nano-composites [Kwon, 2012; Stein, 2012]. These works were selected on the basis of the availability of experimental properties, fitting constants and phenomenological variables required by the models, which are summarized in the Table 1.

According to previous results, the trade-off between accuracy, computational cost, and properties requirements of the models of first, second and third generation is the most crucial issue to choose the suitable computational method for a determined application.

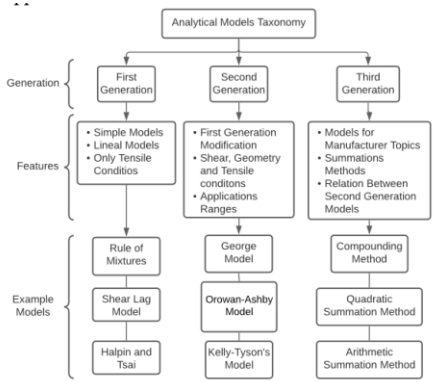


Figure 1: Developed taxonomy for effective tensile strength models of nanocomposites.

Table 1: Variables classification for effective tensile strength analytical models in nanocomposites (EP: Experimental Properties, FC: Fitting Constants, and PV: Phenomenological Variables).

Generation	EP	FC	PV
First	Tensile Strength, Volumetric Fraction	NA	NA
Second	Tensile Strength, Volumetric Fraction	Shear Strength, Thermal Expansion Properties	Reinforcement Direction, Geometry Factors
Third	Tensile Strength, Volumetric Fraction	Shear Strength, Thermal Expansion Properties, First, and Second Generation Conditions	Manufacture conditions, Reinforcement Direction, Geometry Factors

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

XVI SIMPOSIO INTERNACIONAL DE ENERGÍAS
ENERGÍA RENOVABLE

Apoyan:

redin
revista facultad de ingeniería
Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

EFFECT OF ADDITION MWCNT TO A MINERAL LUBRICANT ON DIESEL ENGINE PERFORMANCE

2021 EXPO TECNOLÓGICA

Participantes

Sebastian Meneses Munera1, J. S. Rudas2, Lina Marcela Hoyos Palacio3, A. Toro4, Jesús Antonio Carlos Cornelio5 María Isabel Ardila Marín6.
1Grupo GIIEN Institución Universitaria Pascual Bravo, Colombia, s.meneses1929@pascualbravo.edu.co; 2Grupo GIIEN Institución Universitaria Pascual Bravo, Colombia, ju.rudas@pascualbravo.edu.co; 3Grupo de Investigación Biología de Sistemas, Universidad Pontificia Bolivariana, Colombia, lina.hoyos@upb.edu.co; 4Tribology and Surfaces Group, Universidad Nacional de Colombia, Colombia, aotoro@unal.edu.co; 5Tribology and Surfaces Group, Universidad Nacional de Colombia, Colombia, iqcornelio@gmail.com; 6Grupo GIIEN Institución Universitaria Pascual Bravo, Colombia, isabel.ardila@pascualbravo.edu.co, Colombia.

Keywords: Tribology, Nanolubricants, Pin on disk

Resumen

Regulations becoming increasingly strict both in environmental and safety aspects for automotive industry has brought interest in developing internal combustion engine (ICE) lubricants both aiming to increase the fuel efficiency while decreasing wear to mobile pieces.
Multiwall carbon nanotubes (MWCNTs) have been found promising in tribological applications due to their remarkable mechanical and thermal properties (Zhai et al., 2017). MWCNTs have been widely studied in tribology applications such as ionic fluids (Yu et al., 2015), cutting fluid (Sharma et al., 2019), water-based substitutes for oil-based lubricant (Peng et al., 2007), AP/WP mineral lubricants (Cursaru et al., 2012), ICE evaluation using MWCNTs Nanolubricants have also been reported both in bio lubricant (Habibullah et al., 2014) and mineral based lubricant (Khalil et al., 2016) (Kotia et al., 2018).
Characterization of tribological and rheological properties contrasting various concentrations of MWCNTs is the main strategy to evidence effect of the nanotube's addition to the lubricant, being tribometers the most used experimental devices for lubricant evaluations either being pin on disk (Singh & Bhowmick, 2018), disc on disc (Cornelio et al., 2016) or block on disc (Lijesh et al., 2015).

Methods

This ongoing research aims to evaluate the effect when adding CNTs to a mineral oil lubricant for a Diesel engine by testing both tribological and rheological effects of CNTs adding to 15w40 oil lubricant.
MWCNT used are obtained by chemical vapor deposition (CVD) technique, using nickel catalyst at 700°C, in an arrangement according to Fig 1

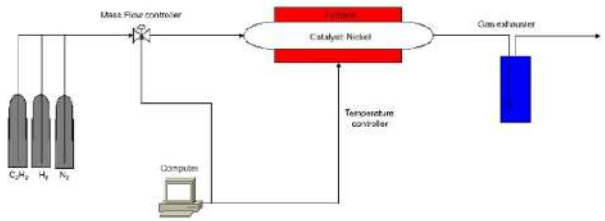


Fig 1. Schematic diagram of CVD for synthesis of CNTs.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

After, obtained nanotubes are purified and functionalized using acid leachates in order to increase dispersion capabilities of nanotubes (Cornelio et al., 2016)(Hoyos-Palacio, 2011).

For tribological characterization the
pin on disk tribometer (ASTM G-99) was used. The test configuration is present in Fig 2. As counter-pairs ASTM A48 steel as a pin on rotating ASTM A48 cast disk with normal load of 50N and at sliding speed of 1m/s. A 19ml of modified lubricant were enough to avoid spilling with the test conditions.

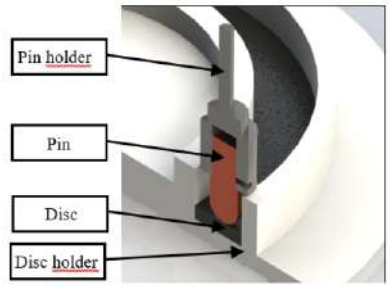


Fig 2: Pin on disc arrangement design
Viscosity is measured by a rotational viscosimeter using samples of 10ml for each concentration tested, a curve from room temperature around 25°C to 80°C is taken for every concentration tested.

Results and discussion

MWCNT concentration levels were set as 0%, 0.1%, 0.5% and 1% according to evidence showing no significant effect obtained from MWCNT over 1% (Joly-Pottuz et al., 2004)
Initial results evidence a general reduction in friction coefficient when CNTs are added to lubricant, suggesting a better performance in ICE could be expected as well.

Future work

Ongoing investigation proposes a factorial design and preliminary results correspond to initial runs.
Wear studies might be as well studied in post mortem test to tribometer samples such as mass comparison or micrography techniques.

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 7 13
2 8 14
3 9 15
16
4 10
17
5 11
18
6 12
19

ORGANIZING COMMITTEE

Apoyan:

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

XVI SIMPOSIO INTERNACIONAL DE ENERGÍA EXPOTECNOLÓGICA2021 (XVI SIE – 2021) "HEAT FLOW MODELING IN THE ALKALINE ACTIVATION PROCESS OF FLY ASH".

Participantes

Mauled Echeverri (1), J. S Rudas (2), J. Molina (3), A. Hoyos (4)
1. Grupo de investigación e innovación en energía – GiiEN, Colombia, mauled.echeverri481@pascualbravo.edu.co; 2. Grupo de investigación e innovación en energía – GiiEN, Colombia, ju.rudas@pascualbravo.edu.co; 3. Grupo de investigación e innovación en energía – GiiEN, Colombia, jarol.molina@pascualbravo.edu.co; 4. Grupo Investigación del Cemento y Materiales de Construcción -CEMATCO, Colombia, aahoyosm@unal.edu.co.

Keywords: Ordinary Portland Cement, alkaline cement, descriptive model.

Resumen

The growing social awareness to environmental impact of the production Ordinary Portland Cement (OPC) is certainly one of the main problems that affect the cement industry. The production of OPC is responsible for the emission of large quantities of Co2 (about 0.87 tonne CO2/tonne cement) [Elahi et al.,2020]. This carbon emission is related to the high energy consumption during manufacture. It is necessary to explore a potential solution to the negative environmental impact generated by the OPC.Studies reveal the alkaline activated cements (AAC) as a cementitious alternative with better manufacturing and application conditions as OPC cement [HoyosMontilla et al., 2021]. Therefore, the study of AAC from multiple approaches, both theoretical and experimental, is among the most relevant scientific and technological priorities worldwide [Elahi et al.,2020] [Zhang et al., 2020].

In this research, the dynamics of the heat flow associated with the process of alkaline activation of fly ash has been studied. A phenomenological model with a semiempirical base was developed in order to predict the performance of some variables that control the alkaline activation, that is to say, NaOH concentration (alkaline activator), heat flow and degree of reaction. The methodology developed was based on three stages: theoretical, simulation and validation. Theoretical stage, a systematic literature review was carried out for the purpose to find models that represent the process of alkaline activation. Simulation stage, a mathematical model is solved using commercial software (Matlab ®). Finally, the proposed model was validated at different operating points by comparing the simulation results of the model developed with models reported in the literature. The equation (1) represents a shrinking-core model [Thomas et al., 2011], the equation (2) represents the degree of reaction, and the equation (3) represents the accumulated heat in the process.

$$\frac{dr}{dt}=\frac{-C_0/a\rho}{\left[\frac{1}{k}+\frac{r_i^2}{D_i}\left(\frac{1}{r_i}-\frac{1}{R}\right)\right]+\frac{xr_i^2}{D_xR^2}+\frac{r_i^2}{D_0}\left(\frac{1}{R}-\frac{1}{r_0}\right)} \quad (1)$$

$$\alpha=1-\left(\frac{r_i}{R}\right)^3 \quad (2)$$

$$Q(t)=Q_{max}\alpha \quad (3)$$

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Where ri, the radius of the unreacted fly ash core. r0, the outside particle radius. R, the initial particle radius. a, number of moles of the alkaline solution reacting per mole of ash consumed. Dx, Di, D0, are the diffusivities through the barrier layer, inner and outer hydrate layers, respectively.C0, is the concentration of species in solution. k, is the first order reaction rate constant. ⊠, is the molar density of fly ash. x, thickness of the barrier formed in the latent period. Qmax, maximum heat released in the alkaline activation.

In this model we can see the kinetics variation of ash which gives account of the advance of the reaction and the degree of reaction in the alkaline activation. Based on this information and knowing the reactions that occur in the process, we can know the concentration of the species involved in alkaline activation. Likewise, the degree of hydration can be related to the heat flow that the system has during the process evolution. This work explores the potential of use a phenomenological model for representing and predicting the degree of hydration and heat flow which constitutes the basis for estimating the evolution of the reaction kinetics on AAC.

Acknowledgements

The authors acknowledges financial support from MINCIENCIAS: Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación "Francisco José de Caldas" (Convocatoria 848 del 2019 Programa de Estancias Postdoctorales) and from Institución Universitaria Pascual Bravo Grand 614, "Talento Pascualino"

References

Elahi, M. M. A., Hossain, M. M., Karim, M. R., Zain, M. F. M., & Shearer, C. (2020). A review on alkaliactivated binders: Materials composition and fresh properties of concrete. Construction and Building Materials, 260. https://doi.org/10.1016/j.conbuildmat.2020.119788 Hoyos-Montilla, A. A., Puertas, F., & Ivan Tobon, J. (2021). Study of the reaction stages of alkaliactivated cementitious materials using microcalorimetry. Advances in Cement Research, 33(1), 1–13. https://doi.org/10.1680/jadcr.19.00025 Thomas, J. J., Biernacki, J. J., Bullard, J. W., Bishnoi, S., Dolado, J. S., Scherer, G. W., & Luttge, A. (2011). Modeling and simulation of cement hydration kinetics and microstructure development. Cement and Concrete Research, 41(12), 1257–1278. https://doi.org/10.1016/j.cemconres.2010.10.004 Zhang, P., Gao, Z., Wang, J., Guo, J., Hu, S., & Ling, Y. (2020). Properties of fresh and hardened fly ash/slag based geopolymer concrete: A review. Journal of Cleaner Production, 270. https://doi.org/10.1016/j.jclepro.2020.12238

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 7 13
2 8 14
3 9 15
4 10 16
5 11 17
6 12 18
19

ORGANIZING COMMITTEE

Apoyan:

XVI SIMPOSIO INTERNACIONAL DE ENERGÍAS ENERGÍA RENOVABLE

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

ESTRATEGIAS DE ESTABILIZACIÓN DE LLAMAS DE MATERIAL RUMINAL SECO PARA APROVECHAMIENTO ENERGÉTICO

2021 EXPO TECNOLÓGICA

Participantes

Oscar Torres-Mercado (1), Andrés colorado (2), Andrés Amell (3), Edwin Chica (4), Jimmy Unfried (5)
1. Grupo de Investigación Ciencia y Tecnología del Gas y Uso Eficiente y Racional de la Energía -GASURE, Facultad de Ingeniería, Universidad de Antioquia, Calle 70 No 52-21, Medellín 050010, Colombia, oscara.torres@udea.edu.co; 2. Grupo de Investigación Ciencia y Tecnología del Gas y Uso Eficiente y Racional de la Energía -GASURE, Facultad de Ingeniería, Universidad de Antioquia, Calle 70 No 52-21, Medellín 050010, Colombia, felipe.colorado@udea.edu.co; ; 3. Grupo de Investigación Ciencia y Tecnología del Gas y Uso Eficiente y Racional de la Energía -GASURE, Facultad de Ingeniería, Universidad de Antioquia, Calle 70 No 52-21, Medellín 050010, Colombia, andres.amell@udea.edu.co; 4. Grupo de Investigación Energía Alternativa (GEA), edwin.chica@udea.edu.co, Facultad de Ingeniería, Universidad de Antioquia, Calle 70 No 52-21, Medellín 050010, Colombia; 3. Universidad de Córdoba, Carrera 6 No. 77-305 Montería, Córdoba, Colombia. jimyunfried@correo.unicordoba.edu.co.

Keywords: Biomasa, combustión, estabilidad, co-combustión, rumen, swirl

Resumen

Los procesos de sacrificio y obtención de la carne y otros subproductos tienen una alta demanda energética debida a los requerimientos de calidad y asepsia de estos; los subproductos obtenidos durante el sacrificio son diversos, entre ellos están la sangre, contenido gastrointestinal, piel, grasa, pelos y cerdas, cuernos, pezuñas y huesos, bilis y cálculos biliares. El contenido ruminal de un bovino adulto, puede pesar entre 30 y 60 kilogramos dependiendo del tiempo de transporte y reposo a que haya sido sometido . El contenido ruminal (MR) es el principal producto de desecho del proceso de beneficio en plantas de bovinos, su manejo y disposición final son un gran reto debido a que la presencia de cantidades importantes de MR en terrenos pantanosos y cuerpos de agua puede generar procesos de eutrofización .

La composición CHON del MR ruminal y algunas características pueden apreciarse en la , con un poder calorífico de 2445 kCal/kg se puede considerar el MR como combustible apto para aporte calórico en los procesos internos de las plantas de beneficio. La sustitución energética o la co-combustión requieren ajustes tecnológicos que permitan unos límites de estabilidad de llama con los que se pueda optimizar el aprovechamiento energético con el menor impacto ambiental posible. Una reacción es estable si permanece encendida (se autosostiene),emite bajas emisiones de monóxido de carbono (CO), hidrocarburos sin quemar (UHC) y dichas emisiones no fluctúan considerablemente con el tiempo .

Metodología

En el presente trabajo se analizan los resultados obtenidos mediante el análisis numérico de diferentes configuraciones propuestas para alcanzar la estabilidad de las llamas en el proceso de co-combustión de MR con gas natural. Se analiza la estabilidad bajo diferentes regímenes de swirl, combinando flujos tipo jet y flujos swirl, el impacto del factor de aireación y de la velocidad de entrada de las especies para diferentes porcentajes de sustitución energética.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Tabla 1 Propiedades del MR

Determinación	Como se recibe	Base seca	Método ASTM
Humedad total, (%)	11.96		D3302/D3302M -17
Cenizas, (%)	20.52	23.31	D3174-12 (2018)
Material volátil, (%)	54.73	62.16	D3175-18
Carbono fijo, (%)	12.79	14.53	D3172-13
Azufre, (%)	0.38	0.43	D5865-13
Poder calorífico, (Kcal/Kg)	2153	2445	D5865-13
Análisis ultimo (C.H.O.N)			D3176-15
Carbono, C, (%)		33.95	D5373-14e2
Hidrógeno, H, (%)		4.80	D5373-14e2
Oxígeno, O, (%)		36.50	D5373-14e2
Nitrógeno, N (%)		1.01	D5373-14e2

Conclusiones

La inclusión de MR tiende a modificar el tamaño y la morfología de la zona de reacción afectando la estabilidad de las llamas.
Los patrones de circulación anular desfavorecen el proceso de combustión generando inestabilidades y apagado de la llama.
El modelo de bajo Swirl permite un mayor porcentaje de sustitución de potencia con MR; sin embargo, es necesario analizar otras estrategias para garantizar una combustión de mayor eficiencia y evitar la emisión de especies contaminantes.

Agradecimientos

Los autores agradecen al Programa Colombia científica como fuente de financiación, en el marco de las convocatorias Ecosistema científico (Contrato No. FP44842- 218-2018).

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 7 13
2 8 14
3 9 15
4 10 16
5 11 17
6 12 18
19

ORGANIZING COMMITTEE

Apoyan:

XV Simposio Internacional de Energías Energía Renovable

redin

revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

Alcaldía de Medellín

RESPUESTA EN FRECUENCIA DE SISTEMA BESS ACOPLADO A RED DE BAJA INERCIA.

EXPO TECNOLÓGICA 2021

Participantes

Alejandro Franco Cano (1), Sergio Danilo Saldarriaga (2)

1. Institución Universitaria Pascual Bravo, Colombia, a.franco397@pascualbravo.edu.co;
2. Institución Universitaria Pascual Bravo, Colombia, s.saldarriagazu@pascualbravo.edu.co;

Resumen

La estabilidad de los sistemas de energía ha sido y sigue siendo de gran preocupación en su funcionamiento. Esto surge del hecho de que en estado estable la velocidad eléctrica media de todos los generadores debe permanecer igual en cualquier parte del sistema [1]. En los sistemas de energía eléctrica tradicionales, la inercia total (H) es la principal fuente de solidez del sistema eléctrico ante las perturbaciones de frecuencia que surgen debido a un desequilibrio de generación y demanda [2]. La transición a una sociedad baja en carbono es la fuerza motriz que impulsa el sistema de energía tradicional para aumentar el volumen de tecnologías no sincrónicas que utilizan principalmente convertidores de potencia (CP) como interfaz a la red eléctrica. Los CP desvinculan la fuente primaria de la red eléctrica, como consecuencia no son capaces de contribuir con la inercia "natural" de la misma manera que los generadores sincrónicos [3]. Debido a esto se buscan alternativas que permitan garantizar la alta penetración de energías renovables sin poner en riesgo la estabilidad del sistema, donde la regulación de frecuencia mediante sistemas BESS juega un rol importante.

Metodología

Este trabajo se aborda como una investigación aplicada descriptiva en la que se busca construir una guía metodológica que permita el análisis de estabilidad de frecuencia en sistemas con alta penetración de tecnologías de generación de energías renovables (Sistemas solares fotovoltaicos, Sistemas eólicos) lo que genera redes con baja inercia.

Mediante el uso del software especializado Digsilent Power Factory, se realiza un trabajo experimental en donde a través de un análisis inductivo por medio de un modelo de prueba Figura 1, se observa el comportamiento de las características eléctricas en especial la frecuencia, cuando el sistema de potencia es sometido a diferentes perturbaciones.

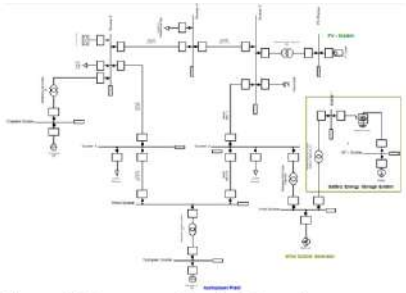


Figura 1: Diagrama unifilar Red de prueba

Resultados y simulaciones

La red consta de 7 elementos principales (generadores, transformadores, cargas, Barras, Líneas de transmisión, reactor y sistema BESS). se hace uso de la herramienta de análisis de estabilidad por medio de simulaciones dinámicas (RMS), donde se definen eventos de perdida de generación, constatando la respuesta de la variable de frecuencia en el tiempo, acoplando y desacoplando los controles asociados al sistema BESS.

La alta penetración de energías renovables trae consigo grandes retos en la operación de los sistemas eléctricos de potencia, es por esto que se hace imperativo el encontrar soluciones tecnológicas que cubran dichas necesidades, es allí donde el uso de servicios auxiliares a través de sistemas BESS juega un papel relevante. Si bien, la emulación de inercia como estrategia de control, es una de las posibles soluciones existentes en lo que respecta a estabilidad de frecuencia, se debe de tener especial cuidado con su aplicación en lo que se refiere a la medición de la frecuencia, puesto que los retardos en la lectura y la naturaleza fluctuante propia de dicha variable, puede generar efectos contrarios a los esperados.

Referencias

[1] AIE International Energy Agency. (2017). Energy Access Outlook 2017. Recuperado de https://www.iea.org/publications/freepublications/publication/WEO2017SpecialReport_EnergyAccessOut
lookresentación de las fuentes bibliográficas que sirvieron de apoyo para la construcción y desarrollo de la investigación (5 referencias).
[2] Zia, M. F., Benbouzid, M., Elbouchikhi, E., Muyeen, S. M., Techato, K., & Guerrero, J. M. (2020). Microgrid
transactive energy: Review, architectures, distributed ledger technologies, and market analysis. Ieee
Access, 8, 19410-19432.
[3] ALIEJA.J, S.M and GONAZALEZ LONGATT. F.M., 2016 investigation on grid-scale BESS providing inertial response support. IEEE Power & Energy Society (PES) sponsored International Conference on Power Systems Technology (POWERCON), Wollongong, NSW, Australia, September 28th October 1st 2016
[4]Andres Peña Asensio, Francisco Gonzalez-Longatt,Santiago Arnaltes and Jose Luis Rodriguez-Amenedo 2020, Analysis of the converter Synchronizing Method for the Contribution of Battery Energy Storage Systems to Inertia Emulation 20 March 2020
[5] Gonzalez-Longatt Francisco and Rueda Torres José Luis, Modelling and Simulation of Power Electronic Converter Dominated Power Systems in Power Factory
[6] Western Electricity Coordinating Council Modeling and Validation Work Group, WECC Battery Storage Dynamic Modeling Guideline Prepared By WECC Renewable Energy Modeling Task Force
[7] DIgSILENT Power Factory, Battery energy storing system (BESS) DIgSILENT Power Factory.
[8] S. Barsali and M. Ceraolo. Dynamical Models of Lead-Acid Batteries: Implementation Issues. IEEE Transactions on on Energy Conversion, 17(1), March 2002.
[9] M. Ceraolo. New Dynamical Models of Lead-Acid Batteries. IEEE Transactions on Power Systems,, 15(4), November 2000.

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

redin

revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

RESULTADOS DE LA IMPLEMENTACIÓN DE MÓDULO EMBEBIDO DE CONVERSIÓN DE TENSIONES EN MOTORES TRIFÁSICOS.

2021 EXPO TECNOLÓGICA

Participantes

AJortin De Jesús Vargas Ortega (1), Carlos Mario Moreno Paniagua (2)
Afiliación
1. Docente Ocasional de la Facultad de Ingeniería de la Institución Universitaria Pascual Bravo, Colombia, jortin.vargas@pascualbravo.edu.co; 2. Docente Ocasional de la Facultad de Ingeniería de la Institución Universitaria Pascual Bravo, Colombia, c.moreno1975@pascualbravo.edu.co

Keywords: Eficiencia energética, acometida monofásica, control de motores, inversor de voltaje.

Resumen

Se presentan los resultados del diseño y la implementación de un módulo electrónico embebido en los motores eléctricos de inducción, que permite la conexión de estos, en instalaciones monofásicas con la finalidad de incrementar la eficiencia en el uso de fuerza motriz en zonas no industriales [Enersinc, 2017]. En estos lugares es creciente la proliferación de pequeños negocios, que utilizan potencia eléctrica, alimentada desde acometidas monofásicas, que generalmente presentan una baja eficiencia en comparación con el uso de las trifásicas. La baja eficiencia está presente en la utilización de motores de inducción alimentados con tensión de una fase.

Métodos

El desarrollo en cuestión consiste en complementar motores trifásicos con dispositivos electrónicos de relativo bajo costo compuestos por rectificador, filtro, inversor, controlador y sensores, disminuyendo las pérdidas eléctricas e incrementando el factor de usabilidad en comparación con otros tipos de soluciones que actualmente se utilizan [Vieira, 2017], para lo cual se plantea un modelo de los motores [Giceva, 2018], se propone un esquema de operación eficiente, se diseña y construye un módulo embebido, se implementa en un motor de laboratorio de la institución y se realizan pruebas para establecer el nivel de incremento de la eficiencia y su correspondencia con los cálculos.

Resultados

Se ha construido el prototipo de un módulo embebido en un motor trifásico, para la prueba y cuantificación del nivel de ahorro de energía eléctrica en diferentes regímenes de trabajo, con respecto a otras técnicas de alimentación [Hareesh, 2017][Sobczynski,2017]. El módulo también varía la velocidad y protege el motor contra subidas de temperatura, sobre corrientes e inversiones de giro súbitas. Por tratarse de un dispositivo montado directamente sobre el motor, este queda protegido contra conexiones indebidas.

Discusión

Para resolver la problemática, se aplican diferentes técnicas, que van desde las poco técnicas, hasta las que tienen un alto componente tecnológico. El factor diferenciador con respecto a las primeras, como el uso de capacitores y roto-generadores, está en el incremento de la eficiencia [Goyal, 2014]. El diferenciador con las soluciones de alto componente tecnológico, está en la disminución del costo y en el factor de usabilidad que evita daños en los variadores de velocidad, debido a la mala manipulación por parte de operarios con competencias deficientes, que por lo general son los trabajadores de los pequeños talleres.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Referencias

CONPES-3867. (2016). Estrategia de preparación institucional para la paz y el postconflicto. Bogotá: Departamento nacional de planeación Colombia.

Enersinc - Departamento_nacional_de_planeación. (2017). Energy demand situation in Colombia, 136.

Giceva, I. Z., Sarac, V. J., Gelev, S. A., & Cingoski, V. T. (2018). Single phasing of three phase induction motors under various load conditions. In 2018 23rd International Scientific-Professional Conference on Information Technology (IT) (pp. 1–4). IEEE. https://doi.org/10.1109/SPIT.2018.8350459

Goyal, A. V., Agarwal, P., Bhatti, J. S., & Agarwal, V. (2014). Three phase induction motor drive using single-phase to three-phase cycloinverter with SVPWM technique. In 2014 Students Conference on Engineering and Systems (pp. 1–5). IEEE. https://doi.org/10.1109/SCES.2014.6880084

Hareesh, A., Manisankar, B., & Jayanand, B. (2017). A novel three phase infinite level inverter (TILI) topology for induction motor drive application. In 2017 Asian Conference on Energy, Power and Transportation Electrification (ACEPT) (pp. 1–8). IEEE. https://doi.org/10.1109/ACEPT.2017.8168619

Sobczynski, D., & Bartman, J. (2018). High-speed three phase motor supplied from a single to three phase converter. In 2018 XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH) (pp. 100–103). IEEE. https://doi.org/10.1109/MEMSTECH.2018.8365711

Vieira, T. E. D., Tofoli, F. L., & de Morais, A. S. (2017). Evaluation of IGBT modules in ASDs submitted to overvoltage. In 2017 IEEE 8th International Symposium on Power Electronics for Distributed Generation Systems (PEDG) (pp. 1–4). IEEE. https://doi.org/10.1109/PEDG.2017.7972497.

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

XVI SIMPOSIO INTERNACIONAL DE ENERGÍAS ENERGÍA RENOVABLE

redin

revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

MODELADO DE LA ADMISIÓN DE LA CARGA EN UNA METODOLOGÍA SECUENCIAL PARA COMBUSTIÓN HCCI

EXPO TECNOLÓGICA 2021

Participantes

Alejandro Franco Cano (1), Sergio Danilo Saldarriaga (2)

1. Institución Universitaria Pascual Bravo, Colombia, a.franco397@pascualbravo.edu.co;
2. Institución Universitaria Pascual Bravo, Colombia, s.saldarriagazu@pascualbravo.edu.co;

Alejandro Franco Cano (1), Sergio Danilo Saldarriaga (2)

1. Institución Universitaria Pascual Bravo, Colombia, a.franco397@pascualbravo.edu.co;
2. Institución Universitaria Pascual Bravo, Colombia, s.saldarriagazu@pascualbravo.edu.co;

Keywords: motores de combustión interna, HCCI, CFD, simulación numérica, cinética química.

Resumen

La combustión de carga homogénea de encendido por compresión (HCCI) es una opción atractiva para solucionar los problemas asociados al proceso de ignición y desarrollo de la combustión en las tecnologías de motores de combustión interna convencionales frente a combustibles con composiciones variables y la generación de contaminantes atmosféricos. La combustión HCCI presenta diferentes desafíos técnicos que han limitado su uso amplio. Los modelos que acoplan el análisis CFD y la cinética química mediante modelos multizona han permitido profundizar en los fenómenos que permiten la ocurrencia de la ignición, el desarrollo de la combustión y las condiciones que favorecen la generación de ciertos contaminantes, siendo una alternativa costo-efectiva para abordar los desafíos técnicos. En la presente ponencia se exponen los resultados de la evaluación del efecto de la inclusión del período de cruce válvulas y de la carrera de admisión sobre la distribución de la masa y la temperatura durante la carrera de compresión, y como esto afecta las predicciones del desarrollo de la combustión y la generación de contaminantes atmosféricos.

Métodos

Experimentalmente, se evaluaron dos condiciones de dosado relativo y temperatura de la carga en la admisión de un motor Diesel modificado de generación, de aspiración natural y operado a 1800 rpm, operando en modo HCCI utilizando biogás (60% gas natural – 40% dióxido de carbono) como combustible. Los experimentos se llevaron a cabo en el banco de motores de la Universidad de Antioquia (Medellín). Estas mismas condiciones se utilizaron para realizar una simulación secuencial CFD-cinética química detallada en el paquete ANSYS para modelar el cruce de válvulas y la carrera de admisión y su efecto sobre la estratificación térmica y másica de la carga durante la carrera de compresión, usando una metodología de reducción de zonas para el análisis de la combustión y la carrera de expansión. El nivel de ajuste entre los experimentos y la simulación se realizó comparando parámetros indicados, presión al interior del cilindro y concentración de contaminantes.

Resultados

Las condiciones de operación (régimen de giro y presión atmosférica) condujeron a altos niveles de temperatura en la admisión (baja densidad de carga) y una combustión retrasada (Figura 1), obteniéndose eficiencias menores a las esperadas con este modo de combustión (Figura 2). El modelado de la carrera compresión presenta un buen ajuste con los resultados experimentales, sin embargo, la condición de combustión retrasada, si bien permite un adecuado ajuste para diversos parámetros indicados, conduce a una deficiencia en la estimación de la generación de monóxido de carbono e hidrocarburos sin quemar (Figura 3).

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

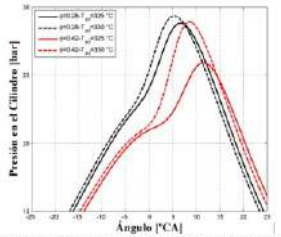


Figura 1. Curvas de presión en el cilindro para las condiciones evaluadas.

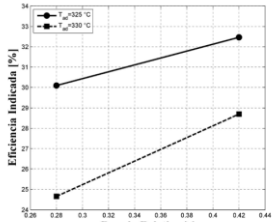


Figura 2. Eficiencia indicada para las condiciones evaluadas.

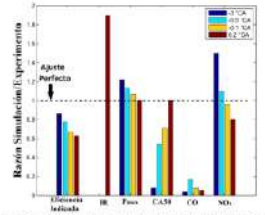


Figura 3. Ajuste entre simulación y experimento para diferentes variables. Φ=0.28 y T_{ad}=325 °C.

Discusión

La condición de operación retrasada reduce la capacidad de la metodología numérica para modelar adecuadamente el proceso de combustión. Sin embargo, esto no afecta el modelado de la carrera de compresión, en la cual, al incluir el cruce de válvula y la carrera de admisión, permite obtener los patrones de flujo y caracterizar la estratificación de la carga, con el fin de entender las condiciones que conducen a una operación ineficiente y que promueven el incremento/reducción de contaminantes.

Referencias

John E. Dec. Advanced Compression-Ignition Engines - Understanding the In-Cylinder Processes. Proceedings of the Combustion Institute, 32 II(2):2727-2742,2009.

Samveg Saxena and Iván Dario Bedoya. Fundamental Phenomena Affecting Low Temperature Combustion and HCCI engines, High Load Limits and Strategies for Extending These Limits. Progress in Energy and Combustion Science, 39(5):457-488,2012.

POSTER

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

XV Simposio Internacional de Energías Energía Renovable

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

Alcaldía de Medellín

SIMULATION OF THE THERMAL PERFORMANCE OF AN IGNITION ENGINE FUELED WITH GASOLINE/ISO-BUTANOL BLENDS

EXPO TECNOLÓGICA 2021

Participantes

Yuhan A. Lenis (1), Kenny A. Gallego (2), Álvaro Delgado Mejía (3), Andrés Morales Rojas (4), Anderson Gallego Montoya (5)
1. I. U. Pascual Bravo, Colombia, yuhan.lenis@pascualbravo.edu.co; 2. I. U. Pascual Bravo, Colombia, k.gallego@pascualbravo.edu.co; 3. I. U. Pascual Bravo, Colombia, alvaro.delgado@pascualbravo.edu.co; 4. I. U. Pascual Bravo, Colombia, andres.morales@pascualbravo.edu.co; 5. I. U. Pascual Bravo, Colombia, anderson.gallego@pascualbravo.edu.co

Keywords: Combustion, thermodynamic model, engine, gasoline, iso-butanol.

Resumen

Simulation of combustion process in internal combustion engines has had an important role in the field of engine development, because it brings a low associated cost and allows to predict some variations in parameters such as mechanical performance and gaseous emissions and to analyse non habitual operative conditions. Therefore, in this work, a zero dimensional thermodynamic model coupled to a thermochemical equilibrium model was used to modelling the iso-butanol + gasoline blends combustion in a spark ignition engine. It was carried out to determine the fuel blends effect on parameters such as, peak pressure, Indicated Mean Effective Pressure (IMEP), fuel consumption and gaseous emissions.

Materials and Methods

This work consisted of a simulation performed in four stages, in the first stage was considered one thermodynamic zone for burned and un-burned species. Second stage included compression and expansion strokes following isentropic processes. In Third stage wall losses were simulated through Woschni heat transfer model [1]. In Forth stage control volume with boundary work was considered. Finally, The energy release is calculated through Wiebe's burning law [2].

Moreover, an energy balance was applied to the control volume of the Figure 1, which is represented with respect to the crankshaft rotation angle is as shown in Equation 1.

$$\frac{dQ}{d\theta} = \frac{dU}{d\theta} + \frac{dW}{d\theta} = \frac{dU}{d\theta} + P\frac{dV}{d\theta} \qquad \text{Eq (1)}$$

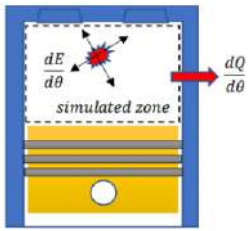


Figure 1. Scheme of the one zone thermodynamic Model

In order to obtain the burned mass fraction, Wiebe's burning law and is expressed as shown in the equation 2

$$x_b = 1 - \exp\left[-a\left(\frac{\theta - \theta_0}{\Delta\theta_0}\right)\right]^m \qquad \text{Eq (2)}$$

Results

The experimental values for both butanol and gasoline and the validation of the model based on the adjustment of the chamber pressure curves is shown in Figure 2.

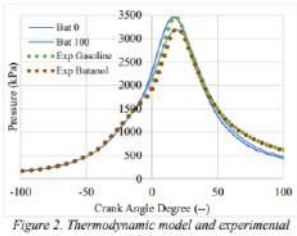


Figure 2. Thermodynamic model and experimental validation

Conclusions

A zero-dimensional thermodynamic model is presented to predict the combustion of isobutanol/gasoline blends inside an Otto cycle engine. According to the validation results, the model is adjusted to the experimental results derived from the use of these two fuels. The increase of iso-butanol in the blends leads to a reduction in chamber pressure and in the performance indicators, which is mainly due to the lower energy density of the iso-butanol fuel compared to gasoline. However, the maximum loss in benefits is close to 10%, which is associated with the air-fuel ratio, which is more favorable for the isobutanol. Although, similar operating conditions are achieved with both fuels, the fuel consumption will increase with that alcohol.

References

C. Rakopoulos and E. Giakoumis, "Secondlaw analyses applied to internal combustion engines operation," Prog. Energy Combust. Sci., vol. 32, no. 1, pp. 2–47, 2006, doi

J. B. Heywood, Internal Combustion Engine Fundamentals, 1st ed. United States of America: McGraw Hill, 1988.

C.-E. J., C. Guzman-Valdivia, O. DésigaOrenday, and M. García Ruiz, "Modelo Termodinámico De Una Zona Con Relaciones Para El Proceso De Combustión En Motores Alimentados Con Biogás," Rev. Mex. Ing. Química, vol. 12, no. 3, pp. 649–660, 2013.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

ESTUDIO COMPARATIVO DEL RENDIMIENTO DE UNA TURBINA TIPO TORNILLO DE ARQUÍMEDES EN ARREGLO TRADICIONAL Y EN ESPIRAL PARA APLICACIONES HIDROCINÉTICAS

Participantes

Ana Montilla (1), Laura Velásquez (1), Johan Betancour (1), Ainhoa Rubio-Clemente (1,2), Edwin Chica (1),

1. Grupo de Investigación Energía Alternativa (GEA), Facultad de Ingeniería, Universidad de Antioquia UdeA, Colombia, ana.montilla@udea.edu.co; lisabel.velasquez@udea.edu.co; johan.betancour@udea.edu.co; edwin.chica@udea.edu.co; 2. Facultad de Ingeniería, Tecnológico de Antioquia-Institución Universitaria TdeA, Colombia, ainhoarubioclem@gmail.com

Keywords: Tornillo de Arquímedes, Turbina hidrocinética, Sistema de energías renovables, Metodología de superficie de respuesta.

Resumen

Para garantizar un suministro energético confiable, sin instalaciones con grandes dimensiones y poco adaptables a regiones con reducida disponibilidad de espacio se han estudiado dispositivos como las turbinas hidrocinéticas, que aprovechan la energía cinética del agua, para producir electricidad [Yuce, 2015]. En este trabajo, se compara el comportamiento de dos turbinas tipo tornillo de Arquímedes (AST, por sus siglas en inglés) para aplicaciones hidrocinéticas, donde se estudian factores geométricos cruciales que afectan el diseño de la turbina como el paso y el ángulo de la hélice. Las turbinas estudiadas fueron una AST con configuración geométrica convencional y otra, en espiral, como se ilustra en las figuras 1 y 2, respectivamente.

Métodos

Una AST consta de un tornillo acoplado a un generador eléctrico y una estructura de soporte. En la configuración de tornillo de Arquímedes convencional se tiene un eje cilíndrico de longitud (L) con una o más hélices de paso (p), diámetro interior (Di) y diámetro exterior (Do) (ver figura 1). Las hélices se inclinan un ángulo (α) con respecto al eje axial. El tornillo en espiral cuenta con tres hélices separadas 120°, definidas por un diámetro exterior (De) y un p. Las hélices se inclinan un ángulo () con respecto al eje axial (ver figura 2).

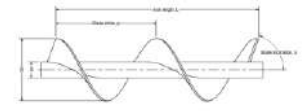


Figura 1. Factores geométricos de interés para el diseño del tornillo con configuración convencional.

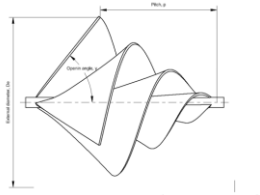


Figura 2. Factores geométricos de interés para el diseño del tornillo con configuración en espiral.

Para el estudio de cada turbina se empleó la dinámica de fluido computacional (CFD, por sus siglas en inglés) empleando la función definida por el usuario (UDF, por sus siglas en inglés) de 6 grados de libertad para modelar la rotación de las turbinas y la metodología de superficie de respuesta (RMS, por sus siglas en inglés), con el fin de obtener la configuración óptima de cada turbina dentro de un dominio experimental utilizando el menor número de experimentos [Bezerra, 2019]. Se determinó que, de acuerdo al número de factores involucrados en el diseño, fueron necesarios 27 experimentos para optimizar la configuración geométrica del tornillo convencional; mientras que el arreglo en espiral requirió tan solo 17 corridos experimentales debido al menor número de factores considerados. El efecto de cada factor geométrico seleccionado sobre la variable respuesta se investigó a tres niveles (-1, 0 y + 1). En las tablas 1 y 2 se presentan los valores correspondientes a estos niveles.

Tabla 1. Factores geométricos de interés para el diseño del tornillo con configuración convencional

Factor independiente	Valores		
	-1	0	+1
Longitud del eje, L (mm)	300	320	340
Paso, p (mm)	120	160	200
Ángulo de inclinación, α (°)	60	70	80
Relación de diámetros, Di/Do	0.2	0.4	0.6

Tabla 2. Factores geométricos de interés para el diseño del tornillo con configuración en espiral

Factor independiente	Valores		
	-1	0	+1
Paso, p (mm)	70	120	170
Diámetro exterior, De (mm)	100	150	200
Ángulo de apertura, γ (°)	30	45	60

La variable respuesta considerada para los análisis fue el coeficiente potencia (Cp), la cual se busca maximizar. Por lo tanto, para cada uno de los experimentos se determinó la eficiencia de las turbinas en términos de Cp, el cual se calculó a partir de la ecuación 1.

$$C_p = \frac{T\omega}{0.5\rho A V^3} \qquad (1)$$

Resultados y discusión

Como resultados se tuvo que para la configuración convencional se logró un valor de Cp igual a 0.5478, con los valores de los factores equivalentes a L, p, y Di/Do iguales a 320 mm, 160 mm, 70° y 0.2, respectivamente. En el caso del tornillo en espiral, el valor de Cp máximo se logró con la geometría p, y De equivalente a 168.9 mm, 51.34° y 245.6 mm, respectivamente, obteniendo un valor de Cp igual a 0.3474.

De acuerdo con los resultados, se observa que los valores de y Di/Do tuvieron la mayor influencia sobre el Cp obtenido en el caso del tornillo con configuración convencional. Una relación Di/Do baja conduce a un área de las hélices más grande; por lo tanto, se puede extraer una mayor cantidad de energía del fluido. Por su parte, en el tornillo en espiral, el De y toman influencia en el valor de Cp. Al tener un ángulo mayor, la turbina logra mayor contacto con el fluido y, por lo tanto, extrae más energía de este.

Referencias

M.A. Bezerra, S.L.C. Ferreira, C.G. Novaes, A.M.P. Dos Santos, G.S. Valasques, U.M.F. da Mata Cerqueira, J.P. dos Santos Alves, Simultaneous optimization of multiple responses and its application in analytical chemistryea review, Talanta 194 (2019) 941e959.

M.I. Yuce, A. Muratoglu, Hydrokinetic energy conversion systems: a technology status review, Renew. Sustain. Energy Rev. 43 (2015) 72e82.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

redin · revista facultad de ingeniería Universidad de Antioquia · Hacia la Acreditación Institucional · Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 7 13
2 8 14
3 9 15
4 10 16
5 11 17
6 12 18
19

ORGANIZING COMMITTEE

Apoyan:

redin
revista facultad de ingeniería
Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

NEW APPROACH FOR FILLING SIMULATION OF DUAL-SCALE PREFORMS USED IN THE MANUFACTURING OF COMPOSITES MATERIALS.

Participantes

Iván David Patiño Arcila (1)

1. Grupo de Investigación e Innovación Ambiental (GIIAM), Colombia, i.patinoar@pascualbravo.edu.co.

Keywords: Fibrous reinforcements, Liquid Composite Molding, Multi-scale filling.

Resumen

Some fibrous reinforcements used in the manufacturing of parts by Liquid Composite Molding (LCM) have a dual-scale nature, which supposes flow imbalances between the bundles and channels at mesoscopic scale, which in turn, cause uncontrolled defects (voids, dry points) and could considerably affect the global flow behavior during the filling of cavities at macroscopic scale. There are two main numerical strategies to tackle this problem: to carry out simultaneous and iterative-corrected simulations at both scales, or to conduct several simulations at the mesoscopic scale for obtaining lumped functions to be used at macroscopic scale. The second strategy is employed in the present work, but it is introduced an important modification in the approach to simulate the liquid absorption into the bundles, which is motivated by some physical incongruences between numerical (Simacek & Advani, 2003, 2007; Tan & Pillai, 2010; Wang & Grove, 2008) and experimental results (Abraham & McIlhagger, 1998; Ben Abdelwahed et al., 2011; Eun & Lee, 2003; Gourichon et al., 2006; Kang & Koelling, 2004; Kedari et al., 2011; Lawrence et al., 2009; Vilá et al., 2015).

Accordingly, in the traditional approach, it is supposed that the tows saturation rate inside the RUC is function of the uniform liquid pressure contained in the channels (Simacek & Advani, 2003, 2007; Tan & Pillai, 2010; Wang & Grove, 2008). On the other hand, the approach proposed here consists of prescribing a pressure gradient along the RUC and imposing Stokes-Darcy matching conditions between the tows and the channel sub-domains to determine the filling of the former ones; applying mass conservation, the saturation rate is established in terms of the difference between the inlet and outlet flow rates of the RUC.

To illustrate the difference between both approaches, two numerical simulations of the RUC filling are compared to each other. Filling simulations of Figs. 1a-c are carried out with the constant-pressure approach, whereas simulations of Fig. 2a-k, with the proposed approach. In both simulations. thev are shown: the ratio of the current time to total simulation time

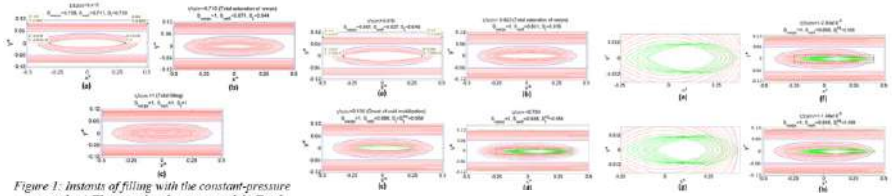


Figure 1: instants of filling with the constant-pressure approach. a) Warps and weft unsaturated. b) Total saturation of warps, c) Total saturation of the RUC

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

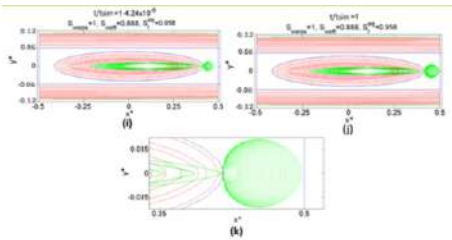


Figure 2: Instants of filling with the proposed Stokes-Darcy methodology assuming full air compressibility. a) Warps and weft unsaturated, b) Total saturation of warps, c) Instant of end of void compression and the onset of void mobilization, d) Motion towards the right extreme of the weft, e) Detail of Figure 6d, f) Arrival of a bubble to right extreme of the weft and onset of void migration, g) Detail of Figure 6f, h) Stage 1 of void migration, i) Stage 2 of void migration, j) Stage 3 of void migration (until the bubble is in the neighborhood of the RUC's edge), k) Detail of void migration.

References

Abraham, D., & McIlhagger, R. (1998). Investigations into various methods of liquid injection to achieve mouldings with minimum void contents and full wet out. Composites Part A, 29, 533–539.

Ben Abdelwahed, M. A., Wielhorski, Y., Bizet, L., & Breard, J. (2011). Bubble Shape and Transport During LCM Processes: Experimental Modeling in a T-Junction Tube. 18th International Conference on Composites Materials, Jeju Island, Korea, 1–6. http://arxiv.org/abs/1207.4099

Eun, S. W., & Lee, W. (2003). Experimental study of the microvoids formation and transport in the vacuum assisted resin transfer molding process. Composites Research, 16(6), 10–15. http://s-space.snu.ac.kr/handle/10371/33769

Gourichon, B., Binetruy, C., & Krawczak, P. (2006). Experimental investigation of high fiber tow count fabric unsaturation during RTM. Composites Science and Technology, 66(7–8), 976–982. https://doi.org/10.1016/j.compscitech.2005.07.032

Kang, K., & Koelling, K. (2004). Void transport in resin transfer molding. Polymer Composites, 25(4), 417–432. https://doi.org/10.1002/pc.20035

Kedari, V. R., Farah, B. I., & Hsiao, K.-T. (2011). Effects of vacuum pressure, inlet pressure, and mold temperature on the void content, volume fraction of polyester/e-glass fiber composites manufactured with VARTM process. Journal of Composite Materials, 45(26), 2727–2742. https://doi.org/10.1177/0021998311415442

Lawrence, J. M., Neacsu, V., & Advani, S. G. (2009). Modeling the impact of capillary pressure and air entrapment on fiber tow saturation during resin infusion in LCM. Composites Part A: Applied Science and Manufacturing, 40(8), 1053–1064. https://doi.org/10.1016/j.compositesa.2009.04.013

Patiño Arcila, I. D., Power, H., Nieto Londoño, C., & Flórez Escobar, W. F. (2016). Boundary element simulation of void formation in fibrous reinforcements based on the Stokes–Darcy formulation. Computer Methods in Applied Mechanics and Engineering, 304, 265–293. https://doi.org/10.1016/j.cma.2016.02.010

Patiño, I. D., Power, H., Nieto-Londoño, C., & Flórez, W. F. (2017). Stokes–Brinkman formulation for prediction of void formation in dual-scale fibrous reinforcements: a BEM/DR-BEM simulation. Computational Mechanics, 59(4), 1–23. https://doi.org/10.1007/s00466-016-1360-5

Patiño, I., Power, H., Nieto, C., & Flórez, W. (2018). Boundary Element Method for the dynamic evolution of intra-tow voids in dual-scale fibrous reinforcements using a Stokes–Darcy formulation. Engineering Analysis with Boundary Elements, 87, 133–152. https://doi.org/10.1016/j.enganabound.2017.11.014

Simacek, P., & Advani, S. G. (2003). A numerical model to predict fiber tow saturation during liquid composite molding. Composites Science and Technology, 63(12), 1725–1736. https://doi.org/10.1016/S0266-3538(03)00155-6

Simacek, P., & Advani, S. G. (2007). Modeling resin flow and fiber tow saturation induced by distribution media collapse in VARTM. Composites Science and Technology, 67(13), 2757–2769. https://doi.org/10.1016/j.compscitech.2007.02.008

Tan, H., & Pillai, K. M. (2010). Modeling Unsaturated Flow in Dual-Scale Fiber Mats of Liquid Composite Molding: Some Recent Developments. In V. Michaud & P. Ermanni (Eds.), The 10th International Conference on Composite Materials (FPCM10) (pp. 1–5).

Vilá, J., Sket, F., Wilde, F., Requena, G., González, C., & LLorca, J. (2015). An in situ investigation of microscopic infusion and void transport during vacuum-assisted infiltration by means of X-ray computed tomography. Composites Science and Technology, 119, 12–19. https://doi.org/10.1016/j.compscitech.2015.09.016

Wang, Y., & Grove, S. M. (2008). Modelling microscopic flow in woven fabric reinforcements and its application in dual-scale resin infusion modelling. Composites Part A: Applied Science and Manufacturing, 39(5), 843–855. https://doi.org/10.1016/j.compositesa.2008.01.014

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

EVALUACIÓN DE CONFIABILIDAD EN MICROGRIDS: COMPARACIÓN DE MÉTODOS DE MODELAMIENTO.

Participantes

Jose López (1), Guisselle García (2), Jorge Vélez (3)
1. Universidad de Sucre, Colombia, Jose.lopez@unisucre.edu.co; 2. Universidad del Norte, Colombia, gagarcia@uninorte.edu.co; 3. Universidad del Norte, Colombia, jvelezv@uninorte.edu.co

Keywords: Microgrids, Generación Distribuida, Simulación de Monte Carlo, Markov

Resumen

La sociedad moderna se enfrenta a un sistema energético insostenible. Por lo que, en los últimos años han surgido soluciones innovadoras, con la finalidad de modernizar los sistemas de distribución (SD). El concepto de Microgrids (MGs) es clave dentro de estas soluciones, pues la implementación de estas mejoraría la calidad del suministro, la eficiencia, la confiabilidad y proporcionaría una energía más limpia y a menores costos [Yoldas et al, 2017].

A pesar de sus ventajas, los investigadores en el área de las MGs en la actualidad se enfrentan a grandes retos. Ejemplo de ello, es la evaluación de la confiabilidad de SD con aplicaciones de MGs, que corresponde a uno de los desafíos más importantes y complejos encontrados [Yoldas et al, 2017], [López et al, 2020].

Métodos

Para el análisis de confiabilidad, éste trabajo estudia una colección de métodos reportados en la literatura para el modelamiento de sistemas estocásticos. Allí se reportan ventajas y desventajas. Con ello se presenta una selección del método que mejor podría aportar en el caso de MGs.

Resultados

Según reportes de la literatura, los métodos podrían dividirse en dos categorías, métodos analíticos y simulación de Monte Carlo (SMC). Después de su análisis se infiere que los métodos analíticos se vuelven inviables para SD con MGs (reales) complejos. Sin embargo, los métodos analíticos tienen un papel importante que jugar en la evaluación de confiabilidad del SD con MGs, se identificó que la tendencia es usar métodos avanzados (aproximados), los cuales pueden ser híbridos entre diferentes métodos [Elmakias, 2011].

Discusión

Para los modelos estudiados los resultados revelan que, para sistemas a gran escala y cuando se involucran condiciones de operación complejas, siendo este el caso de los SD con MGs, el método de simulación de Monte Carlo es a menudo preferible sobre los métodos analíticos. Esto debido a las siguientes características: fuerte adaptabilidad, estructura simple del programa y dimensiones del sistema de estudio irrelevantes.

Por otro lado, dado que la producción de energía fotovoltaica y energía eólica en las MGs, son aleatoria e intermitente, no es fácil evaluar la capacidad de suministro de energía de estos generadores distribuidos (DG) mediante un método analítico, siendo más adecuado la SMCS. Además, este método puede simular la salida aleatoria de las DG en una secuencia de tiempo, logrando establecer un modelo de probabilidad más realista.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Referencias

Yolda☒, Y., Önen, A., Muyeen, S. M., Vasilakos, A. V., & Alan, ☒. (2017). Enhancing smart grid with microgrids: Challenges and opportunities. Renewable and Sustainable Energy Reviews, 72, 205-214.

López-Prado, J. L., Vélez, J. I., & Garcia-Llinás, G. A. (2020). Reliability Evaluation in Distribution Networks with Microgrids: Review and Classification of the Literature. Energies, 13(23), 6189.

D. Elmakias, New Computational Methods in Power System Reliability. 2011.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

redin

revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

THERMAL FINITE ELEMENT ANALYSIS IMPLEMENTED WITH OPEN SOURCE TOOLS AND HIGH-PERFORMANCE COMPUTING

EXPO TECNOLÓGICA 2021

Participantes

Esteban Foronda (1), Álvaro Delgado (2), Luis Olmos (3), Francisco Ramírez (4)
Institución Universitaria Pascual Bravo, Medellín - Colombia
1. esteban.foronda@pascualbravo.edu.co; 2. alvaro.delgado@pascualbravo.edu.co; 3. luis.olmos@pascualbravo.edu.co; 4. francisco.ramirez@pascualbravo.edu.co

Keywords: FEA; Python; Firedrake; Heat sink; Heat transfer; HPC.

Resumen

The modelling of heat transfer phenomena in thermal systems has been thoroughly explored in industry and academia by using commercial software. However, when the thermal problem introduces complexities in geometry and physics, the availability of licenses for high-performance computing might represent a limitation to achieve efficient results. Thus, finite element analysis using open source tools platforms hasbecome a successful methodology to solve numerical problems. In this work, multiple open source tools are integrated to solve the heat transfer equation for heat sinks with application in electronics and assessing the performance of parallel computing to reduce the processing time.

Methodology

A complete finite element analysis has been implemented to solve the weak form of the heat transfer equation including conduction, and surface-toambient convection and radiation. The preprocessing includes the geometry definition from a CAD file and the mesh creation using the open source mesh generator GMSH [Geuzaine & Remacle, 2009]. The solution process is performed in Firedrake [Rathgeber et al., 2016] through its Application Programming Interface (API) for Python. High-performance computing is achieved by using the MPI (Message Passing Interface) capabilities of the solver PETSc [Abhyankar et al., 2018] to parallelize the solution operations in multiple CPUs. The results are postprocessed in Paraview [Ahrens et al., 2005]. The heat transfer of a complex heat sink for application in electronics has been modelled as a case study to test the proposed methodology using tenths of millions of finite elements. A comparison study is made by changing several simulation parameters such as FEA model size and parallel settings. Finally, our proposed open source tool is compared with commercial software.

Results

The temperature field of a heat sink for electronics has been assessed with the indicated Python open source tools, as shown in Figure 1. The selected geometry contained microchannels to enhance the cooling capabilities of the heat sink, which required local mesh refinement to capture the small radius curvature. Thus, the number of degrees of freedom increased significantly during the mesh convergence analysis and the computation time was compared for different configurations of parallelism used to solve the problem.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Discussion

Open source tools can be integrated successfully in Python scripts to solve differential equations with the finite element method. Firedrake and its solver PETSc are robust enough to implement high-performance computing for large-scale problems for both linear and nonlinear solvers and maximize the available hardware resources of workstations and clusters. The proposed methodology has proven successful for heat transfer problems but can be extended to other physics. Due to the fact that these open source tools are fairly new (compared to traditional commercial software), the biggest challenge resides in the initial configuration of each library and the setup to enhance the performance of the hardware. After the configuration stage is verified successfully, this methodology can be easily implemented to other engineering problems, both in academia and industry. Thus, the authors herein consider relevant to increase the publications around this topic, in such a way that these open source tools are extended to other applications and research fields.

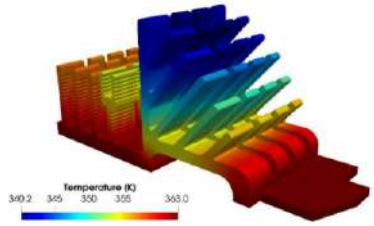


Figure 1: Temperature field of a heat sink with microchannels

References

Abhyankar, S., Brown, J., Constantinescu, E. M., Ghosh, D., Smith, B. F., & Zhang, H. (2018). PETSc/TS: A Modern Scalable ODE/DAE Solver Library. Numerical Analysis.Ahrens, J., Geveci, B., & Law, C. (2005). ParaView: An End-User Tool for Large Data Visualization. Elsevier.Geuzaine, C., & Remacle, J.-F. (2009). Gmsh: a threedimensional finite element mesh generator with built-in pre- and post-processing facilities. International Journal for Numerical Methods in Engineering, 79(11), 1309–1331.Rathgeber, F., Ham, D. A., Mitchell, L., Lange, M., Luporini, F., McRae, A. T. T., Bercea, G.-T., Markall, G. R., & Kelly, P. H. J. (2016). Firedrake: automating the finite element method by composing abstractions. ACM Transactions on Mathematical Software, 43(3), 24:1--24:27. https://doi.org/10.1145/299844

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

PROTOTIPO DE RED PARA LA IDENTIFICACIÓN Y SEGUIMIENTO DE PERSONAS EN UN ENTORNO EDUCATIVO CON IoT.

Participantes

Bayron Ospina Cifuentes1, Alber Montoya Benitez2, Diego Saldarriaga3, Samuel Quiroz4,

1. Docente Facultad de Ingenierías Instituto Tecnológico Metropolitano, Colombia, bayronospina@itm.edu.co; 2. Docente Facultad de Ingenierías Instituto Tecnológico Metropolitano, Colombia, albermontoya@itm.edu.co; 3. Estudiante Tecnología en Gestión de Redes de Telecomunicaciones Instituto Tecnológico Metropolitano, Colombia, diegosaldarriaga281327@correo.itm.edu.co; 4. Estudiante Tecnología en Gestión de Redes de Telecomunicaciones Instituto Tecnológico Metropolitano, Colombia, samuelquiroz280631@correo.itm.edu.co.

Keywords: Inteligencia Ambiental, Internet de las Cosas, Sensorizado, Seguimiento a Personas.

Resumen

Las Instituciones Educativas no cuentan con un sistema que permita conocer el aforo y la ubicación en tiempo real de las personas que a diario ingresan a sus campus universitarios. Este tipo de información es relevante para localizar o enviar información relacionada con en eventos críticos como emergencia o desastres naturales, lo que permite tener planes de evacuación sistematizados y dinámicos en función de la afluencia de personas. Adicionalmente, la trazabilidad de las personas se puede usar para controlar la concentración de personas en espacios cerrados, algo que se ha visto necesario en los últimos meses con la pandemia del COVID 19.

A este respecto, existen herramientas, como la inteligencia ambiental (Aml), basada en un elemento computacional que se enfoca en brindar calidad de vida y mejor ambiente laboral en entornos empresariales. Aml, se refiere a redes inteligentes que reconocen la presencia y las preferencias humanas y se ajustan a entornos inteligentes para satisfacer las necesidades y requerimientos inmediatos. El factor diferenciador es la presencia de módulos inteligentes para la toma de decisiones en los entornos de IoT. Las tecnologías subyacentes incluyen computación generalizada, comunicación, ubicación, conectividad perfecta de dispositivos inteligentes, redes de sensores, inteligencia artificial (AI), machine learning (ML), y en general la interacción humano-computadora (HCI). Las aplicaciones y escenarios de Aml con IoT incluyen hogares inteligentes, vehículos autónomos sin conductor, sistemas de salud, carreteras inteligentes, sector industrial, gestión inteligente de instalaciones, sector educativo y servicios de emergencia como se presenta en este proyecto .

De acuerdo a los beneficios de tener un sistema automático de control de aforo y ubicación en tiempo real de personas basado en Aml, en este trabajo, se propone un prototipo de red que simula un entorno educativo monitoreado por sensores para la identificación y seguimiento de personas simuladas como objetos basados en Internet de las Cosas, lo que brinda una referencia para el monitoreo y asistencia de las personas en entornos educativos.

La metodología usada se compone de las fases operacional, experimental y evaluación. La fase operacional, considera el proceso de diseño e implementación del prototipo de red que permita simular el entorno educativo con Aml basado en IoT. La fase experimental, se basa en un método de análisis estadístico del proceso experimental . La fase de evaluación comprende la verificación y análisis de resultados, lo que permite identificar problemas y confiabilidad del sistema de monitoreo propuesto.

Los resultados del proyecto muestran la confiabilidad en el control de asistencia, control de aforo, reducción en un 70% del tiempo para verificar el número de personas en un espacio determinado. Además, se puedo establecer en tiempo real la ubicación y hacer seguimiento de las personas que ingresan al recinto con una exactitud superior al 90%.

Referencias

Montebello, M. Ambient intelligence. In A. Kameas & K. Stathis (Eds.), Studies in Computational Intelligence (Vol. 840, pp. 11–34). Springer International Publishing. https://doi.org/10.1007/978-3-030-21882-9 2, 2019.

Montgomery, D. C. Design and Analysis of Experiments - Douglas C. Montgomery-Ninth Edition. https://books.google.es/books?hl=es&lr=&id=Py7bDgAAQBAJ&oi=fnd&pg=PA1&dq=D.C.+Montgomery.+Design+and+Analysis+of+Experiments,+5a+edición.+Wiley,+2001.+%5B396%5D+R.L.+Plackett,+J.P.+Burman.+Biometrika,+33+(1946)+305-325.+%5B397%5D&ots=X6x3oZJM5a&sig=8sCYfs, 2017.

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

ORGANIZING COMMITTEE

Apoyan:

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

Alcaldía de Medellín

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

XV SIMPOSIO INTERNACIONAL DE ENERGÍAS ENERGÍA RENOVABLE

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

MODELO DINÁMICO DE UN SISTEMA DE GASIFICACIÓN DE LECHO FIJO ESTRATIFICADO

EXPO TECNOLÓGICA 2021

Participantes

Yuhan Arley Lenis (1), José Reinaldo Garzón (2), Iván Portnoy de la Ossa (3), Arnaldo Verdeza Villalobos (4), Celso Lins de Oliveira (5)
1. Institución Universitaria Pascual Bravo, Colombia, yuhan.lenis@pascualbravo.edu.co;
2. Institución Universitaria Pascual Bravo, Colombia, jose.garzon044@pascualbravo.edu.co;
3. Corporación Universidad de la costa, Colombia, iportnoy@cuc.edu.co; 4. Universidad Simón Bolívar, Colombia, averdeza@unisimonbilivar.edu.co; 5. Universidade de São Paulo - USP/ FZEA, Brasil, celsooli@usp.br

Keywords: Biomasa, reactor de lecho fijo, control, modelo de diagnóstico

Resumen

Los problemas ambientales asociados al uso de recursos de origen fósil, han fomentado el desarrollo de tecnologías de transformación de recursos renovables. Es el caso de los sistemas de gasificación, que si bien fueron desarrollados hace cerca de 200 años, durante las últimas décadas se viene reconsiderando su uso para la transformación de materiales carbonosos. Este proceso permite transformar un combustible sólido en una mezcla de gases combustibles de bajo poder calorífico que pueden ser usados en las máquinas térmicas convencionales [Tinuat, 2008]. El diseño de los sistemas de gasificación de lecho fijo es bastante simple, siendo el componente más crítico, el empleado para el movimiento de la carga al interior reactor y la ubicación del frente de llama. En este sentido, los equipos comerciales de bajo costo, emplean sistemas de control on-off que, aunque logran mantener bajo control el proceso, inducen inestabilidades que afectan la calidad del gas producido [Verdeza, 2019]. Es así como en este estudio se presenta el desarrollo de un modelo dinámico de un reactor de lecho fijo, que puede ser empleado para el diseño y optimización de estrategias de control continuo en lazo cerrado, que procuren una mejora en la calidad del gas producido y la estabilidad del proceso.

Métodos

El modelo considera que la velocidad del frente de llama responde como una función de primer orden más tiempo muerto ante las variaciones del flujo volumétrico ⊠) de aire. De igual manera se simulan como perturbaciones los efectos de los cambios de densidad ⊠), humedad de la biomasa ⊠) y el encendido del sistema de vibración ⊠), ver Figura 1.

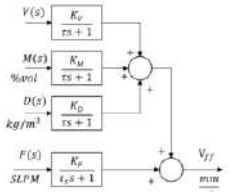


Figura 1: Diagrama del modelo del sistema de gasificación.

Las ganancias de las funciones empleadas fueron obtenidas a partir de pruebas experimentales y de resultados reportados en [Lenis, 2013], mientras que las variables relacionadas con la dinámica son derivadas de los resultados experimentales reportados

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Resultados y discusión

La Figura 2, muestra la respuesta del modelo ante perturbaciones toda vez que el frente de llama está en una posición estable. La variación en el flujo de aire ocasiona un aumento en la velocidad de propagación del frente de llama, por lo que, con esta perturbación, se tiene un desplazamiento hacia la parte superior del reactor. Cuando desaparece la perturbación, el frente de llama vuelve a una posición estable.

Comportamientos similares se obtienen para las demás perturbaciones. La figura 2, muestra la respuesta ante la intermitencia en la agitación, resultado consistentecon lo reportado previamente en [Verdeza, 2019]. De esta manera se concluye que, el modelo predice adecuadamente el desempeño del proceso en modo dinámico, resaltándose su utilidad para el desarrollo de sistemas de control, lo que eventualmente facilitará el desarrollo local y la masificación de la tecnología.

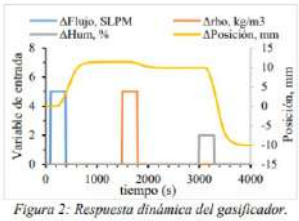


Figura 2: Respuesta dinámica del gasificador.

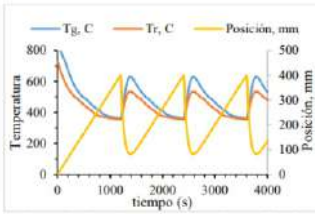


Figura 3: Respuesta ante variación on-off en la agitación.

Referencias

Verdeza, A., Lenis, Y. A., Bula, A., Mendoza, J., and Gomez, R., 2019, "Performance Analysis of a Commercial Fixed Bed Downdraft Gasifier Using Palm Kernel Shells," 9(December), pp. 79–88.

Lenis, Y. A. A., Osorio, L. F. F., and Pérez, J. F. F., 2013, "Fixed Bed Gasification of Wood Species with Potential as Energy Crops in Colombia: The Effect of the Physicochemical Properties," Energy Sources, Part A Recover. Util. Environ. Eff., 35(17), pp. 1608–1617

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

XV SIMPOSIO INTERNACIONAL DE ENERGÍAS RENOVABLES

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

EMOTION DETECTION FROM GALVANIC SKIN RESPONSE USING MACHINE LEARNING IN A HUMAN-ROBOT SCENARIO

2021 EXPO TECNOLÓGICA

Participantes

YSebastian Giraldo-Betancur (1), Rubén D. Fonnegra (2)
1. Institución Universitaria Pascual Bravo, Colombia,
s.giraldo1230@pascualbravo.edu.co;
2. Institución Universitaria Pascual Bravo, Colombia,
ruben.fonnegra@pascualbravo.edu.co

Keywords: Emotion Detection, Human-Robot Interaction, Machine Learning

Resumen

Affective computing has been a growing research field during recent years since lots of human's decisions are based on their emotions and judgements (Pei et al., 2021.). As a consequence, this area has been explored from the analysis of multiple information resources considering the different physical manifestation of emotions. This work proposes the use of galvanic skin response (GSR) to estimate the emotion in participants during a human-robot interaction. To achieve this, an experiment is proposed during a human-human and human-robot interaction under a Jenga game setup, and a stimulus is elicited (throw or not throw the tower). GSR signals and Positive and Negative Affect Schedule (PANAS) scores are recorded (Wedderhoff et al., 2021). Then, multiple machine learning models are used to estimate the emotional affection under previous conditions. Results for machine learning models show effectiveness to estimate the emotional state of participants.

Methodology

The experimental design consists in an evaluation considering the emotional reaction (throw/no throw) and interaction (human/robot). GSR signals are taken from 23 different participants, after a relaxation phase. Thereupon, signals are processed to extract statistical features (Table 1) using the TEAP toolbox (Soleymani et al., 2017) and labelled according to PANAS scales as positive or negative affection (valence). After feature extraction, different machine learning algorithms were employed to recognize emotional affection. Among them, an artificial neural network (ANN), a support vector machine (SVM), a hierarchical decision tree (HDT), a k nearest neighbours classifier (kNN) and a logistic regression (LR) were used to determine the user's intention of emotion.

Table 1: Statistical features extracted from GSR signals

Features extracted - Num of features
Maximum and minimum peak values of the signal (2)
Mean, Standard deviation and variance of the signal (3)
Number of local minima (1)
Number of local maxima (1)
The value of the amplitude of the signal (1)
Average between each pair-wise local minima-maxima across resistance (1)
Average between each pair-wise local minima- maxima across time (1)

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Results

To analyze results, several approaches are considered for the PANAS scores and machine learning models.

Emotional analysis.

Emotional analysis consists in estimating the behaviour of participants for throwing or no-throwing conditions during the experiment (reaction). Results for the first setup are displayed in Figure 1, where there exists a significant difference in PANAS scores after and before the experiment for participants whose tower was thrown during the trial. On the other hand, results for the machine learning models in terms of accuracy for the first setup are presented in Figure 2. The models show effectiveness to detect emotional affection in participants, obtaining best performance in the ANN.

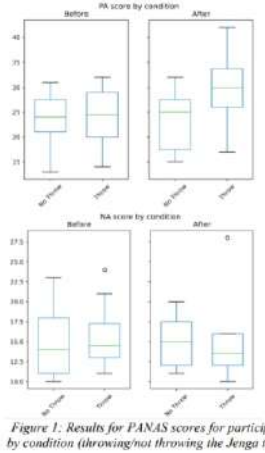


Figure 1: Results for PANAS scores for participants by condition (throwing/not throwing the Jenga tower)

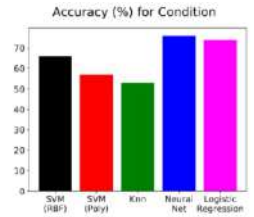


Figure 2: Performance on emotional detection for machine learning models considering the condition

Referencias

Pei, G., & Li, T. (2021). A Literature Review of EEGBased Affective Computing in Marketing. Frontiers in Psychology, 12, 720.

Soleymani, M., Villaro-Dixon, F., Pun, T., & Chanel, G. (2017). Toolbox for emotional feature extraction from physiological signals (TEAP). Frontiers in ICT, 4, 1.

Wedderhoff, N., Gnambs, T., Wedderhoff, O., Burgard, T., & Bošnjak, M. (2021). On the structure of affect: A meta-analytic investigation of the dimensionality and the cross-national applicability of the Positive and Negative Affect Schedule (PANAS). Zeitschrift für Psychologie, 229(1), 24

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

XVI SIMPOSIO INTERNACIONAL DE ENERGÍAS ENERGÍA RENOVABLE

redin

revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

IDENTIFICACIÓN DE LOS CONSUMOS ENERGÉTICOS EN LA INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO.

2021 EXPO TECNOLÓGICA

Participantes

Yolber José Montes Canole1, Alejandro Villa Jaramillo2, Marco Antonio Monsalve Cadavid3, Sergio Danilo Saldarriaga Zuluaga4

1Institución Universitaria Pascual Bravo, Colombia, yolber.montes840@pascualbravo.edu.co
2Institución Universitaria Pascual Bravo, Colombia, alejandro.villa733@pascualbravo.edu.co
3Institución Universitaria Pascual Bravo, Colombia, marco.monsalve@pascualbravo.edu.co
4Institución Universitaria Pascual Bravo, Colombia, s.saldarriagazu@pascualbravo.edu.co

Keywords: Emotion Detection, Human-Robot Interaction, Machine Learning

Resumen

El objetivo del proyecto es hacer un estudio de los consumos en el bloque 6 de la Institución Universitaria Pascual Bravo para determinar qué acciones se pueden implementar con el propósito de hacer un mejor uso de los recursos energéticos y con ello disminuir los consumos eléctricos. Como base del proyecto, está la comparación entre las facturas de energía antes y durante la pandemia que indican que la reducción en consumos de energía es de aproximadamente el 44%, lo que lleva a pensar que hay equipos poco eficientes o cargas que no se gestionan de manera adecuada.

Figura 1: Comparativo consumos 2019 vs 2020

La metodología planteada se fundamenta en la norma ISO 50001 que determina los requerimientos necesarios en un sistema de gestión energética con el objeto de mejorar el rendimiento energético de una instalación [ISO 50001, 2018].

Figura 2: flujograma de aplicación de la metodología

En la aplicación de la metodología, se comenzó por revisar la iluminación del bloque 6 debido a que es de los principales consumos energéticos que se realizan dentro de una instalación de tipo educativo, el objetivo era verificar el estado y la tecnología de las luminarias actuales para entrar a proponer un cambio que permita generar ahorros.

Resultados

Del recorrido inicial realizado en el bloque 6, se evidenció que hay una referencia de luminaria de tecnología fluorescente que se usa en todas las aulas y oficinas y que representan una gran oportunidad de mejora en términos de ahorro de energía.

Tabla 1: Características de la iluminación en el bloque 6

Bloque de estudio	6
Número de luminarias	300
Consumo actual [kWh/mes]	4243
Consumo retrofit [kWh/mes]	2496
Precio [$/kWh]	460,44
Ahorro mensual	$804.481
Ahorro anual	$9'653.769

Referencias

Institución Universitaria Pascual Bravo. Plan de desarrollo 2019-2022, (2018).

ICONTEC, ISO 50001, 2018

Ministerio de Minas y Energía -MME- & Unidad de Planeación Minero-Energética -UPME-. Balance Energético Colombiano, 2016.

Ministerio de Minas y Energía -MME- & Unidad de Planeación Minero-Energética -UPME-. PLAN DE ACCIÓN INDICATIVO DE EFICIENCIA ENERGÉTICA –PAI PROURE 2017- 2022. (2016).

POSTER

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 7 13
2 8 14
3 9 15
4 10 16
5 11 17
6 12 18
19

ORGANIZING COMMITTEE

Apoyan:

XV SIMPOSIO INTERNACIONAL DE ENERGÍAS
ENERGÍA RENOVABLE

redin

revista facultad de ingeniería
Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

ANÁLISIS TEÓRICO – EXPERIMENTAL DEL DESEMPEÑO TÉRMICO DE UN HELIÓSTATO DE BAJA POTENCIA

2021 EXPO TECNOLÓGICA

Participantes

Álvaro Delgado Mejía (1), Laura Catalina Gaviria Sierra (2), Sebastián Zapata Molina (3), Victor Cifuentes Durango (4)

1Docente Facultad de Ingeniería . 2Estudiante de Ingeniería Mecánica. 3Estudiante de Ingeniería Mecánica, 4Estudiante Tecnología en Mecatrónica.

Keywords: Eficiencia Energética

Resumen

Se ha diseñado y construido un sistema de concentración solar tipo helióstato, al cual se le hicieron pruebas experimentales para determinar su eficiencia, a condiciones de Medellín.

METODOLOGÍA

Se hicieron mediciones de radiación solar entrante, velocidad del viento y temperaturas superficiales para calcular las pérdidas de calor y determinar la eficiencia térmica del helióstato.

Figura 1: Prototipo del helióstato durante las pruebas experimentales.

Ecuaciones

La radiación útil se calcula restándole a la radiación global las pérdidas de calor por convección y por radiación (Ec 1) y con ellas se calcula la eficiencia térmica (Ec 2)

$$Rad_{útil} = I_{global} - h\,A_s\,(T_s - T_{aire}) - \epsilon\,\sigma\,A_s\,T_s^4 \quad \text{Ec. 1}$$

$$Eficiencia = \frac{Rad_{útil}}{I_{global}} * 100 \quad \text{Ec. 2}$$

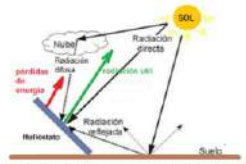


Fig 2. Balance de energía para cálculo de eficiencia

Resultados

Tabla 1. Resultados obtenidos en un intervalo de tiempo de 10:30 AM a 1:30 PM

Hora	I global	Q convección	Q radiación	Rad. Util	Eficiencia
	W	W	W	W	%
10:30 a. m.	207	13.8	17.4	175.8	84.9
11:00 a. m.	207	43.7	20.8	142.5	68.9
11:30 a. m.	207	57.5	22.1	127.4	61.5
12:00 p. m.	207	31.51	19.5	156.0	75.4
01:00 p. m.	207	22.08	18.4	166.5	80.4
01:30 p. m.	207	38.64	20.3	148.1	71.5

Referencias

- D. A. Baharoon, et al, "Historical development of concentrating solar power technologies to generate clean electricity efficiently – A review," Renew. Sustain. Energy Rev., vol. 41, pp. 996–1027, 2015.
- P. Schöttl et al., "Solar field heliostat selection based on polygon optimization and boundaries," AIP Conf. Proc., vol. 2126, no. July, 2019, doi: 10.1063/1.5117565.
- P. Schöttl et al., "Solar field heliostat selection based on polygon optimization and boundaries," AIP Conf. Proc., vol. 2126, no. July, 2019, doi: 10.1063/1.5117565.
- P. Schöttl et al., "Solar field heliostat selection based on polygon optimization and boundaries," AIP Conf. Proc., vol. 2126, no. July, 2019,

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

XV SIMPOSIO INTERNACIONAL DE ENERGÍAS RENOVABLES

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

MEMBRANA DE SEPARACIÓN CH4/CO2 PARA EL INCREMENTO DE EFICIENCIA ENERGÉTICA EN LA FOTOSÍNTESIS ARTIFICIAL

2021 EXPO TECNOLÓGICA

Participantes

Miriam Béjar Sánchez1
, A. Aguilar-Elguézabal
1
, Ricardo Beltran-Chacon1
1Centro de Investigación en Materiales Avanzados (CIMAV), Miguel de Cervantes 120, Complejo Industrial Chihuahua, Chihuahua, México.

Keywords: Eficiencia Energética

Resumen

La fotosíntesis artificial es una herramienta para la disminución de CO2 en el ambiente y una forma de producción de combustibles tales como el Ch4 a partir de energía solar. Sin embargo, la posible oxidación de Ch4 a Co2 al continuar en contacto con el fotocatalizador limita el desempeño en este proceso. Como posible solución, se pretende sintetizar una membrana de óxido de grafeno (GO), la cual, será permeable para el CH4 evitando su oxidación, mientras el Co2 tendrá una menor permeancia y será retenido. Los grupos funcionales oxigenados producen fuerzas electrostáticas y de Van der Walls con el Co2, lo que produce una adsorción en el GO, mientras tanto, el Ch4 al ser una molecular apolar no interactúa con el GO y logra pasar a través de la membrana. Por último, se analiza si se favorece la integración de la membrana en la fotosíntesis artificial obteniendo la producción de Ch4 y la eficiencia energética, comparando los resultados de un sistema con membrana y sin membrana de GO

Métodos

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

Resultados

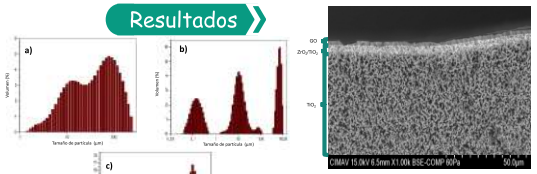


Figura 1. Distribución de tamaño de partícula a) GOP1000, b) GOP500 y c) GOP400.

Figura 2. Estructura y composición de las membranas de GO.

Tabla 1. Composición de C y espesor de las capas superficiales de GO.

	GOP1000	GOP500	GOP400
Espesor (µm)	3.0	2.0	1.7
Permeancia CH_4 (mol/Pa m² s)	1.2×10^{-6}	3.7×10^{-7}	4.0×10^{-7}
Permeancia CO_2 (mol/Pa m² s)	8.5×10^{-7}	2.5×10^{-7}	2.6×10^{-7}
Selectividad CH_4/CO_2	1.4	1.5	1.5

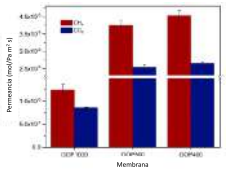


Figura 3. Permeancia para el CH_4 y CO_2 en las tres diferentes membranas sintetizadas de GO.

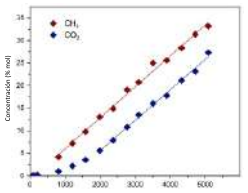


Figura 4. Relación de la concentración de CH_4 y CO_2 con respecto al tiempo en el permeado.

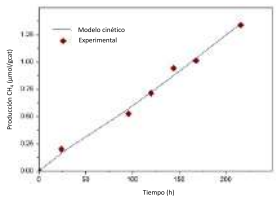


Figura 5. Comparación del modelo cinético con los datos experimentales en la producción de CH_4.

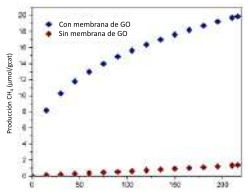


Figura 6. Producción teórica de CH_4 con respecto al tiempo integrando membrana de GO y sin membrana de GO.

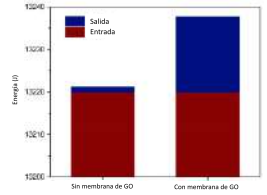


Figura 7. Energía de entrada y de salida en el proceso fotocatalítico de producción de CH_4 con y sin membrana de GO en el sistema.

Discusión

- En este trabajo se sientan las bases para desarrollar un sistema de fotosíntesis artificial donde se obtiene mayor producción de CH_4 a partir de la integración de una membrana.
- Las tres membranas a base de GO sintetizadas tuvieron una permeancia mayor para el CH_4 y menor para el CO_2.
- Al incluir la membrana de GO a la producción fotocatalítica de CH4 se logra mayor producción de CH4 y por lo tanto, la eficiencia energética.
- El máximo factor de separación obtenido con los métodos desarrollados fue de 1.5, a pesar de no ser un valor alto de separación que se encuentra dentro del rango obtenido por otras referencias.

Bibliografía

Chi, C., et al., 28(9): p. 2921-2927.
Koci, K., et al., 31: p. 395-407.
Boffa, V., 2008: Enschede: University of Twente.

POSTER

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

redin revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

Alcaldía de Medellín

"MODELO PARA LA EVALUACIÓN DE LA GESTIÓN DE CIENCIA, TECNOLOGÍA E INNOVACIÓN EN ORGANIZACIONES"

2021 EXPO TECNOLÓGICA

Participantes

Juan Camilo Zapata Giraldo1, Erika Juliett Restrepo Zapata 2, John Jairo Castro Maldonado3

1Investigador SENNOVA, Colombia, jczapatag@sena.edu.co
2 Investigador SENNOVA, Colombia, erestrepoz@sena.edu.co
3Investigador SENNOVA, Colombia, jcastrom@sena.edu.co

Keywords: Desarrollo sostenible

Resumen

La inversión en actividades de ciencia, tecnología e innovación – ACTI cada día toma más relevancia para el fortalecimiento de la productividad y competitividad de las organizaciones y las regiones [Robledo, 2019]. Este trabajo surge a partir de la necesidad de definir un estándar para la medición de las actividades de CTeI en organizaciones, sumado a la carencia de modelos y pronósticos que permitan medir el impacto de dichas actividades y la ausencia de elementos e información que permitan tomar decisiones antes de iniciar su proceso de desarrollo. Por esta razón la falta de una adecuada medición conlleva a despilfarrar los recursos que se podrían destinar a otros proyectos que generen mayor impacto [Adams, 2006; Yam, 2004]. Es por esto que se hace necesario realizar de forma sistémica una medición, seguimiento y evaluación a la Gestión en Ciencia, Tecnología e Innovación - CTeI de tal manera que las empresas puedan definir su impacto dentro del ecosistema de CTeI de la región [Molina-Molina, 2020].

Para lograr el objetivo planteado se realizó la adquisición de información a través de diferentes herramientas como: entrevistas a expertos y revisión de bases de datos, para posteriormente definir un modelo conceptual como herramienta de apoyo en la Gestión de CTeI a partir de la definición de las distintas capacidades organizacionales, para esto se compararon diferentes modelos implementados en la industria. Finalmente, a partir de los resultados obtenidos, se proponen una serie de indicadores, que permiten el seguimiento y el análisis de tendencias de las actividades teniendo en cuenta los insumos, recursos, capacidades, procesos, actividades y productos.

Introducción

En la actualidad, las organizaciones necesitan medir y conocer el estado actual de las distintas actividades de CTeI. Sin embargo, la medición del impacto es un desafío, puesto que es intangible por naturaleza [Eisenhardt, 2000; Romijn, 2000].

De ahí que, a nivel internacional la Organización para la Cooperación y el Desarrollo Económico (OCDE) ha elaborado diferentes manuales con lineamientos que pueden ser utilizados para la definición de indicadores que permitan la medición de las actividades de CTeI [Restrepo, 2016].

De acuerdo a la OCDE las ACTI (ver Figura 1) se pueden definir como aquellas que se relacionan con producir, promover, difundir y aplicar conocimientos científicos en los campos que involucran la CTeI [Fuentes, 2016].

Figura 1: Clasificación Actividades de Ciencia y Tecnología OCDE. Fuente: Adaptado de [OCDE, 2016; Robledo, 2019]

Método

Este proyecto envuelve una investigación con enfoque mixto, que incluye revisión bibliográfica, en donde se analizaron diferentes trabajos con el objetivo de identificar modelos y metodologías para la evaluación de indicadores de CTeI, además, se llevaron a cabo entrevistas a expertos y el desarrollo de un prototipo,. A continuación (**ver Figura 2**) se presenta las distintas fases del Proceso metodológico:

Figura 2: Proceso metodológico
Fuente: Elaboración propia

Resultados

El modelo conceptual para evaluar (Ver Figura 3) la Gestión de Ciencia, tecnología e innovación se realizó a través de indicadores de insumos y productos [Khatibi, 2017; Saunila, 2012; Yam, 2004]. Siendo los insumos de CTeI todos aquellos recursos que respaldan las actividades de CTeI. Por otro lado, los indicadores de salida CTeI, son los resultados producto de las ACTI [Adams, 2006; Nerkar, 2005; Saunila, 2012]

Figura 3: Modelo insumo-producto para la definición de indicadores
Fuente: Adaptado de (O'Onofrio & Gellman, 2009)

De acuerdo con [Saunila, 2012], además, de las entradas y salidas de las ACTI, es importante evaluar las capacidades y los procesos relacionados de forma directa e indirecta con las ACTI, las cuales difieren entre organizaciones, de ahí que, para la definición de las métricas, es necesario una comprensión de los procesos que intervienen en el desarrollo de dichas actividades, siendo las medidas más adecuadas, aquellas que permitan focalizar la innovación [Birchall, 2011]. Así pues, la medición se considera un proceso integral o sistémico, de manera que todas las cosas que suceden en la organización tienen un impacto en el desempeño [Adams, 2006; OCDE, 2016; Robledo, 2019].

Dimensión de medida	Indicador propuesto
Entrada/Insumos	Gasto en ACTI
	Intensidad de I+D
	El número de personal de ACTI
	Intensidad del personal de ACTI
	Intensidad de introducción de tecnología
	Tasa de crecimiento anual promedio de ACTI
Salida/Productos	El número de patentes
	Ingresos por ventas de nuevos productos
	Ingresos por transferencia de tecnología
	El número de nuevos productos
	Ciclo de nuevos productos

Tabla 1: Propuesta de indicadores para la medición de ACTI. Fuente: Elaboración propia

Modelo de analítica de datos

Ahora bien quienes toman de decisiones requieren datos precisos y oportunos para comprender mejor las oportunidades y los problemas de la organización, por esta razón, ha tomado relevancia el uso de sistemas información como soporte a decisiones, lo que permite integrar y utilizar los datos para mejorar la calidad de las métricas y la evaluación de la gestión [Khatibi, 2017]. Teniendo en cuenta lo anterior, se propuso una serie de indicadores para la medición de ACTI (ver tabla 1), a partir de estos indicadores se propone un modelo preliminar usando herramientas de analítica de datos.

Discusión

En línea con lo anterior, el modelo conceptual para medición y análisis del impacto de las actividades de ciencia tecnología e innovación y sus efectos se basa en una revisión de la literatura y las consideraciones de los diferentes investigadores. El modelo de analítica de datos se encuentra en una etapa de validación en una organización de Educación Superior. Para futuras investigaciones, es necesario llevar a cabo más estudios de casos en diferentes organizaciones para evaluar la idoneidad del modelo. Por ejemplo, se puede probar este mismo modelo para medir su desempeño innovador, con empresas de diferentes sectores y diferentes tamaños. El contraste de resultados se puede probar con evidencia empírica y permitirá determinar en qué categoría se encuentra una empresa específica en esa escala.

Referencias

Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: A review. International Journal of Management Reviews, 8(1), 21–47. https://doi.org/10.1111/j.1468-2370.2006.00118.x

Birchall, D., Chanaron, J.-J., Tovstiga, G., & Hillenkamp, J. (2011). Innovation performance measurement: current practices, issues and management challenges. International Journal of Technology Management, 54(1), 1–20.

D'Onofrio, M. G., & Gellman, I. (2009). Fuentes de información para el análisis de resultados e impacto de programas de becas de posgrado en ciencias e ingeniería en Iberoamérica. Revista Iberoamericana de Ciencia, Tecnología y Sociedad-CTS, 5(13), 149–159.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? Strategic Management Journal, 21(10–11), 1105–1121. https://doi.org/10.1002/1097-0266

Khatibi, V., Hatamabadi, A., & Montazeri, G. A. (2017). A Business Intelligence Approach to Monitoring and Trend Analysis of National R&D Indicators. Knowledge Management Journal, 4(14), 144–157. https://doi.org/10.32065/JKMR.04.2017.1.1.0.0

Molina-Molina, S., Alvarez-Lopez, G., Estrada-Hernandez, L., & Estrada-Hernandez, N. (2020). Indicadores de ciencia, tecnología e innovación: hacia la configuración de un sistema de medición. Revista Interamericana de Bibliotecología, 43(3), e20. https://doi.org/10.17533/udea.rib.v43n3e08

Nerkar, A., & Paruchuri, S. (2005). Evolution of R&D capabilities: The role of knowledge networks within a firm. Management Science, 51(5), 771–785. https://doi.org/10.1287/mnsc.1040.0364

OCDE. (2016). OECD Science, Technology and Innovation Outlook 2016. https://doi.org/10.1787/sti_in_outlook-2016-en

Parmar, I. (2020). Business Intelligence: Concepts, components, techniques and benefits. Journal of Theoretical and Applied Information Technology, 98(1), 85–99.

Restrepo Cifuni, G. E., & Zapata Izaguirre, O. E. (2016). Indicadores de gestión para procesos de investigación y extensión en instituciones de educación superior. Revista Ciencias Estratégicas, 24(35), 455–482. https://repositorio.upb.edu.co/bitstream/handle/20.500.11912/3009/Indicadores%20de%20gestion%20para%20procesos.pdf?sequence=1&isAllowed=y

Robledo Velásquez, J. (2019). Gestión estratégica de la innovación. Introducción a la Gestión de la Tecnología y la Innovación (empresarial). ESDS, 23–28. http://www.eumed.net/libros-gratis/2019/jrv/index.htm

Romijn, H., & Albaladejo, M. (2002). Determinants of innovation capability in small UK firms: An empirical analysis. Chalmers Centre for Innovation Studies Working Paper.

Saunila, M., & Ukko, J. (2012). A conceptual framework for the measurement of innovation capability and its effects. Baltic Journal of Management, 7(4), 355–375. https://doi.org/10.1108/17465261211272130

Yam, R. C. M., Guan, J. C., Pun, K. F., & Tang, E. P. Y. (2004). An audit of technological innovation capabilities in Chinese firms: some empirical findings in Beijing, China. Research Policy, 33(8), 1123–1140. https://doi.org/10.1016/j.respol.2004.03.004

NOTA EDITORIAL: Las opiniones y los artículos publicados en el libro son de responsabilidad exclusiva de los autores; así mismo: estos se responsabilizaran de obtener el permiso correspondiente para poner material publicado en otro lugar.

POSTER

Memorias
7 16 18 19
POSTER

INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO

SENA | SENNOVA
Sistema de Investigación, Desarrollo Tecnológico e Innovación

MODELADO DE LA ADMISIÓN DE LA CARGA EN UNA METODOLOGÍA SECUENCIAL PARA COMBUSTIÓN HCCI

Sebastián H. Quintana (1), Andrés Morales (2), Iván Bedoya (3)

1. Consultor en eficiencia energética, Energética SAS, Colombia, sebastian@energeticasas.com; 2. Docente ocasional, grupo GIIEN, Institución Universitaria Pascual Bravo, Colombia, andres.morales@pascualbravo.edu.co; 3. Profesor asociado, grupo GASURE, Universidad de Antioquia, Colombia, iban.bedoya@udea.edu.co

Palabras clave: motores de combustión interna, HCCI, CFD, simulación numérica, cinética química.

Introducción

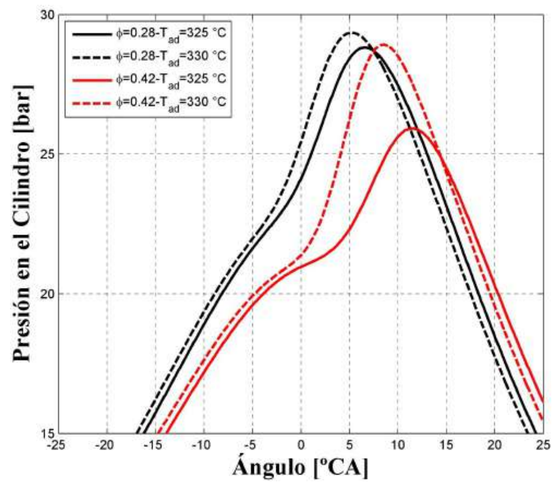
La combustión de carga homogénea de encendido por compresión (HCCI) es una opción atractiva para solucionar los problemas asociados al proceso de ignición y desarrollo de la combustión en las tecnologías de motores de combustión interna convencionales frente a combustibles con composiciones variables y la generación de contaminantes atmosféricos. La combustión HCCI presenta diferentes desafíos técnicos que han limitado su uso amplio. Los modelos que acoplan el análisis CFD y la cinética química mediante modelos multizona han permitido profundizar en los fenómenos que permiten la ocurrencia de la ignición, el desarrollo de la combustión y las condiciones que favorecen la generación de ciertos contaminantes, siendo una alternativa costo-efectiva para abordar los desafíos técnicos. En la presente ponencia se exponen los resultados de la evaluación del efecto de la inclusión del período de cruce válvulas y de la carrera de admisión sobre la distribución de la masa y la temperatura durante la carrera de compresión, y como esto afecta las predicciones del desarrollo de la combustión y la generación de contaminantes atmosféricos.

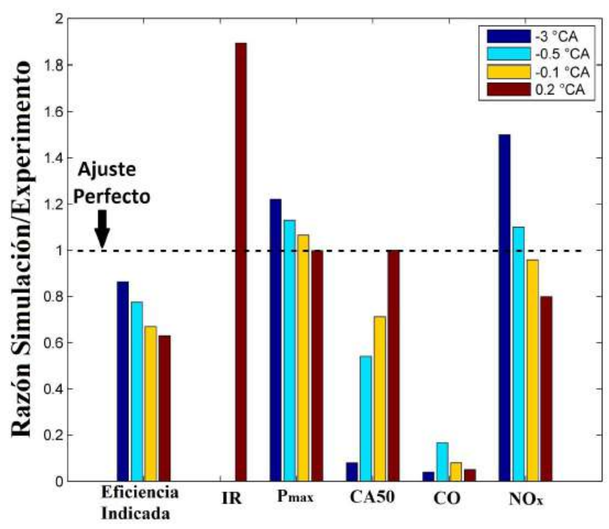
Métodos

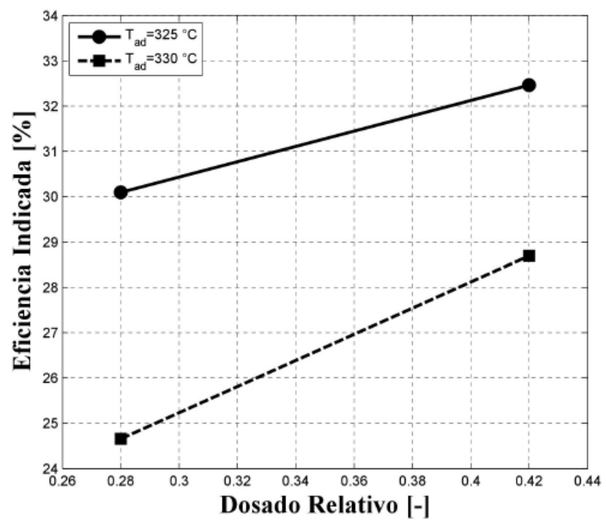
Experimentalmente, se evaluaron dos condiciones de dosado relativo y temperatura de la carga en la admisión de un motor Diesel modificado de generación, de aspiración natural y operado a 1800 rpm, operando en modo HCCI utilizando biogás (60% gas natural – 40% dióxido de carbono) como combustible. Los experimentos se llevaron a cabo en el banco de motores de la Universidad de Antioquia (Medellín). Estas mismas condiciones se utilizaron para realizar una simulación secuencial CFD-cinética química detallada en el paquete ANSYS para modelar el cruce de válvulas y la carrera de admisión y su efecto sobre la estratificación térmica y másica de la carga durante la carrera de compresión, usando una metodología de reducción de zonas para el análisis de la combustión y la carrera de expansión. El nivel de ajuste entre los experimentos y la simulación se realizó comparando parámetros indicados, presión al interior del cilindro y concentración de contaminantes.

Resultados

Las condiciones de operación (régimen de giro y presión atmosférica) condujeron a altos niveles de temperatura en la admisión (baja densidad de carga) y una combustión retrasada (Figura 1), obteniéndose eficiencias menores a las esperadas con este modo de combustión (Figura 2). El modelado de la carrera compresión presenta un buen ajuste con los resultados experimentales, sin embargo, condición de combustión retrasada, si bien permite un adecuado ajuste para diversos parámetros indicados, conduce a una deficiencia en la estimación de la generación de monóxido de carbono e hidrocarburos sin quemar (Figura 3).







embargo, esto no afecta el modelado de la carrera de compresión, en la cual, al incluir el cruce de válvula y la carrera de admisión, permite obtener los patrones de flujo y caracterizar la estratificación de la carga, con el fin de entender las condiciones que conducen a una operación ineficiente y que promueven el incremento/reducción de contaminantes.

Referencias

John E. Dec. Advanced Compression-Ignition Engines - Understanding the- In Cylinder Processes. Proceedings of the Combustion Institute, 32 II(2):2727-2742,2009.

Samveg Saxena and Iván Dario Bedoya. Fundamental Phenomena Affecting Low Temperature Combustion and HCCI engines, High Load Limits and Strategies for Extending These Limits.Progress in Energy and Combustion Science, 39(5):457-488,2012.

INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO
Memorias
7 16 18 19
POSTER

IDENTIFICACIÓN DE LOS CONSUMOS ENERGÉTICOS EN LA INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO.

Yolber José Montes Canole[1], Alejandro Villa Jaramillo[2], Marco Antonio Monsalve Cadavid[3], Sergio Danilo Saldarriaga Zuluaga[4]

[1]Institución Universitaria Pascual Bravo, Colombia, yolber.montes840@pascualbravo.edu.co
[2]Institución Universitaria Pascual Bravo, Colombia, alejandro.villa733@pascualbravo.edu.co
[3]Institución Universitaria Pascual Bravo, Colombia, marco.monsalve@pascualbravo.edu.co
[4]Institución Universitaria Pascual Bravo, Colombia, s.saldarriagazu@pascualbravo.edu.co

Eje temático:

Resumen y metodología

El objetivo del proyecto es hacer un estudio de los consumos en el bloque 6 de la Institución Universitaria Pascual Bravo para determinar qué acciones se pueden implementar con el propósito de hacer un mejor uso de los recursos energéticos y con ello disminuir los consumos eléctricos.

Como base del proyecto, está la comparación entre las facturas de energía antes y durante la pandemia que indican que la reducción en consumos de energía es de aproximadamente el 44%, lo que lleva a pensar que hay equipos poco eficientes o cargas que no se gestionan de manera adecuada.

Figura 1: Comparativo consumos 2019 vs 2020

La metodología planteada se fundamenta en la norma ISO 50001 que determina los requerimientos necesarios en un sistema de gestión energética con el objeto de mejorar el rendimiento energético de una instalación [ISO 50001, 2018].

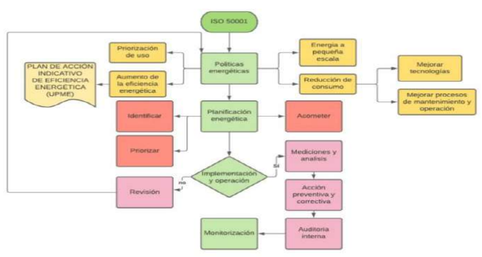


Figura 2: flujograma de aplicación de la metodología

En la aplicación de la metodología, se comenzó por revisar la iluminación del bloque 6 debido a que es de los principales consumos energéticos que se realizan dentro de una instalación de tipo educativo, el objetivo era verificar el estado y la tecnología de las luminarias actuales para entrar a proponer un cambio que permita generar ahorros.

Resultados

Del recorrido inicial realizado en el bloque 6, se evidenció que hay una referencia de luminaria de tecnología fluorescente que se usa en todas las aulas y oficinas y que representan una gran oportunidad de mejora en términos de ahorro de energía.

Tabla 1: Características de la iluminación en el bloque 6

Bloque de estudio	6
Número de luminarias	300
Consumo actual [kWh/mes]	4243
Consumo retrofit [kWh/mes]	2496
Precio [$/kWh]	460,44
Ahorro mensual	$804.481
Ahorro anual	$9’653.769

Referencias

Institución Universitaria Pascual Bravo. Plan de desarrollo 2019-2022, (2018).

ICONTEC, ISO 50001, 2018

Ministerio de Minas y Energía -MME- & Unidad de Planeación Minero-Energética -UPME-. Balance Energético Colombiano, 2016.

Ministerio de Minas y Energía -MME- & Unidad de Planeación Minero-Energética -UPME-. PLAN DE ACCIÓN INDICATIVO DE EFICIENCIA ENERGÉTICA –PAI PROURE 2017- 2022. (2016).

VIGILADA Mineducación
Apoyan:
ISSNe: 2422-2208 | ISSN: 0122-350X REVISTA cintex
redin revista facultad de ingeniería Universidad de Antioquia
2021 EXPO TECNOLÓGICA
Hacia la Acreditación Institucional
Alcaldía de Medellín

INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO
Memorias
7 16 18 19
POSTER

MEMBRANA DE SEPARACIÓN CH4/CO2 PARA EL INCREMENTO DE EFICIENCIA ENERGÉTICA EN LA FOTOSÍNTESIS ARTIFICIAL

Miriam Béjar Sánchez[1], A. Aguilar-Elguézabal[1], Ricardo Beltran-Chacon[1]

[1]Centro de Investigación en Materiales Avanzados (CIMAV), Miguel de Cervantes 120, Complejo Industrial Chihuahua, Chihuahua, México.

Introducción

La fotosíntesis artificial es una herramienta para la disminución de CO_2 en el ambiente y una forma de producción de combustibles tales como el CH_4 a partir de energía solar. Sin embargo, la posible oxidación de CH_4 a CO_2 al continuar en contacto con el fotocatalizador limita el desempeño en este proceso. Como posible solución, se pretende sintetizar una membrana de óxido de grafeno (GO), la cual, será permeable para el CH_4 evitando su oxidación, mientras el CO_2 tendrá una menor permeancia y será retenido. Los grupos funcionales oxigenados producen fuerzas electrostáticas y de Van der Walls con el CO_2, lo que produce una adsorción en el GO, mientras tanto, el CH_4 al ser una molecular apolar no interactúa con el GO y logra pasar a través de la membrana. Por último, se analiza si se favorece la integración de la membrana en la fotosíntesis artificial obteniendo la producción de CH_4 y la eficiencia energética, comparando los resultados de un sistema con membrana y sin membrana de GO.

Métodos

Purificación de GO	Obtención de dispersiones de 400 (GOP400), 500 (GOP500) y 1000 (GOP1000) mg/l de concentración inicial con láminas de mayor tamaño por ajuste de pH y decantación
Preparación de membranas GO	Depósito de capa superficial de GO con las dispersiones purificadas GOP400, GOP500 y GOP1000 sobre sustratos de TiO_2/ZrO_2 por spin coating
Pruebas desempeño de separación	Pruebas de permeación y selectividad para cada membrana con respecto al CO_2 y CH_4
Producción de CH_4	Ajuste de modelo cinético Langmuir-Hinshelwood en la producción experimental de CH_4 sin membrana de GO y cálculo de producción de CH_4 con membrana de GO
Eficiencia energética	Cálculo de eficiencia energética a partir de la relación de la energía que proporciona el combustible y la energía necesaria para producir el combustible

Resultados

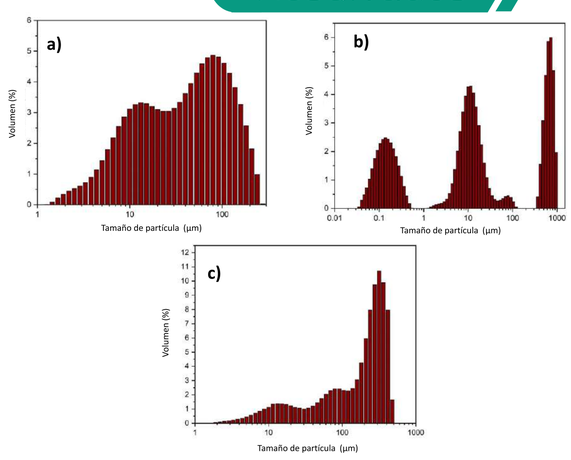


Figura 1. Distribución de tamaño de partícula a) GOP1000, b) GOP500 y c) GOP400.

Tabla 1. Composición de C y espesor de las capas superficiales de GO.

	GOP1000	GOP500	GOP400
Espesor (µm)	3.0	2.0	1.7
Permeancia CH_4 (mol/Pa m² s)	1.2×10^{-6}	3.7×10^{-5}	4.0×10^{-5}
Permeancia CO_2 (mol/Pa m² s)	8.5×10^{-7}	2.5×10^{-5}	2.6×10^{-5}
Selectividad CH_4/CO_2	1.4	1.5	1.5

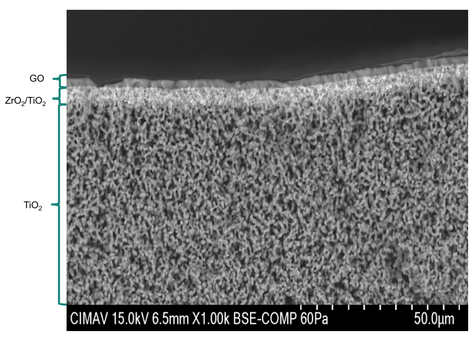


Figura 2. Estructura y composición de las membranas de GO.

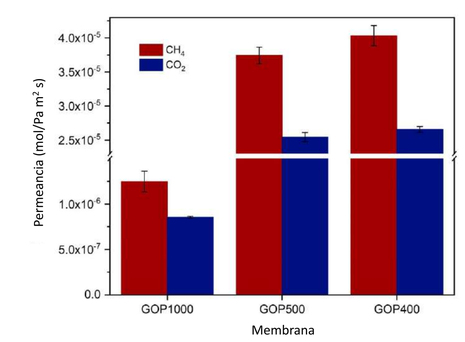


Figura 3. Permeancia para el CH_4 y CO_2 en las tres diferentes membranas sintetizadas de GO.

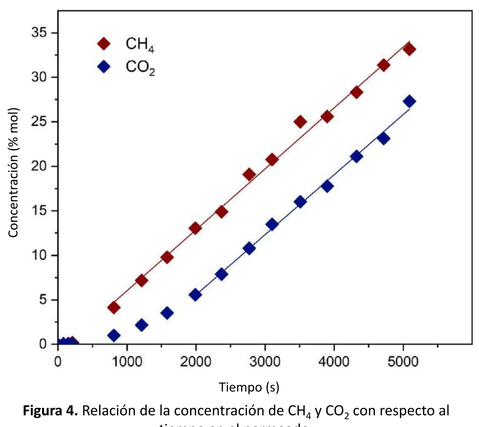


Figura 4. Relación de la concentración de CH_4 y CO_2 con respecto al tiempo en el permeado.

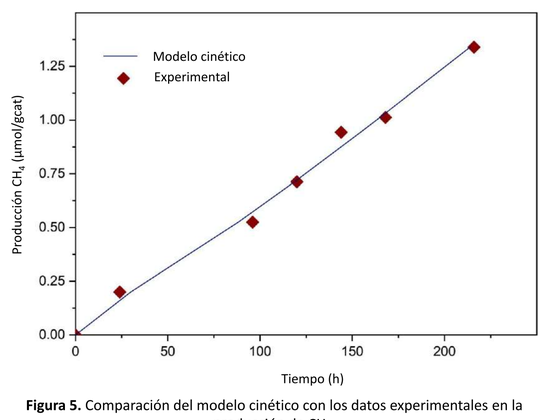


Figura 5. Comparación del modelo cinético con los datos experimentales en la producción de CH_4.

Discusión

- En este trabajo se sientan las bases para desarrollar un sistema de fotosíntesis artificial donde se obtiene mayor producción de CH_4 a partir de la integración de una membrana.
- Las tres membranas a base de GO sintetizadas tuvieron una permeancia mayor para el CH_4 y menor para el CO_2.
- Al incluir la membrana de GO a la producción fotocatalítica de CH4 se logra mayor producción de CH4 y por lo tanto, la eficiencia energética.
- El máximo factor de separación obtenido con los métodos desarrollados fue de 1.5, a pesar de no ser un valor alto de separación que se encuentra dentro del rango obtenido por otras referencias.

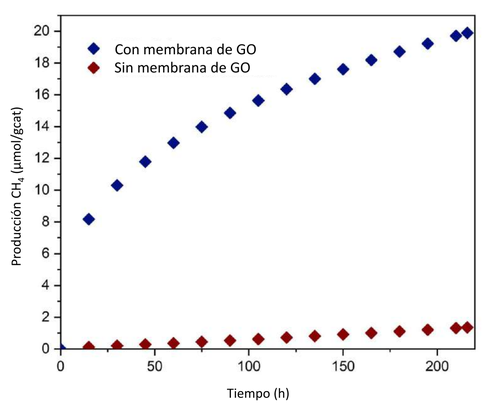


Figura 6. Producción teórica de CH_4 con respecto al tiempo integrando membrana de GO y sin membrana de GO.

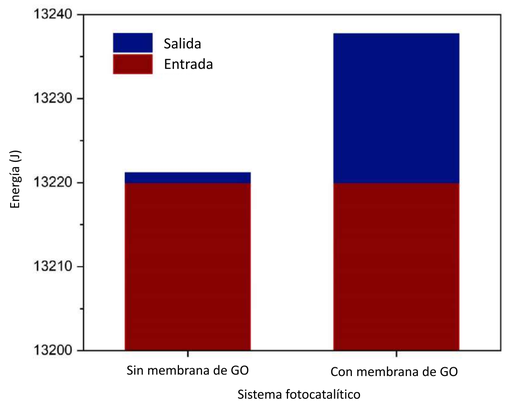


Figura 7. Energía de entrada y de salida en el proceso fotocatalítico de producción de CH_4 con y sin membrana de GO en el sistema.

Bibliografía

Chi, C., et al., 28(9): p. 2921-2927.
Koci, K., et al., 31: p. 395-407.
Boffa, V., 2008: Enschede: University of Twente.

Apoyan:
ISSNe: 2422-2208 | ISSN: 0122-350X REVISTA cintex
redin revista facultad de ingeniería Universidad de Antioquia
EXPO TECNOLÓGICA 2021
Hacia la Acreditación Institucional
Alcaldía de Medellín
VIGILADA Mineducación

7 16 18 19

POSTER

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO®

SENA | SENNOVA
Sistema de Investigación, Desarrollo Tecnológico e Innovación

"MODELO PARA LA EVALUACIÓN DE LA GESTIÓN DE CIENCIA, TECNOLOGÍA E INNOVACIÓN EN ORGANIZACIONES"

Juan Camilo Zapata Giraldo[1], Erika Juliett Restrepo Zapata [2], John Jairo Castro Maldonado[3]

[1]Investigador SENNOVA, Colombia, jczapatag@sena.edu.co
[2] Investigador SENNOVA, Colombia, erestrepoz@sena.edu.co
[3]Investigador SENNOVA, Colombia, jcastrom@sena.edu.co

Eje temático: Desarrollo sostenible

Resumen

La inversión en actividades de ciencia, tecnología e innovación – ACTI cada día toma más relevancia para el fortalecimiento de la productividad y competitividad de las organizaciones y las regiones [Robledo, 2019]. Este trabajo surge a partir de la necesidad de definir un estándar para la medición de las actividades de CTeI en organizaciones, sumado a la carencia de modelos y pronósticos que permitan medir el impacto de dichas actividades y la ausencia de elementos e información que permitan tomar decisiones antes de iniciar su proceso de desarrollo. Por esta razón la falta de una adecuada medición conlleva a despilfarrar los recursos que se podrían destinar a otros proyectos que generen mayor impacto [Adams, 2006; Yam, 2004]. Es por esto que se hace necesario realizar de forma sistémica una medición, seguimiento y evaluación a la Gestión en Ciencia, Tecnología e Innovación - CTeI de tal manera que las empresas puedan definir su impacto dentro del ecosistema de CTeI de la región [Molina-Molina, 2020].
Para lograr el objetivo planteado se realizó la adquisición de información a través de diferentes herramientas como: entrevistas a expertos y revisión de bases de datos, para posteriormente definir un modelo conceptual como herramienta de apoyo en la Gestión de CTeI a partir de la definición de las distintas capacidades organizacionales, para esto se compararon diferentes modelos implementados en la industria. Finalmente, a partir de los resultados obtenidos, se proponen una serie de indicadores, que permiten el seguimiento y el análisis de tendencias de las actividades teniendo en cuenta los insumos, recursos, capacidades, procesos, actividades y productos.

Introducción

En la actualidad, las organizaciones necesitan medir y conocer el estado actual de las distintas actividades de CTeI. Sin embargo, la medición del impacto es un desafío, puesto que es intangible por naturaleza [Eisenhardt, 2000; Romijn, 2000].
De ahí que, a nivel internacional la Organización para la Cooperación y el Desarrollo Económico (OCDE) ha elaborado diferentes manuales con lineamientos que pueden ser utilizados para la definición de indicadores que permitan la medición de las actividades de CTeI [Restrepo, 2016].
De acuerdo a la OCDE las ACTI (ver Figura 1) se pueden definir como aquellas que se relacionan con producir, promover, difundir y aplicar conocimientos científicos en los campos que involucran la CTeI [Fuentes, 2016].

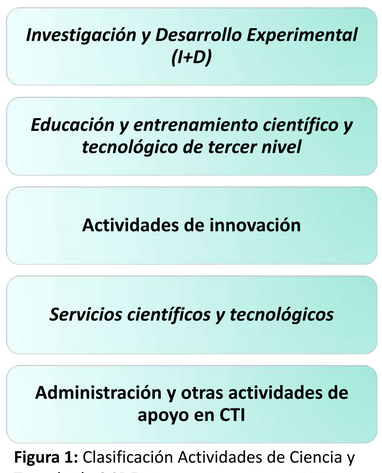


Figura 1: Clasificación Actividades de Ciencia y Tecnología OCDE. Fuente: Adaptado de [OCDE, 2016; Robledo, 2019]

Método

Este proyecto envuelve una investigación con enfoque mixto, que incluye revisión bibliográfica, en donde se analizaron diferentes trabajos con el objetivo de identificar modelos y metodologías para la evaluación de indicadores de CTeI, además, se llevaron a cabo entrevistas a expertos y el desarrolló de un prototipo,. A continuación (**ver Figura 2**) se presenta las distintas fases del Proceso metodológico:

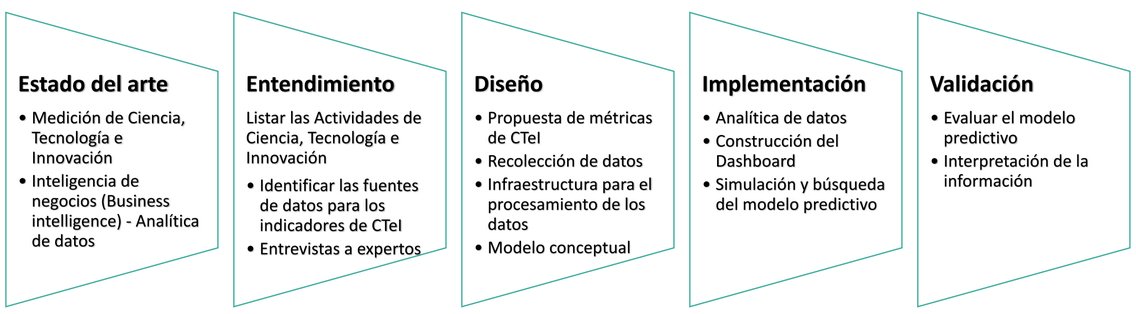


Figura 2: Proceso metodológico
Fuente: Elaboración propia

Resultados

El modelo conceptual para evaluar (Ver Figura 3) la Gestión de Ciencia, tecnología e innovación se realizó a través de indicadores de insumos y productos [Khatibi, 2017; Saunila, 2012; Yam, 2004]. Siendo los insumos de CTeI todos aquellos recursos que respaldan las actividades de CTeI. Por otro lado, los indicadores de salida CTeI, son los resultados producto de las ACTI [Adams, 2006; Nerkar, 2005; Saunila, 2012]

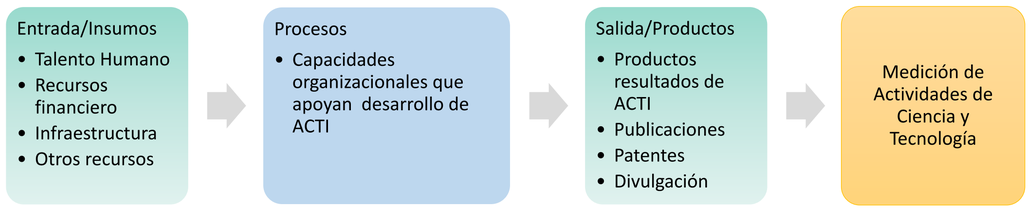


Figura 3: Modelo insumo-producto para la definición de indicadores
Fuente: Adaptado de (D'Onofrio & Gelfman, 2009)

De acuerdo con [Saunila, 2012], además, de las entradas y salidas de las ACTI, es importante evaluar las capacidades y los procesos relacionados de forma directa e indirecta con las ACTI, las cuales difieren entre organizaciones, de ahí que, para la definición de las métricas, es necesario una comprensión de los procesos que intervienen en el desarrollo de dichas actividades, siendo las medidas más adecuadas, aquellas que permitan focalizar la innovación [Birchall, 2011]. Así pues, la medición se considera un proceso integral o sistémico, de manera que todas las cosas que suceden en la organización tienen un impacto en el desempeño [Adams, 2006; OCDE, 2016; Robledo, 2019].

Dimensión de medida	Indicador propuesto
Entrada/Insumos	Gasto en ACTI
	Intensidad de I+D
	El número de personal de ACTI
	Intensidad del personal de ACTI
	Intensidad de introducción de tecnología
	Tasa de crecimiento anual promedio de ACTI
Salida/Productos	El número de patentes
	Ingresos por ventas de nuevos productos
	Ingresos por transferencia de tecnología
	El número de nuevos productos
	Ciclo de nuevos productos

Tabla 1: Propuesta de indicadores para la medición de ACTI. Fuente: Elaboración propia

Modelo de analítica de datos

Ahora bien quienes toman de decisiones requieren datos precisos y oportunos para comprender mejor las oportunidades y los problemas de la organización, por esta razón, ha tomado relevancia el uso de sistemas información como soporte a decisiones, lo que permite integrar y utilizar los datos para mejorar la calidad de las métricas y la evaluación de la gestión [Khatibi, 2017]. Teniendo en cuenta lo anterior, se propuso una serie de indicadores para la medición de ACTI (ver tabla 1), a partir de estos indicadores se propone un modelo preliminar usando herramientas de analítica de datos.

Discusión

En línea con lo anterior, el modelo conceptual para medición y análisis del impacto de las actividades de ciencia tecnología e innovación y sus efectos se basa en una revisión de la literatura y las consideraciones de los diferentes investigadores. El modelo de analítica de datos se encuentra en una etapa de validación en una organización de Educación Superior. Para futuras investigaciones, es necesario llevar a cabo más estudios de casos en diferentes organizaciones para evaluar la idoneidad del modelo. Por ejemplo, se puede probar este mismo modelo para medir su desempeño innovador, con empresas de diferentes sectores y diferentes tamaños. El contraste de resultados se puede probar con evidencia empírica y permitirá determinar en qué categoría se encuentra una empresa específica en esa escala.

Referencias

Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: A review. *International Journal of Management Reviews*, *8*(1), 21–47. https://doi.org/10.1111/J.1468-2370.2006.00119.X

Birchall, D., Chanaron, J.-J., Tovstiga, G., & Hillenbrand, C. (2011). Innovation performance measurement: current practices, issues and management challenges. *International Journal of Technology Management*, *56*(1), 1–20.

D'Onofrio, M. G., & Gelfman, J. (2009). Fuentes de información para el análisis de resultados e impactos de programas de becas de posgrado en ciencias e ingeniería en Iberoamérica. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad-CTS*, *5*(13), 103–130.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic Management Journal Strat. Mgmt. J*, *21*, 1105–1121. https://doi.org/10.1002/1097-0266

Khatibi, V., Keramati, A., & Montazer, G. A. (2017). A Business Intelligence Approach to Monitoring and Trend Analysis of National R&D Indicators. *EMJ - Engineering Management Journal*, *29*(4), 244–257. https://doi.org/10.1080/10429247.2017.1380578

Molina-Molina, S., Álvarez-Argaez, S., Estrada-Hernández, J., & Estrada-Hernández, M. (2020). Indicadores de ciencia, tecnología e innovación: hacia la configuración de un sistema de medición. *Revista Interamericana de Bibliotecología*, *43*(3), eI9. https://doi.org/10.17533/udea.rib.v43n3eI9

Nerkar, A., & Paruchuri, S. (2005). Evolution of R&D capabilities: The role of knowledge networks within a firm. *Management Science*, *51*(5), 771–785. https://doi.org/10.1287/mnsc.1040.0354

OCDE. (2016). *OECD Science, Technology and Innovation Outlook 2016*. https://doi.org/10.1787/sti_in_outlook-2016-en

Ranjan, J. (2009). Business intelligence: Concepts, components, techniques and benefits. *Journal of theoretical and applied information technology*, *9*(1), 60–70.

Restrepo Ortiz, G. E., & Zabala Mendoza, D. E. (2016). Indicadores de gestión para proyectos de investigación y extensión en instituciones de Educación Superior. *Revista Ciencias Estratégicas*, *24*(36), 451–461. https://repository.upb.edu.co/bitstream/handle/20.500.11912/7980/Artículo 12.pdf?sequence=1&isAllowed=y

Robledo Velásquez, J. (2019). Gestión estratégica de la innovación. *Introducción a la Gestión de la Tecnología y la Innovación empresarial*, *53*(0), 21–25. http://www.elsevier.com/locate/scp

Romijn, H., & Albaladejo, M. (2000). Determinants of innovation capability in small UK firms: an empirical analysis. *Eindhoven Centre for Innovation Studies Working Paper*.

Saunila, M., & Ukko, J. (2012). A conceptual framework for the measurement of innovation capability and its effects. *Baltic Journal of Management*, *7*(4), 355–375. https://doi.org/10.1108/17465261211272139

Yam, R. C. M., Guan, J. C., Pun, K. F., & Tang, E. P. Y. (2004). An audit of technological innovation capabilities in Chinese firms: Some empirical findings in Beijing, China. *Research Policy*, *33*(8), 1123–1140. https://doi.org/10.1016/j.respol.2004.05.004

INSTITUCIÓN UNIVERSITARIA
PASCUAL BRAVO

Memorias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

ORGANIZING COMMITTEE

Apoyan:

redin

revista facultad de ingeniería Universidad de Antioquia

Hacia la Acreditación Institucional

Alcaldía de Medellín

ISSNe: 2422-2208 | ISSN: 0122-350X
REVISTA cintex

ORGANIZING COMMITTEE

EXPO TECNOLÓGICA 2021

Participantes

General Chair
PhD. ANDRÉS BENAVIDES ARÉVALO

Program Chair
PhD. ROSALBA RÍOS GALVIS

SYMPOSIUM CHAIRS
M. Sc. JESUS IGNACIO CALLE PEREZ
M. Sc. OSCAR JULIÁN GALEANO ECHEVERRI
M. Sc. CARLOS ALBERTO VALENCIA HERNÁNDEZ

Vice-Chair Poster Session
M. Sc. JHONY MAURICIO GUTIERREZ
M. Sc. OSCAR IGNACIO BOTERO HENAO

Publications Chair
PhD. JULIÁN ALBERTO PATIÑO MURILLO
Revista CINTEX,
FONDO EDITORIAL PASCUAL BRAVO.

compilation: PhD. ROSALBA RÍOS GALVIS

AGRADECIMIENTOS

GRACIAS A TODOS LOS QUE PARTICIPARON EN LA REALIZACIÓN DE ESTE EVENTO, QUE SE HA CONVERTIDO EN TODA UNA PROEZA DE INNOVACIÓN

ISSN 2422-3891

ANEXOS 2. Avances de redacción de artículo científico orientado en la visualización de mapas para actividades de ciencia tecnología e innovación.

Clustering and geospatial visualization of innovation indicators in the SENA Regional Antioquia

List the author names here

Type the author addresses here

Type the corresponding author's e-mail address here

Abstract. The activities of Science, Technology and Innovation are related to the execution of actions that involve research, experimental development, support for education and training and provision of scientific and technological services, innovation activities, scientific and technological services, administration and other management activities, which are carried out by entities public and private entities to contribute to the generation and appropriation of knowledge. In this sense, due to the exponential increase that has had the generation of new knowledge products at the educational level and in order to respond to an evident need to contribute to the development of the productive sector in Colombia, through the SENNOVA Research System of SENA both human resources are potentiated as technological with which the entity has to contribute to the economic and social growth of the region. Taking into account the above, in the Antioquia region of SENA, work is being done on the development of the visualization of a data analytics model for the measurement of innovation indicators, where the different categories of Science, Technology and Innovation activities are graphically identified that have been working in recent years in the entity, with the purpose of facilitating the measurement and analysis of the results obtained from the perspective of both descriptive statistics and data science. Among the indicators used are the management and productivity processes of projects executed in the entity in a period of time, to which techniques such as K-means grouping and Principal Component Analysis (PCA) were applied, which allowed the selected indicators to be grouped. Python Folium was used as a visualization tool that allows to generate comparisons graphically between the different SENA centers of the Antioquia Regional in a given period of time. It was observed that, through the techniques used, there is greater ease of interpretation, comparison and analysis of statistical results and data analytics in the measurement of indicators of Science, Technology and Innovation activities compared to information visualization tools conventionally used.

1. Introduction

Comenzamos definiendo las actividades de Ciencia, Tecnología e innovación (ACTI) como aquellas que se relacionan con producir, promover, difundir y aplicar conocimientos científicos en los campos que involucran la ciencia y la tecnología [1]. En los últimos años ha surgido la necesidad de proporcionar de forma sistemática información sobre la recolección, análisis, publicación y normalización de los datos asociados a estas actividades [2], de allí que se empiecen a diseñar indicadores que aporten a la toma de decisiones estratégicas sobre su gestión.

A nivel internacional la Organización para la Cooperación y el Desarrollo Económico (OCDE) ha elaborado diferentes manuales con lineamientos que pueden ser utilizados para la definición de indicadores que permitan la medición de ACTI [3], como el Manual de Frascati y el Manual de Oslo, que compuestos por una serie de propuestas para la recolección e interpretación de datos sobre innovación tecnológica, por otro lado, el Manual de Bogotá que contiene pautas para construir indicadores de innovación tecnológica en Latinoamérica y el Caribe, entre otros [4].

Esta investigación tiene como objetivo proponer un método de medición de Indicadores de Ciencia, Tecnología e Innovación mediante analítica de datos, la fundamentación teórica de la propuesta se basó en técnicas estadísticas no paramétricas supervisadas, entre las que encontramos métodos como clustering basado en algoritmos k-means y el uso de análisis de componentes principales (PCA). Para ello se simuló una base de datos sintéticos de 500 proyectos distintos, que permitieron caracterizar los 16 centros del SENA de la regional Antioquia, de los cuales posteriormente se crean mapas utilizando como herramienta de base Folium [10].

2. Methodology implemented

El enfoque desarrollado en nuestra investigación es de tipo mixto este implica la recolección y el análisis de datos cuantitativos y cualitativos, permitiendo integrarlos de manera conjunta, para realizar inferencias de la información realizada y lograr una mayor comprensión del objeto de estudio propuesto [5], ya que el tema se abordó por primera vez en este contexto, el método de estudio es fenomenológico ya que intenta interpretar y comprender el proceso de desarrollo, ejecución y validación del modelo de evaluación y valoración de capacidades de (CTeI) en los centros de formación de la regional Antioquia del SENA, asimismo, el proceso de investigación es inductivo - deductivo. Los fundamentos expuestos en el marco del sistema de I + D [6] pretenden medir las capacidades de ciencia, tecnología e innovación desde cuatro dimensiones, que son entrada, proceso, salida y entorno.

Sin embargo, la propuesta se realizó teniendo en cuenta sólo las dimensiones de entrada y salida, dado que; en primer lugar, ignoramos la variable entorno debido a que las políticas de (CTeI) son cambiantes de acuerdo con la industria y la región. En segundo lugar, el proceso juega un papel clave en la formación de capacidades, sin embargo, está excluido por dos razones: la primera referida a la dificultad en cuantificar para algunas organizaciones los procesos y actividades de ciencia, tecnología e innovación y la segunda en cuanto a la intensidad de los insumos de (ACTI) y los resultados de producción que podrían representar el proceso de (CTeI) en conjunto. Se propuso un modelo de medición de indicadores para que las distintas organizaciones evalúen el impacto de las (ACTI), (see Figure 1) se realizó a través de indicadores de insumos y productos [7]. Siendo los insumos de (CTeI) todos aquellos recursos que respaldan las actividades de (CTeI). Por otro lado, los indicadores de salida (CTeI), son los resultados producto de las ACTI [8].

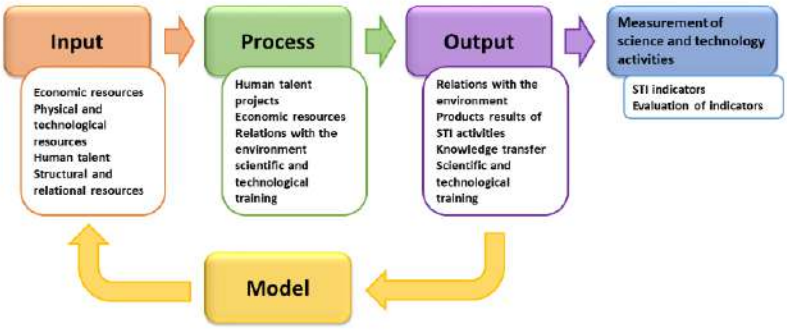


Figure 1. Conceptual model of the evaluation of science, technology and innovation management

3. Mathematical theory

3.1. *Clustering method*

La solución generalmente buscada es una partición de los m objetos, o sea, un conjunto de grupos donde un objeto pertenezca a un sólo grupo y el conjunto de dichos grupos contenga a todos los objetos, se parte de una muestra Ξ de m individuos, $X_1, ..., X_m$, cada uno de los cuales está representado por un vector $n-$ dimensional, $X_j = (x_{j1}, x_{j2}, ..., x_{jn})'$, $j = 1, ..., m$ y debemos encontrar una partición de la muestra en regiones $\omega_1, ..., \omega_c$ de forma que:

$$\bigcup_{i=1}^{c} \omega_i = \Xi$$
$$\omega_i \bigcap \omega_j = \emptyset \ ; \ i \neq j$$

3.2. *K-means method*

A menudo nos encontramos con el problema de particionar un conjunto de datos determinado en varios grupos: los puntos de datos en el mismo grupo comparten más similitudes. Mediante este algoritmo se busca minimizar la varianza total del sistema, Por lo tanto, dado un conjunto de datos $\{x_i\}_{i=1}^{n} \in R^p$, deseamos encontrar k centroides tales que se minimice la distancia total entre estos puntos de datos y sus centroides correspondientes, donde $\{c_j\}_{j=1}^{k}$ son los centroides y $\{\Psi_j\}_{i=1}^{k}$ son sus particiones correspondientes. Aqui $\{\Psi_j\}_{i=1}^{k}$ son mutuamente disyuntos y $\bigcup_{j=1}^{k} \Psi_j = [n]$. La suma total de la distancia es:

$$f(c_j, \Psi_j) := \sum_{j=1}^{k} \sum_{i \in \Psi_j} d(x_i, c_j)$$

Se llama función k-means (coste), que depende de los centroides $\{c_j\}_{j=1}^{k}$ y $\{\Psi_j\}_{i=1}^{k}$. En nuestro caso $d(x_i, c_j)$ se denota como la distancia, que mide la similitud entre x y y.

3.3. *Principal components method*

Es una técnica estadística multivariada cuyo objeto es representar un conjunto de variables observadas en un grupo de individuos por un número pequeño de variables nuevas construidas mediante combinaciones lineales de las variables originales. A continuación, definimos las componentes principales en términos poblacionales:

Sea $X = (X_1, ..., X_p)'$ un vector aleatorio $p-$ dimensional y sean μ y Σ su esperanza y covarianza

$$E[X] = \mu = 0 \quad Cov[X] = \Sigma = (\sigma_{ij})$$

Sea $\alpha = (\alpha_1, ..., \alpha_p)' \in R^p$ tal que $\alpha'\alpha = 1$. Se consideran combinaciones lineales

$$\alpha' X = (\alpha_1, ..., \alpha_p)\begin{pmatrix} X_1 \\ \vdots X_p \end{pmatrix} = \alpha_1 X_1 + ... + \alpha_p X_p = \sum_{i=1}^{p} \alpha_i X_i$$

Cuya varianza vendrá dada por

$$Var[\alpha' X] = E\left[(\alpha' X)^2\right] = E[\alpha' X X' \alpha] = \alpha' \Sigma \alpha$$

3.4. Cálculo de las componentes principales

Sea una matriz X con $n\ x\ p$ datos

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

De una población con matriz de covarianza Σ o correlación para las p variables involucradas
Por notación nos referiremos a las variables de este conjunto como elementos del vector aleatorio

$$\bar{x} = (x_1, x_{2,\ldots,} x_p)$$

Con el ACP intentaremos construir las siguientes combinaciones lineales:

$$\begin{aligned} y_1 &= \bar{a}_1' \bar{x} = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p, \\ y_2 &= \bar{a}_2' \bar{x} = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2p}x_p, \\ &\vdots \\ y_p &= \bar{a}_p' \bar{x} = a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p \end{aligned} \qquad \text{Ec. (1)}$$

Donde $\bar{a}_1'$, $\bar{a}_2'$, ..., $\bar{a}_p'$ son desconocidos y los $px1$ vectores contienen los pesos de las combinaciones lineales. Necesitamos que la primera componente principal y_1 explique la mayor varianza posible, que resultara de las combinaciones lineales de las variables iniciales en $\bar{x}$ y ahora se necesita imponer una restricción para la magnitud de los elementos de $\bar{a}_1$ ya que, si no se hace, la $var(y_1)$ aumentará desmedidamente, lo cual impide encontrar un vector $\bar{a}_1$ La elección que por lo general se hace es asumir que la longitud de 1, esto es $\|\bar{a}_1\|$ una vez que se encuentre la solución y se encuentre este vector, óptimo de pesos $\bar{a}_1$ la combinación lineal $\bar{a}_1' \bar{x}$ es llamado, la primera componente principal. El siguiente paso es encontrar un segundo vector $\bar{a}_2$ tal que la segunda componente $\bar{a}_2' \bar{x}$ tenga las siguientes propiedades:

1. Debe de explicar la mayor varianza posible de los datos restantes (considerando de nuevo la restricción.

2. No debe estar correlacionado con la primera componente principal.

Se busca que no tengan correlación, ya que buscamos terminar con el menor número de combinaciones lineales que expliquen la porción más grande posible de las varianzas de las variables iniciales. Una vez encontrado seta conjunto de valores $\bar{a}_2$, las combinaciones lineales$\bar{a}_2' \bar{x}$ se denominará la segunda componente principal. Podemos continuar con este comportamiento hasta que encontremos p vectores $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p$ que

aparecen en el lado derecho de las ecuaciones (1), por construcción las sucesivas componentes principales no están correlacionadas y formalmente hay tantas componentes principales como variables estudiadas.

4. Resultados de simulación

Se implementó diferentes etapas para la determinación de indicadores que evalúan la productividad académica y los procesos de gestión para centro de formación a partir de datos sintéticos. Se realiza inicialmente una etapa de clustering tanto para productividad y procesos utilizando k-means clustering, cuyos parámetros de validación externos e internos fueron medidos y se muestran en la Tabla 1. Estos valores exhiben y validan la aleatoriedad de los datos.

	Externos				**Internos**		
	Homogeneidad	**Completitud**	**índice Rand**	**Medida V**	**Silueta**	**Calinski-Harabasz**	**Davies-Bouldin**
Procesos	0.183	1.00	0	0.309	0.301	121.774	1.328
Productividad	0.197	1.00	0	0.330	0.171	41.810	1.891

Para la clasificación de clúster se utilizó el método de componentes principales(PCA) tanto para la etapa de procesos como productividad en centros de formación. En la Figura 2 se puede observar la tendencia de las variables de procesos y productividad respectivamente. En ella se generaron clasificación de tres tipos de clúster clasificados en rango de alto, medio y bajo. Para la determinación de estas categorías de clasificación, se consideró una comparativa entre la gráfica de componentes principales y el gráfico correlaciones, donde el factor principal para esta clasificación fue la cantidad de vectores, así como su magnitud en los cuadrantes del mapa de correlaciones.

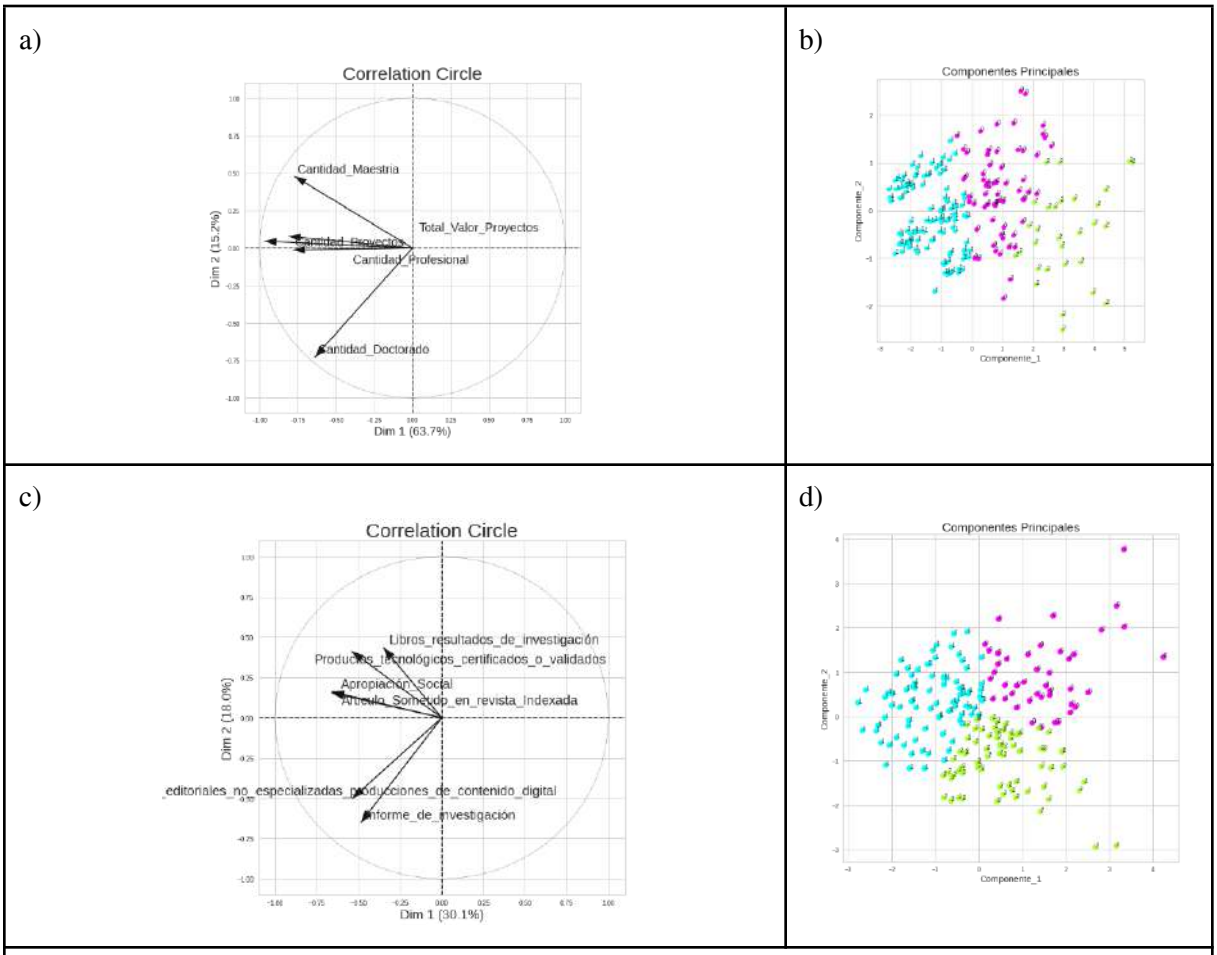


Figure 2. Gráfica para clasificación de clúster: a) Correlaciones de variables de productividad de centros de formación; b) Componentes principales redimensionados y clasificados por tipo clúster de productividad en centros de formación; c) Correlaciones de variables de procesos de centros de formación; d) Componentes principales redimensionados y clasificados por tipo clúster de procesos en centros de formación.

Determinada esta clasificación de clúster, se procede realizar mapas de geoespaciales que representen de manera gráfica la evolución de los centros de formación en alta, media y baja productividad- procesos, a lo largo del tiempo.

5. Visualización Geoespacial de resultados

Realizado el proceso de clasificación a partir de PCA y Clustering, se generaron tablas de datos que ubican espacialmente las clasificaciones generadas de alta, media y baja actividad de las dimensiones de gestión y productividad para un año específico, como se ilustra en la Figura 3. En ellas se puede observar con colores de rojo como bajo; morado como medio y azul magenta como alto.

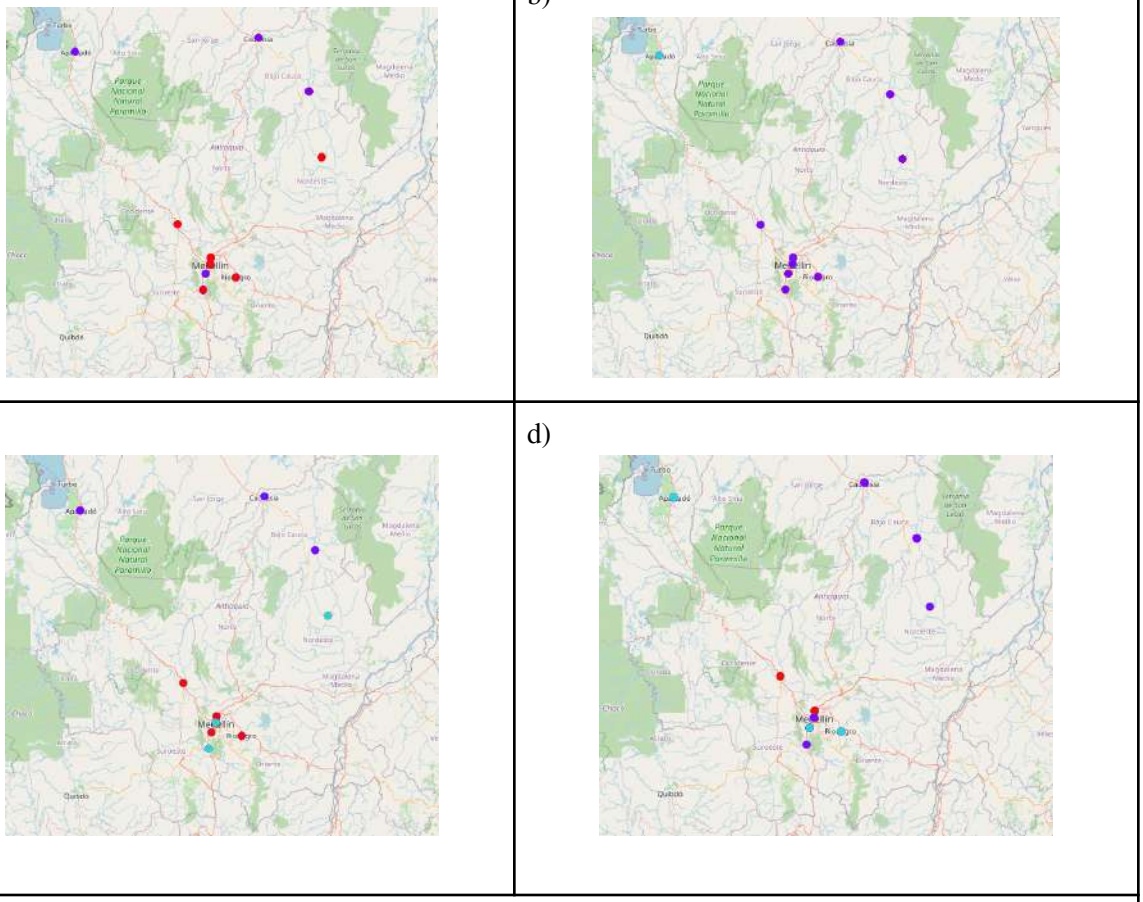


Figure 3. Mapas de actividades de productividad, producción científica y divulgativa, así como de gestión por centros formación en el departamento de Antioquia a lo largo del tiempo. Rojo: Bajo, Morado: Medio y Magenta: Alto. a) productividad de centros de formación 2018; b) productividad de centros de formación 2022.c) procesos de gestión en centros de formación 2018; d) procesos de gestión en centros de formación 2022.

6. Conclusiones

Procesos de clusterización a partir de K-Means y clasificación de clúster por medio de PCA, se muestran como un camino adecuado para la determinación de indicadores que orientan procesos de gestión y productividad investigativa en centros formación, esto a partir de una estrategia de datos sintéticos. Por otro lado, la generación de mapas geoespaciales facilita la visualización de los resultados en la analítica de datos, que generan alertas de alta, media y baja productividad y procesos de gestión respectivamente, así como su relación con el contexto a lo largo del tiempo para centros de formación SENA regional Antioquia, y destaca respecto a métodos tradicionales de visualización. De esto, a pesar de la aleatoriedad ofrecida por datos sintéticos, esta nos permitió generar una metodología que será replicable para generadores de indicadores de impacto en CTI para centros de formación tanto educativos como a nivel de industria. Finalmente se propone iniciar con el proceso de recolección de información a través de una aplicación con algoritmos para los centros de formación del SENA.

Referencias

1. A reference (Cuadrado & Angulo Cuentas, 2016).
   This reference has two entries but the second one is not numbered (it uses the 'Reference (no number)' style.
0. Another reference (Aguilar, 2017)
0. More references (Ortiz & Mendoza, 2016)
0. (Fuentes Cuadrado & Angulo Cuentas, 2016).
0. (Hernández-Sampieri y Mendoza, 2018)
0. (Li & Xie, 2010)
0. (Khatibi et al., 2017; Saunila & Ukko, 2012; Yam et al., 2004b)
0. (Adams et al., 2006; Nerkar & Paruchuri, 2005; Saunila, 2016).
0. Boris Mirkin, Mathematical classification and clustering. https://www.hse.ru/data/2010/04/04/1226332230/mcc.pdf
10. https://drive.google.com/file/d/1XqBcJNuisTujhhD89EJYQUrQeOZQKnZ_/view?usp=sharing

ANEXO 3. Modelo finalizado de clusterización- K means y clasificación PCA no supervisados para las dimensiones de productividad y procesos de centros de formación en la regional Antioquia para aplicación en API

Procesos_Mapas_API

```
#librerías
import numpy as np
import pandas as pd

import folium
from folium import plugins

from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors


#Cargar archivos desde drive
from google.colab import drive
drive.mount('/content/drive')
```

```
     Mounted at /content/drive
```

```
from google.colab import files
path1 = "/content/drive/MyDrive/Proyecto SGPS 9748/TPProceso.xlsx"
path2 = "/content/drive/MyDrive/Proyecto SGPS 9748/Datos.xlsx"
df = pd.read_excel(path1)
df_centros = pd.read_excel(path2,sheet_name= 'Centros_Formacion')
df
```

	Centro_Año	Q_Proyectos	QNA_5007	QNA_5009	QNA_5010	TV_Proyectos
0	9101-2010	1	1	2	0	4.652924e+08
1	9127-2010	2	2	4	1	3.920052e+08

```
df.columns
```

```
Index(['Centro_Año', 'Q_Proyectos', 'QNA_5007', 'QNA_5009', 'QNA_5010',
       'TV_Proyectos'],
      dtype='object')
```

	...	...	...	...	...	...	...

```
scaler = StandardScaler()
df[['Q_Proyectos', 'QNA_5007', 'QNA_5009', 'QNA_5010',
       'TV_Proyectos']] = scaler.fit_transform(df[['Q_Proyectos', 'QNA_5007', 'QNA_5009', 'QN
       'TV_Proyectos']])
```

```
# variables
y = df[['Centro_Año']]
x = df.drop(['Centro_Año'], axis= 1)
#K-means a la base de datos
clustering=KMeans(n_clusters=3, max_iter=300,random_state=0) #Crea el modelo
clustering.fit(x) #Aplica el modelo a la base de datos
```

```
KMeans(n_clusters=3, random_state=0)
```

```
#creando columna de clusters
df["KMeans_Clusters"]=clustering.labels_
```

```
# la asignación de etiqueta de cluster cambia al ser compilado, por lo que no debe ser consid
# para ser corregido, debe realizarse a partir del PCA y el mapa de correlaciones.
# Es posbile corregirse, posiblemente con algun condicional en los signos y magnitudes del ve
produc= df['KMeans_Clusters'].replace({1:2,0: 1, 2:0})
produc= df['KMeans_Clusters'].replace({2: 'Alta Gestión',1: 'Media Gestión', 0: 'Baja Gestión
df['Gestión'] = pd.concat([produc], axis=1)
centros_años = df["Centro_Año"].str.split("-",expand=True)
centros_años.columns = ['Codigo_Centro', 'Año']
df = pd.concat([df, centros_años], axis=1)
```

```
df_map = df[['Codigo_Centro','Año','KMeans_Clusters','Gestión']]
df_map['Codigo_Centro'] = df_map['Codigo_Centro'].astype(int)
df_map['Año'] = df_map['Año'].astype(int)
df_map =df_map.merge(df_centros,on='Codigo_Centro',how='inner')
df_map = df_map.drop(['Codigo_Regional','Codigo_Municipio','Codigo_Departamento'],axis= 1)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
  This is separate from the ipykernel package so we can avoid doing imports until
```

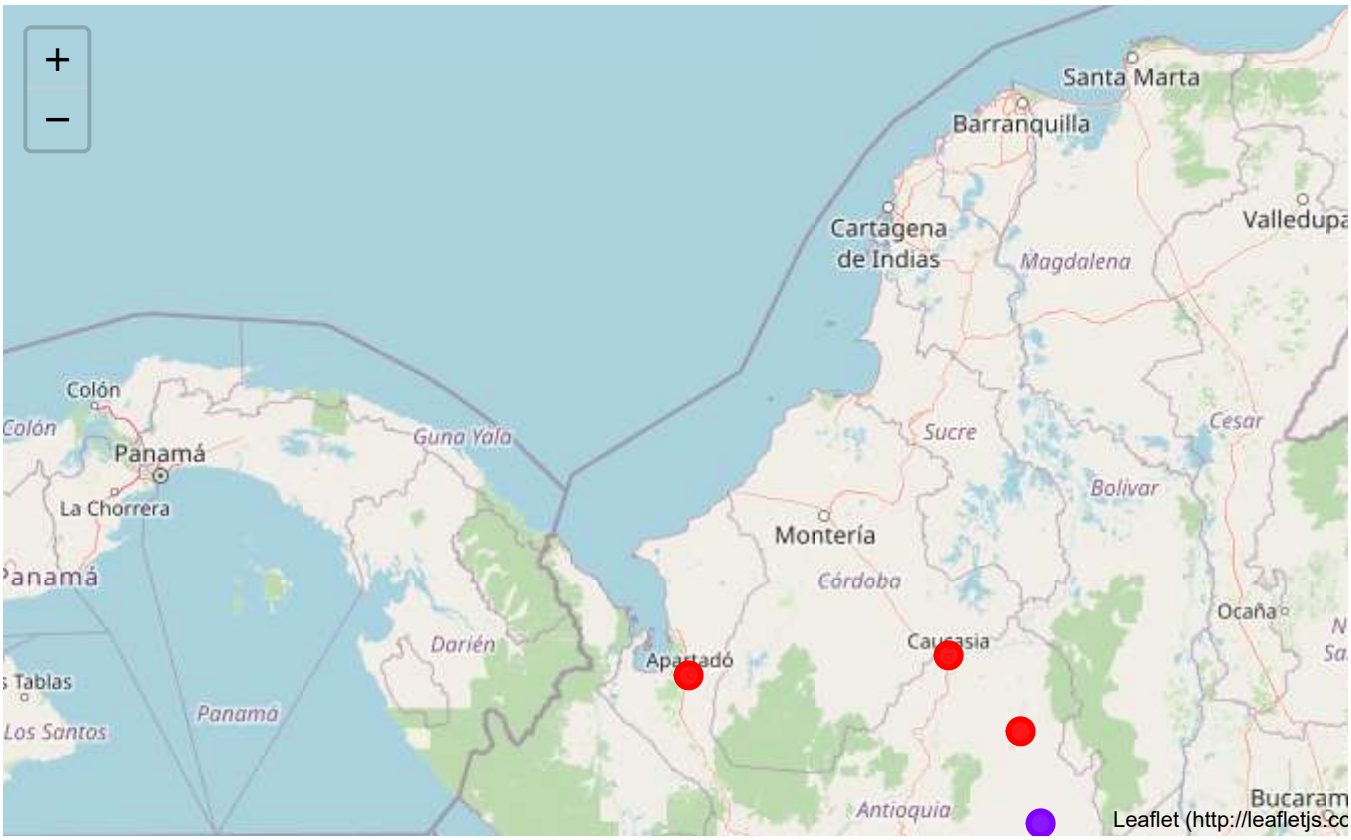
```
# Separación de centros por años
df_map_2010 = df_map[(df_map['Año']<2011)&(df_map['Año']>=2010)]
df_map_2010 = df_map_2010.drop(['Año'],axis = 1)


latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lng, cluster, cent, gest in zip(df_map_2010['latitud'],
                            df_map_2010['longitud'],
                            df_map_2010['KMeans_Clusters'],
                            df_map_2010['Nombre_Centro'],
                            df_map_2010['Gestión']) :


    label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
    folium.vector_layers.CircleMarker(
        [lat, lng],
        radius=6,
        #popup=label,
        tooltip = str(cent) +' , '+ str(gest),
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.9).add_to(map_clusters)

map_clusters
```

```
# Separación de centros por años
df_map_2014 = df_map[(df_map['Año']<2015)&(df_map['Año']>=2014)]
df_map_2014 = df_map_2014.drop(['Año'],axis = 1)


latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
```

```
for lat, lng, cluster, cent, gest in zip(df_map_2014['latitud'],
                            df_map_2014['longitud'],
                            df_map_2014['KMeans_Clusters'],
                            df_map_2014['Nombre_Centro'],
                            df_map_2014['Gestión']) :


    label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
    folium.vector_layers.CircleMarker(
        [lat, lng],
        radius=6,
        #popup=label,
        tooltip = str(cent) +' , '+ str(gest),
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.9).add_to(map_clusters)

map_clusters
```

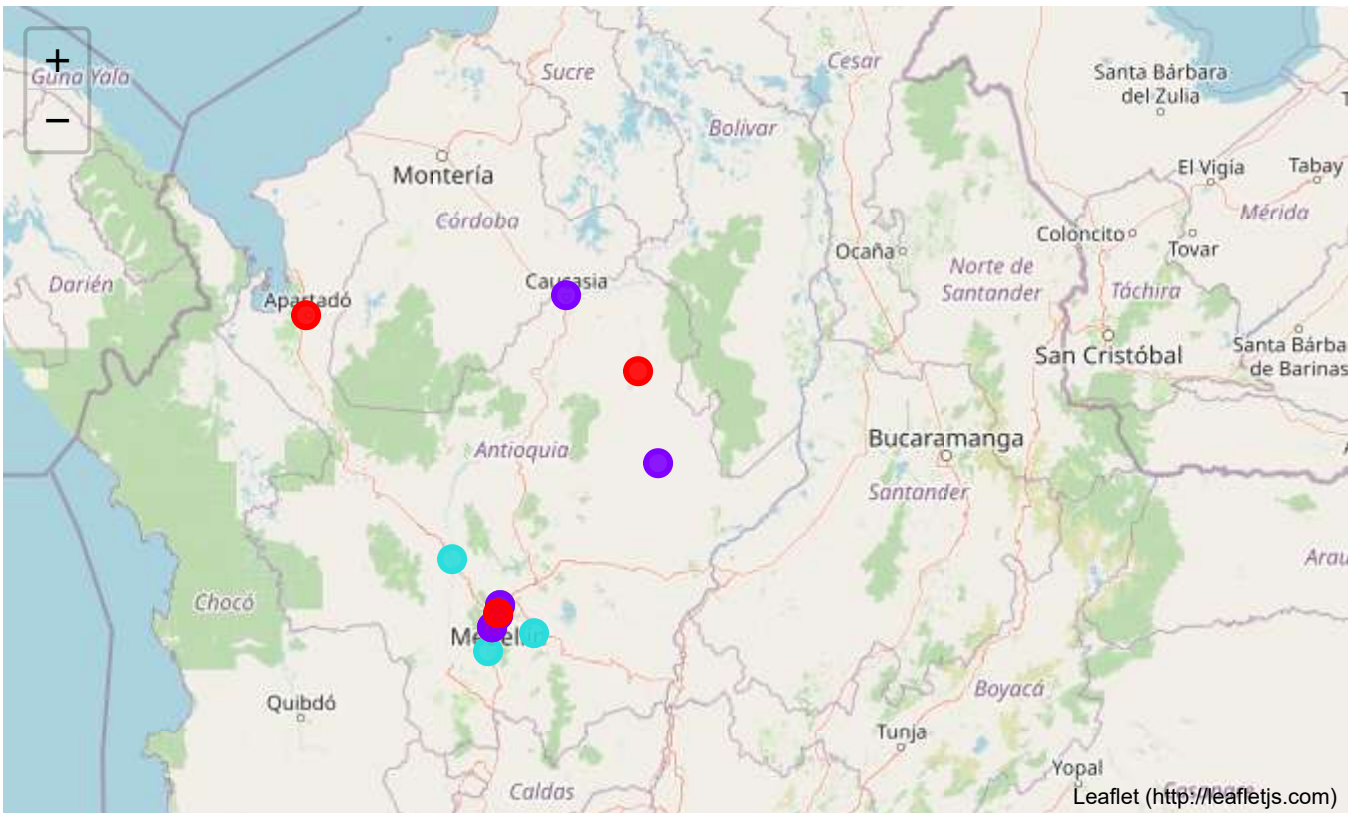
```
# Separación de centros por años
df_map_2018 = df_map[(df_map['Año']<2019)&(df_map['Año']>=2018)]
df_map_2018 = df_map_2018.drop(['Año'],axis = 1)
```

```
latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lng, cluster, cent, gest in zip(df_map_2018['latitud'],
                            df_map_2018['longitud'],
                            df_map_2018['KMeans_Clusters'],
                            df_map_2018['Nombre_Centro'],
                            df_map_2018['Gestión']) :


    label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
    folium.vector_layers.CircleMarker(
        [lat, lng],
        radius=6,
        #popup=label,
        tooltip = str(cent) +' , '+ str(gest),
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.9).add_to(map_clusters)

map_clusters
```




```
# Separación de centros por años
df_map_2021 = df_map[(df_map['Año']<2022)&(df_map['Año']>=2021)]
df_map_2021 = df_map_2021.drop(['Año'],axis = 1)


latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lng, cluster, cent, gest in zip(df_map_2021['latitud'],
                            df_map_2021['longitud'],
                            df_map_2021['KMeans_Clusters'],
                            df_map_2021['Nombre_Centro'],
                           df_map_2021['Gestión']) :
```

```
        label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
        folium.vector_layers.CircleMarker(
            [lat, lng],
            radius=6,
            #popup=label,
            tooltip = str(cent) +' , '+ str(gest),
            color=rainbow[cluster-1],
            fill=True,
            fill_color=rainbow[cluster-1],
            fill_opacity=0.9).add_to(map_clusters)

map_clusters
```



✓ 0 s completado a las 20:54

Productividad_Mapas_API

```
#librerías
import numpy as np
import pandas as pd

import folium
from folium import plugins

from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors


#Cargar archivos desde drive
from google.colab import drive
drive.mount('/content/drive')
```

```
     Mounted at /content/drive
```

```
#Path en Drive
from google.colab import files
path = "/content/drive/MyDrive/Proyecto SGPS 9748/Datos.xlsx"
path2 = '/content/drive/MyDrive/Proyecto SGPS 9748/PCPA.xlsx'

#Dataframes
principal_proyectos= pd.read_excel(path,sheet_name= 'Principal_Proyectos')
principal_talento_humano = pd.read_excel(path,sheet_name= 'Principal_Talento_Humano')
principal_productos = pd.read_excel(path,sheet_name= 'Principal_Productos')
df = pd.read_excel(path2)
df_centros = pd.read_excel(path,sheet_name= 'Centros_Formacion')


#Escalar datos: desviación estandar
scaler = StandardScaler()
df[[ 'P_8020', 'P_8025', 'P_8019', 'P_8039', 'P_8016','P_8031']] = scaler.fit_transform(df[['


# variables
y = df[['Centro_Año']]
x = df.drop(['Centro_Año'], axis= 1)
#K-means a la base de datos
clustering=KMeans(n_clusters=3, max_iter=300,random_state=0) #Crea el modelo
clustering.fit(x) #Aplica el modelo a la base de datos
```

```
KMeans(n_clusters=3, random_state=0)
```

```
#creando columna de clusters
df["KMeans_Clusters"]=clustering.labels_
```

```
df
```

	Centro_Año	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031	KMeans_Cl
0	9101-2010	-0.560125	0.857305	-0.505342	2.814094	-1.143520	0.780675	
1	9127-2010	-0.560125	-1.274003	-1.170602	0.168528	0.930772	0.068480	
2	9201-2010	-1.189888	-1.274003	-1.170602	-1.154255	-1.143520	0.068480	
3	9202-2010	-1.189888	-0.563567	-1.170602	0.168528	-1.143520	-0.643715	
4	9203-2010	2.588687	0.146869	0.159918	1.491311	0.239341	0.780675	
...	...	...	...	...	...	...	...	
203	9501-2022	0.699400	2.988612	2.820960	-0.492864	-0.452089	-0.643715	
204	9502-2022	0.069637	0.857305	0.159918	-0.492864	-0.452089	-0.643715	
205	9503-2022	0.069637	-1.274003	0.159918	-0.492864	0.239341	-0.643715	
206	9504-2022	1.329162	0.146869	-1.170602	-0.492864	-0.452089	-0.643715	
207	9549-2022	-0.560125	-1.274003	0.825179	0.829919	0.239341	1.492871	

208 rows × 8 columns

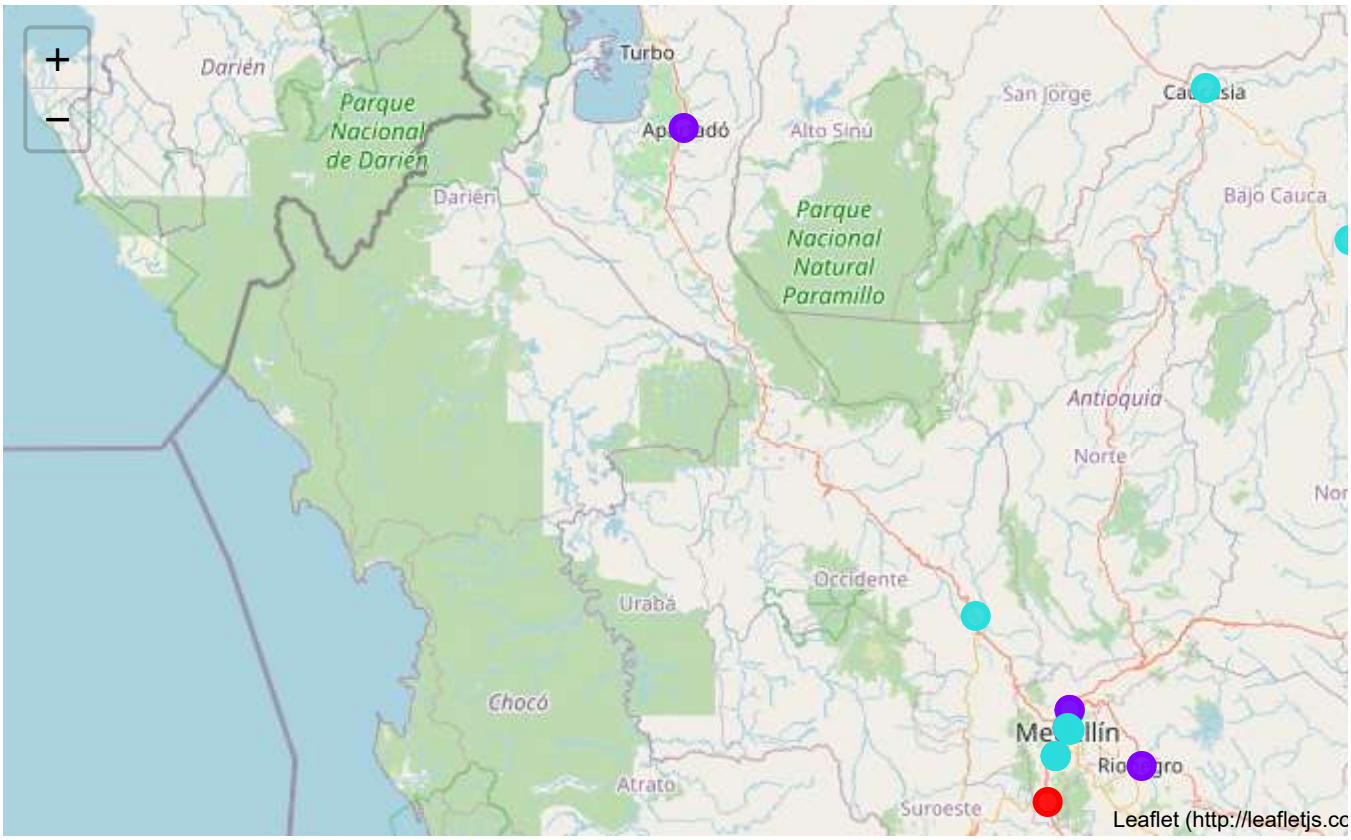
```
# la asignación de etiqueta de cluster cambia al ser compilado, por lo que no debe ser consid
# para ser corregido, debe realizarse a partir del PCA y el mapa de correlaciones.
# Es posbile corregirse, posiblemente con algun condicional en los signos y magnitudes del ve
produc= df['KMeans_Clusters'].replace({0:0,2:1, 1:2})
produc= df['KMeans_Clusters'].replace({0: 'Baja Productividad', 1: 'Media Productividad',2: '
df['Productividad'] = pd.concat([produc], axis=1)
centros_años = df["Centro_Año"].str.split("-",expand=True)
centros_años.columns = ['Codigo_Centro', 'Año']
df = pd.concat([df, centros_años], axis=1)
```

```
df_map = df[['Codigo_Centro','Año','KMeans_Clusters','Productividad']]
df_map['Codigo_Centro'] = df_map['Codigo_Centro'].astype(int)
df_map['Año'] = df_map['Año'].astype(int)
df_map =df_map.merge(df_centros,on='Codigo_Centro',how='inner')
df_map = df_map.drop(['Codigo_Regional','Codigo_Municipio','Codigo_Departamento'],axis= 1)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
  This is separate from the ipykernel package so we can avoid doing imports until
```

```
# Separación de centros por años
df_map_2010 = df_map[(df_map['Año']<2011)&(df_map['Año']>=2010)]
df_map_2010 = df_map_2010.drop(['Año'],axis = 1)


latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lng, cluster, cent, gest in zip(df_map_2010['latitud'],
                            df_map_2010['longitud'],
                            df_map_2010['KMeans_Clusters'],
                            df_map_2010['Nombre_Centro'],
                            df_map_2010['Productividad']) :


    label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
    folium.vector_layers.CircleMarker(
        [lat, lng],
        radius=6,
        #popup=label,
        tooltip = str(cent) +' , '+ str(gest),
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.9).add_to(map_clusters)

map_clusters
```

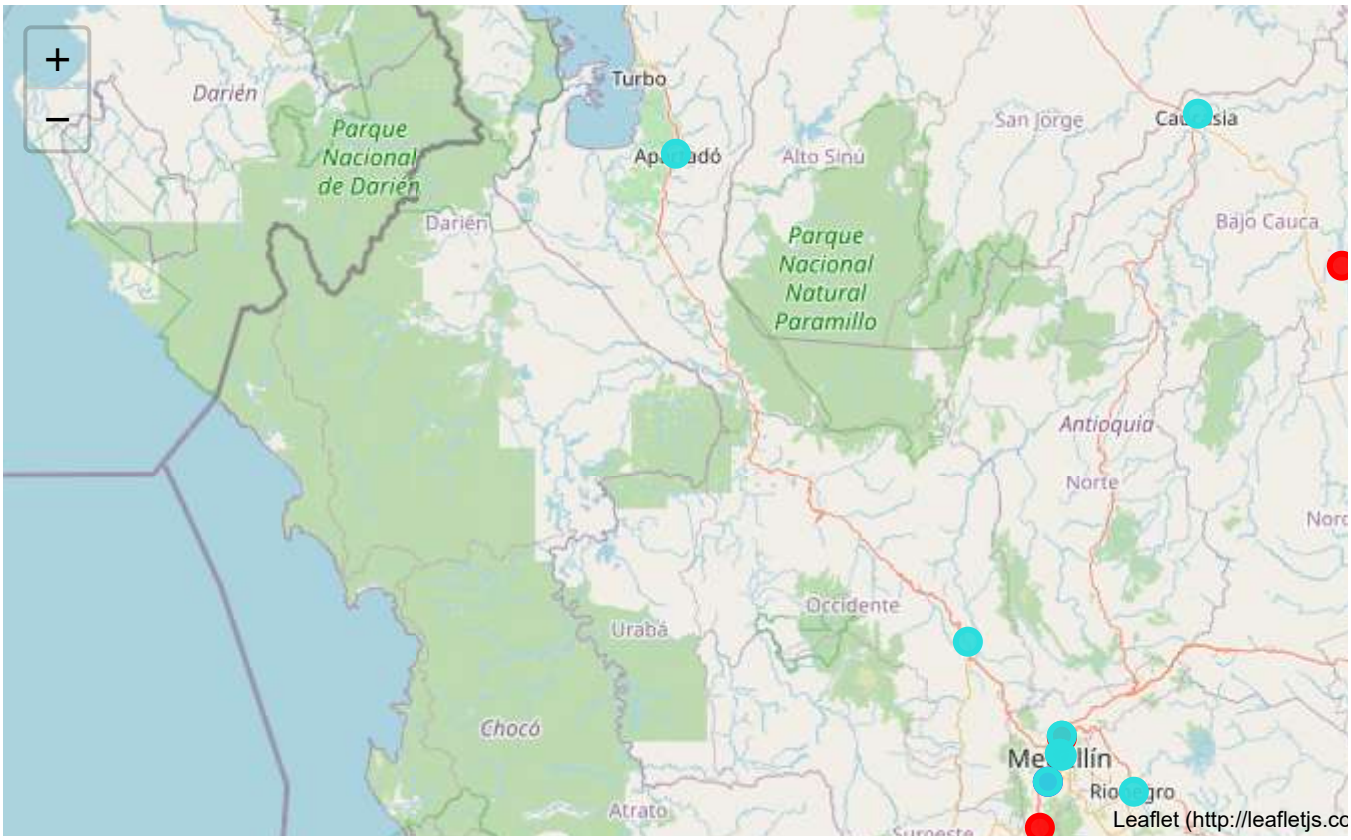



```
# Separación de centros por años
df_map_2014 = df_map[(df_map['Año']<2015)&(df_map['Año']>=2014)]
df_map_2014 = df_map_2014.drop(['Año'],axis = 1)


latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
```

```
for lat, lng, cluster, cent, gest in zip(df_map_2014['latitud'],
                            df_map_2014['longitud'],
                            df_map_2014['KMeans_Clusters'],
                            df_map_2014['Nombre_Centro'],
                            df_map_2014['Productividad']) :


    label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
    folium.vector_layers.CircleMarker(
        [lat, lng],
        radius=6,
        #popup=label,
        tooltip = str(cent) +' , '+ str(gest),
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.9).add_to(map_clusters)

map_clusters
```

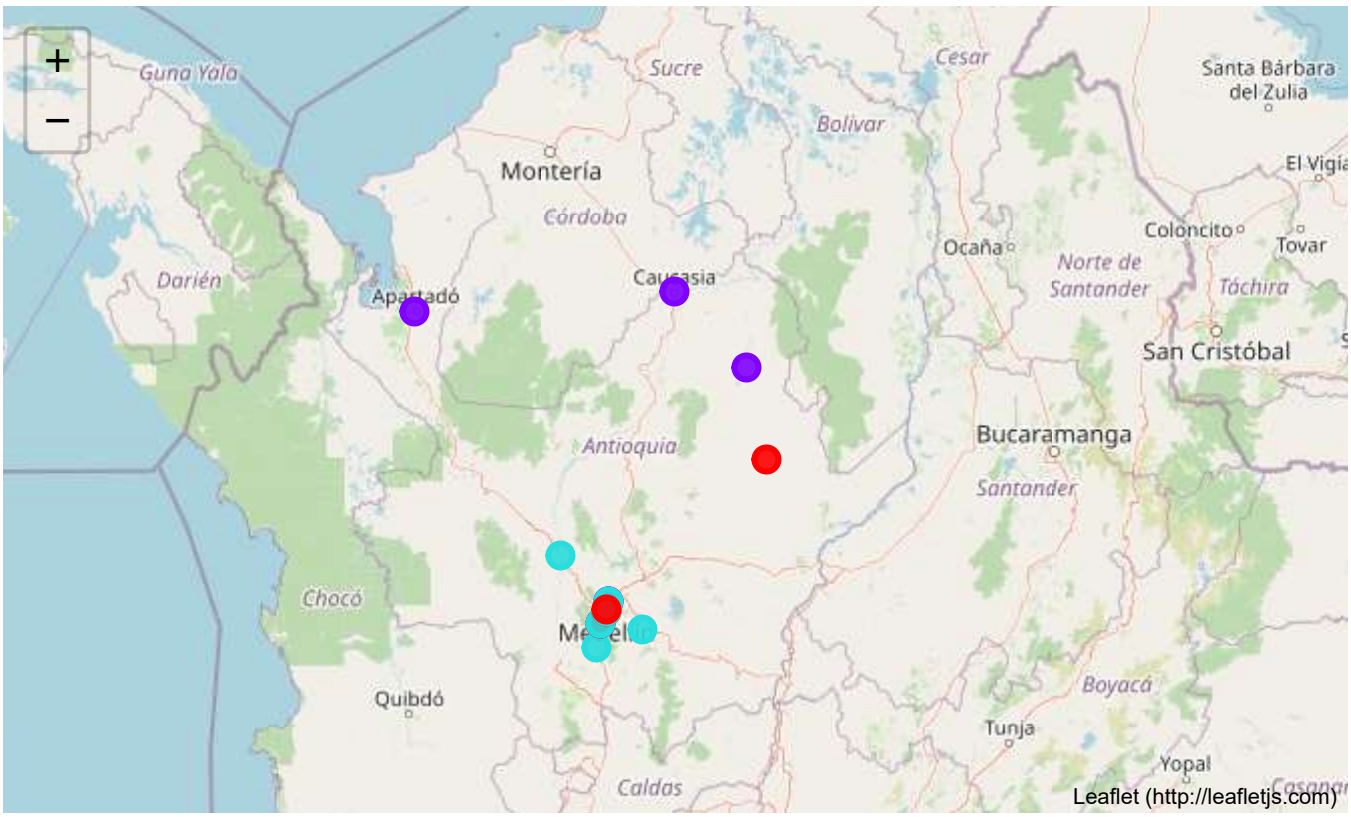
```
# Separación de centros por años
df_map_2018 = df_map[(df_map['Año']<2019)&(df_map['Año']>=2018)]
df_map_2018 = df_map_2018.drop(['Año'],axis = 1)
```

```
latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lng, cluster, cent, gest in zip(df_map_2018['latitud'],
                            df_map_2018['longitud'],
                            df_map_2018['KMeans_Clusters'],
                            df_map_2018['Nombre_Centro'],
                            df_map_2018['Productividad']) :


    label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
    folium.vector_layers.CircleMarker(
        [lat, lng],
        radius=6,
        #popup=label,
        tooltip = str(cent) +' , '+ str(gest),
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.9).add_to(map_clusters)

map_clusters
```




```
# Separación de centros por años
df_map_2021 = df_map[(df_map['Año']<2022)&(df_map['Año']>=2021)]
df_map_2021 = df_map_2021.drop(['Año'],axis = 1)


latitude = 6.51033052
longitude = -75.48477721

map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

# set color scheme for the clusters
x = np.arange(2)
ys = [i + x + (i*x)**2 for i in range(4)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lng, cluster, cent, gest in zip(df_map_2021['latitud'],
                            df_map_2021['longitud'],
                            df_map_2021['KMeans_Clusters'],
                            df_map_2021['Nombre_Centro'],
                            df_map_2021['Productividad']) :
```

```
    label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
    folium.vector_layers.CircleMarker(
        [lat, lng],
        radius=6,
        #popup=label,
        tooltip = str(cent) +' , '+ str(gest),
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.9).add_to(map_clusters)

map_clusters
```



✓ 0 s completado a las 20:54

ANEXOS 4. Presentación de avances técnicos julio 2022.

El empleo es de todos

Mintrabajo

Proyecto SGPS 9748

Prototipo Innovación

Presentación de avances 12

Modelo de evaluación y valoración de la Gestión de CTeI para los Centros SENA Regional Antioquia aplicando analítica de datos

4 de julio del 2022

SENA

GC-F-004v.05

Contenido

1. Base de datos procesos
2. Base de datos productividad
3. Modelo de analítica de datos para medición de gestión por centros
   - Clústering método k-means para medir nivel de gestión
   - Clustering K-means y PCA para medir nivel de gestión
   - Representación en el plano factorial de variables de interés

GC-F-004v.05

Contenido

3. Modelo de analítica de datos para medición de gestión por centros
   - Dendograma de centros para medición de gestión
   - Clustering jerárquico para medir nivel de gestión
4. Modelo de analítica de datos para medición de productividad por centros
   - Clustering método k-means para medir nivel de productividad

GC-F-004v.05

Contenido

4. Modelo de analítica de datos para medición de productividad por centros

- Clustering K-means y PCA para medir nivel de productividad
- Representación en el plano factorial de variables de interés
- Dendograma de centros para medición de productividad
- Clustering jerárquico para medir nivel de productividad

GC-F-004v.05

1. Base de datos Procesos

A continuación se evidencia la tabla de procesos en R stadistics y en excel

Filter

	Centro_Año	Q_Proyectos	QNA_5007	QNA_5009	QNA_5010	TV_Proyectos
1	9101-2010	1	1	2	0	465292449
2	9127-2010	2	2	4	1	392005181
3	9201-2010	1	4	1	1	455662749
4	9202-2010	3	5	5	2	407185228
5	9203-2010	2	4	3	1	290617466
6	9204-2010	3	5	7	1	564001653
7	9205-2010	1	3	1	1	508395371
8	9206-2010	1	3	2	1	300860419
9	9301-2010	3	6	6	0	652146153
10	9401-2010	3	2	4	1	584555799
11	9402-2010	2	2	6	2	469064097
12	9501-2010	2	1	4	1	550483170
13	9502-2010	3	3	7	1	518339355
14	9503-2010	2	1	4	2	385085638
15	9504-2010	1	1	3	0	356850041
16	9549-2010	2	1	3	1	362870121
17	9101-2011	3	5	6	1	541412297
18	9127-2011	4	6	11	0	625084148

Showing 1 to 18 of 208 entries, 6 total columns

Tabla 1. Tabla procesos en R

	A	B	C	D	E	F
1	Centro_Año	Q_Proyectos	QNA_5007	QNA_5009	QNA_5010	TV_Proyectos
2	9101-2010	1	1	2	0	465292449
3	9127-2010	2	2	4	1	392005181
4	9201-2010	1	4	1	1	455662749
5	9202-2010	3	5	5	2	407185228
6	9203-2010	2	4	3	1	290617466
7	9204-2010	3	5	7	1	564001653
8	9205-2010	1	3	1	1	508395371
9	9206-2010	1	3	2	1	300860419
10	9301-2010	3	6	6	0	652146153
11	9401-2010	3	2	4	1	584555799
12	9402-2010	2	2	6	2	469064097
13	9501-2010	2	1	4	1	550483170
14	9502-2010	3	3	7	1	518339355
15	9503-2010	2	1	4	2	385085638
16	9504-2010	1	1	3	0	356850041
17	9549-2010	2	1	3	1	362870121
18	9101-2011	3	5	6	1	541412297
19	9127-2011	4	6	11	0	625084148
20	9201-2011	2	1	1	2	450792096
21	9202-2011	1	1	1	1	329325789
22	9203-2011	3	3	9	1	657768885
23	9204-2011	1	0	2	0	398642441

Sheet1

Tabla 2. Tabla procesos en excel

2. Base de datos de productividad

A continuación se evidencia la tabla de productividad en R stadistics y en excel

Filter

	Centro_Año	Cantidad_Proyectos	Fecha_inicio_Proyecto	Cantidad_Productos	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031
1	9101-2010	4	2010	14	1	3	1	6	0	3
2	9127-2010	2	2010	8	1	0	0	2	3	2
3	9201-2010	1	2010	2	0	0	0	0	0	2
4	9202-2010	1	2010	4	0	1	0	2	0	1
5	9203-2010	5	2010	19	6	2	2	4	2	3
6	9204-2010	2	2010	7	0	1	2	0	1	3
7	9205-2010	1	2010	2	0	0	1	0	0	1
8	9206-2010	4	2010	14	2	4	2	1	4	1
9	9301-2010	2	2010	8	0	2	2	1	3	0
10	9401-2010	1	2010	4	1	0	2	0	0	1
11	9402-2010	2	2010	7	2	1	1	1	0	2
12	9501-2010	3	2010	10	1	0	1	5	0	3
13	9502-2010	1	2010	3	0	1	0	0	2	0
14	9503-2010	3	2010	10	1	2	5	0	1	1
15	9504-2010	4	2010	14	2	4	2	2	2	2
16	9549-2010	3	2010	11	2	1	1	2	0	5
17	9101-2011	2	2011	6	0	0	1	3	1	1
18	9127-2011	2	2011	11	2	0	3	0	0	6

Showing 1 to 18 of 208 entries, 10 total columns

Tabla 3. Tabla productividad en R

2. Base de datos de productividad

A continuación se evidencia la tabla de productividad en R stadistics y en excel

	A	B	C	D	E	F	G	H	I	J
1	Centro_Año	Cantidad_Proyectos	Fecha_inicio_Proyecto	Cantidad_Productos	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031
2	9101-2010	4	2010	14	1	3	1	6	0	3
3	9127-2010	2	2010	8	1	0	0	2	3	2
4	9201-2010	1	2010	2	0	0	0	0	0	2
5	9202-2010	1	2010	4	0	1	0	2	0	1
6	9203-2010	5	2010	19	6	2	2	4	2	3
7	9204-2010	2	2010	7	0	1	2	0	1	3
8	9205-2010	1	2010	2	0	0	1	0	0	1
9	9206-2010	4	2010	14	2	4	2	1	4	1
10	9301-2010	2	2010	8	0	2	2	1	3	0
11	9401-2010	1	2010	4	1	0	2	0	0	1
12	9402-2010	2	2010	7	2	1	1	1	0	2
13	9501-2010	3	2010	10	1	0	1	5	0	3
14	9502-2010	1	2010	3	0	1	0	0	2	0
15	9503-2010	3	2010	10	1	2	5	0	1	1
16	9504-2010	4	2010	14	2	4	2	2	2	2
17	9549-2010	3	2010	11	2	1	1	2	0	5
18	9101-2011	2	2011	6	0	0	1	3	1	1
19	9127-2011	2	2011	11	2	0	3	0	0	6
20	9201-2011	1	2011	6	1	0	2	2	0	1
21	9202-2011	4	2011	12	4	2	1	0	2	3
22	9203-2011	1	2011	9	2	2	2	2	1	0
23	9204-2011	4	2011	13	2	3	3	0	3	2

Sheet1

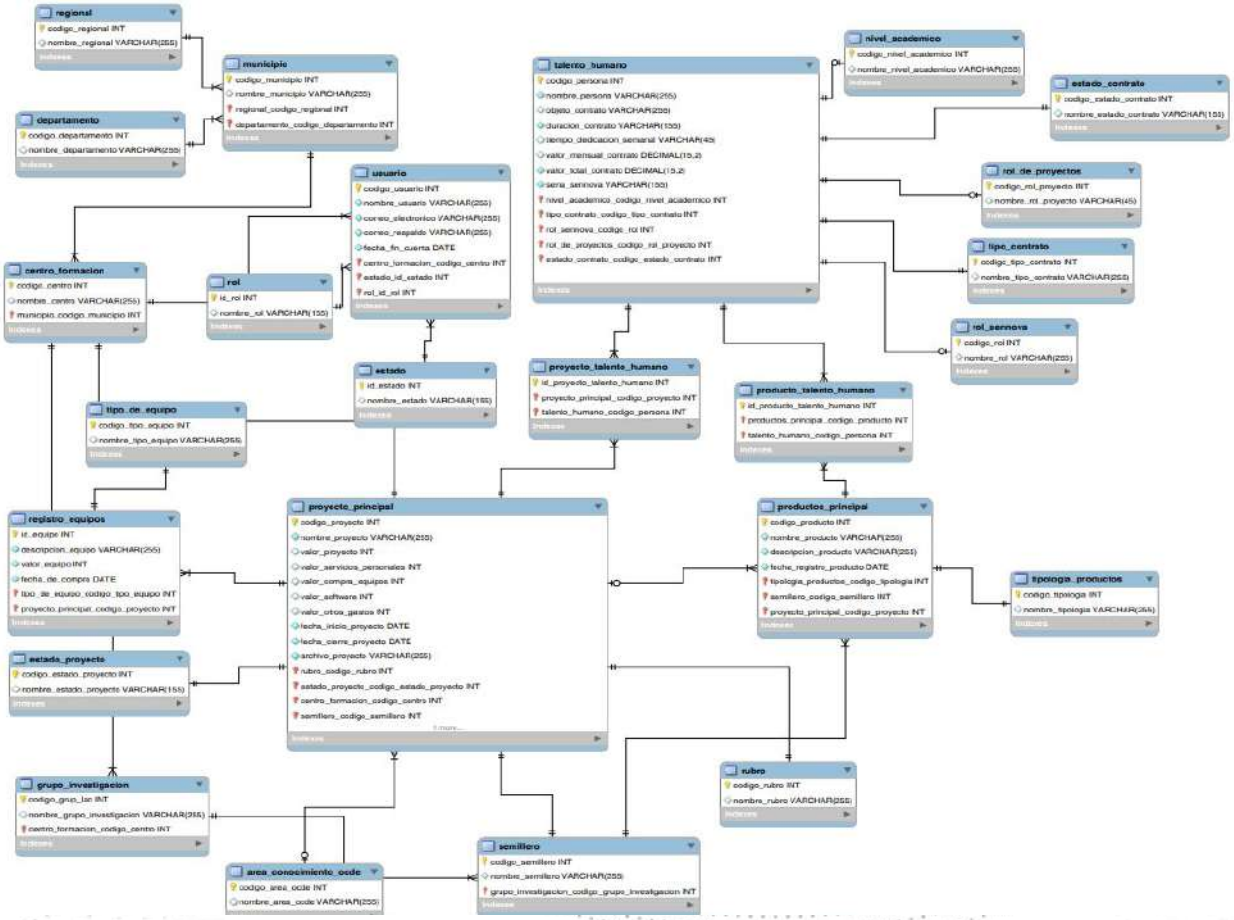
Tabla 3. Tabla productividad en excel

Indicadores y correspondencia de variables de entrada y salida del modelo

Nro	DIMENSIÓN (Proceso de entradas y salidas)	CATEGORIA	INDICADOR (Variables)	INDICADOR FINAL
1	Insumos para actividades de Ciencia, Tecnología e Innovación (ACTI) (Entradas en el modelo)	RECURSOS ECONÓMICOS	GASTO/INVERSIÓN en ACTI	Suma de los diferentes rubros ejecutados.
2		RECURSOS FÍSICOS Y TECNOLÓGICOS	NÚMERO DE LABORATORIOS PARA ACTI	AMBIENTES DE APRENDIZAJE (Laboratorios dotados con recursos de SENNOVA)
3			INVERSIÓN DE EQUIPOS DE PARA ACTI	Suma del rubro de equipos y maquinaria.
4			NÚMERO DE LICENCIAS (Campo de conocimiento segun la OCDE) (Por programas de educacion)	-Software de aplicaciones especializadas -Software de programación
5		TALENTO HUMANO	NUMERO DE PERSONAS DEDICADAS A ACTI	
6		RECURSOS ESTRUCTURALES Y RELACIONALES	NÚMERO DE UNIDADES ORGANIZATIVAS (GRUPOS DE INVESTIGACIÓN - SEMILLEROS)	
7			NÚMERO DE PROGRAMAS IMPACTADOS DE EDUCACIÓN SUPERIOR OFERTADOS (PROGRAMAS DEL SENA POR ACTI)	
8			NÚMERO DE CONVENIOS DE COLABORACIÓN PARA ACTI	

9	Proceso (Planta del modelo)	PROYECTOS	NÚMERO DE PROYECTOS
10		TALENTO HUMANO	PROPORCIÓN DE GASTO EN ACTI POR PERSONAS DEDICAS A ACTI
11		RECURSOS ECONÓMICOS	PORCENTAJE DE GASTO EN ACTI en relación con el PRESUPUESTO TOTAL
12			Porcentaje de gasto en INVESTIGACIÓN APLICADA en relación con el gasto total por ACTI
13		RELACIONES CON EL ENTORNO	EVENTOS DE MOVILIDAD PARA ACTI
14		FORMACIÓN CIENTÍFICA Y TECNOLÓGICA	Formacion de aprendices en semilleros de investigacion
15	Resultado (Salidas del modelo)	RELACIONES CON EL ENTORNO	Convenios con otras instituciones
16		PRODUCTOS RESULTADOS DE LAS ACTIVIDADES DE CTI	NÚMERO DE PRODUCTOS DE I+D (Productos de I+D por tipologia Colciencias)
17		TRANFERENCIA DE CONOCIMIENTO	NÚMERO DE EVENTOS DE CTI ORGANIZADOS POR LA UNIVERSIDAD (Número de eventos de CTI por clasificación ACTI)
19		FORMACIÓN CIENTÍFICA Y TECNOLÓGICA	Numero de aprendices con formacion cientifica
20	Productividad (Rendimiento del modelo)	TALENTO HUMANO	PROPORCIÓN DE PRODUCTOS DE CTI POR PERSONA DEDICADAS A ACTI, EN RELACIÓN CON EL TOTAL DE PERSONAL
21			PROPORCIÓN DE PRODUCTOS DE CTI POR UNIDAD ORGANIZATIVA
22		RELACIONES CON EL ENTORNO	PROPORCIÓN DE EVENTOS DE MOVILIDAD PARA ACTI POR PERSONAS DEDICADAS A ACTI

Modelo relacional



GC-F-004v.05

3. Modelo de analítica de datos para medición de gestión por centros

Clústering método k-means para medir nivel de gestión

Antecedentes de la agrupación en clústeres de k-medias: **Metodo Bootstrapping**

Considere un conjunto de datos que contiene n instancias/observaciones cada una descrita por m atributos. El problema de agrupamiento de k-medias es dividir el n instancias en k clústeres, con los clústeres particionando las instancias $(x_1, x_2, \dots, x_n)$ en los subconjuntos $Q_{1..k}$ Los subconjuntos se pueden resumir como puntos $(\boldsymbol{C_{1..k}})$ en el espacio m dimensional, comúnmente conocido como **centroides** o centros de conglomerados, cuyas coordenadas son el promedio de todos los puntos pertenecientes al subconjunto. Nos referiremos a esta colección de centroides obtenidos como un modelo de aplicación del algoritmo de agrupamiento. El agrupamiento de K-means también se puede considerar como una cuantificación vectorial con el objetivo de minimizar el error de cuantificación del vector (también conocido como distorsión) que se muestra en la ecuación (1). La solución matemática trivial es tener un clúster para cada instancia, pero normalmente $k \ll n$.

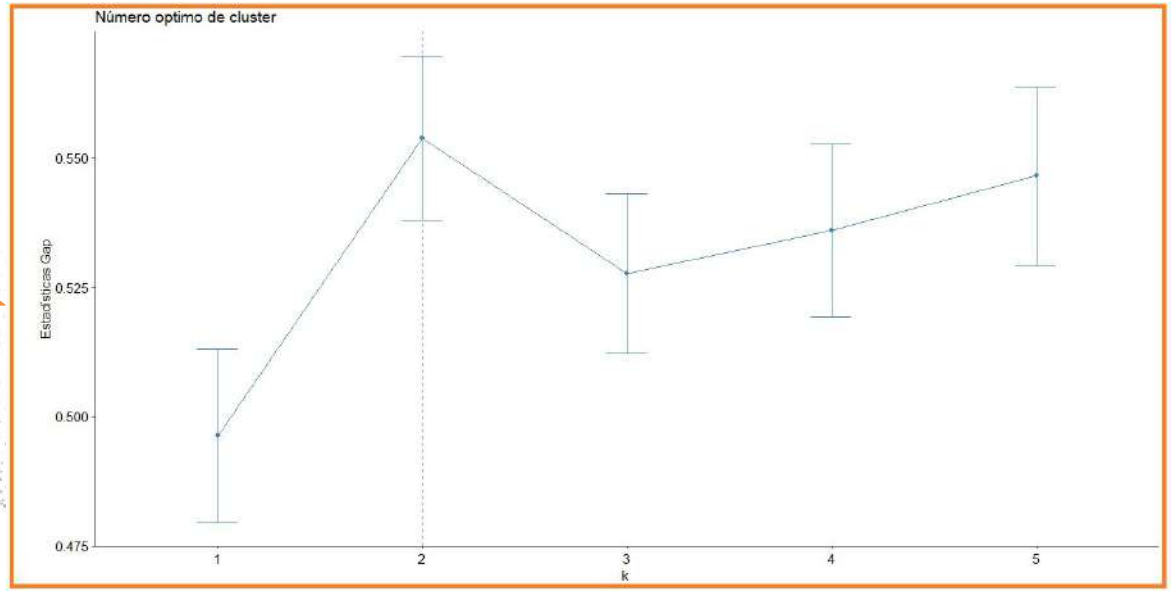
$$VQ = \frac{1}{2}\sum_{i=1}^{n} D(x_i, C_{f(x_i)})$$

Donde D es una función distancia y $f(x)$ devuelve el índice del clúster en la instancia i

Clústering método k-means para medir nivel de gestión

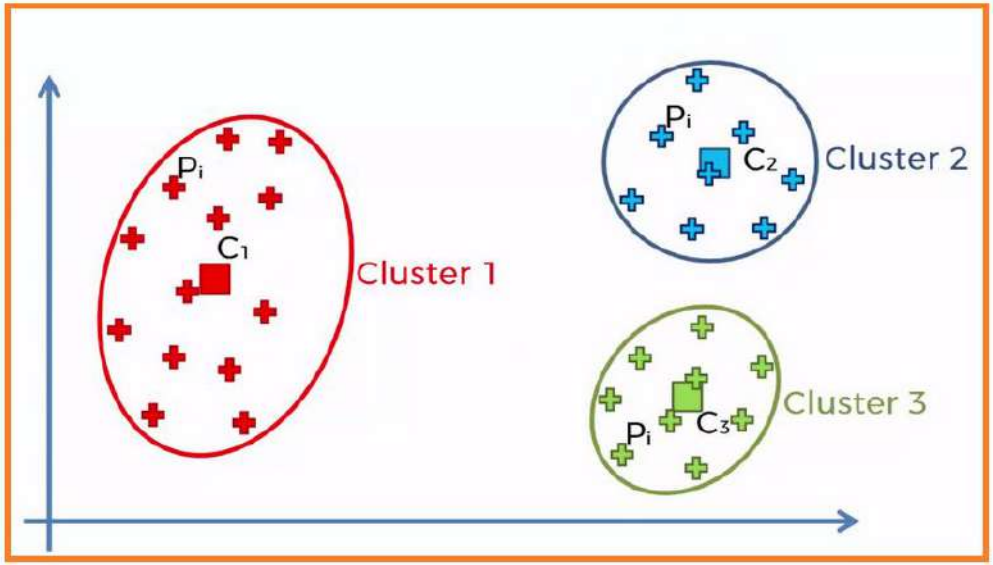
Calculamos el k óptimo por Bootstrapping

```
#=============================================================================
#Calculo del óptimo método de Bootstrapping
#=============================================================================
fviz_nbclust(datos_tp, kmeans, method = "gap_stat", k.max = 5,
             diss = get_dist(datos_tp, method = "euclidean"), nstart = 100)+
  labs(title= "Número optimo de cluster") +
  xlab("k ") +
  ylab("Estadísticas Gap")
#=============================================================================
```



Clústering método k-means para medir nivel de gestión

Calculamos el k óptimo por **Método de Elbow**: El método del codo se usa con el error cuadrático (sse) o la suma cuadrática del grupo (wcss). Para nuestro caso, usaré wcss para encontrar la cantidad óptima de clústeres. El siguiente código es una manera fácil de obtener el valor de wcss para diferentes números de clústeres:



Matemáticamente

$$WCSS = \sum_{P_i \text{ in Cluster } 1} distancia\ (P_i, C_1)^2 + \sum_{P_i \text{ in Cluster } 2} distancia\ (P_i, C_2)^2 + \sum_{P_i \text{ in Cluster } 3} distancia\ (P_i, C_3)^2$$

Clústering método k-means para medir nivel de gestión

Calculamos la suma cuadrática de los grupos

```
#=======================================================
#Estimación de número de cluster:
#Esta función aplica el algoritmo kmeans y devuelve la suma total de
#cuadrados internos
#=======================================================
calcular_totwithinss <- function(n_clusters, datos_tp, iter.max=1000, nstart=50){
  cluster_kmeans <- kmeans(centers = n_clusters, x = datos_tp, iter.max = iter.max,
                           nstart = nstart)
  return(cluster_kmeans$tot.withinss)
}
#=======================================================
```

```
#=======================================================
#Se aplica esta función para diferentes valores de k
#=======================================================
library(purrr)
total_withinss <- map_dbl(.x = 1:15,
                          .f = calcular_totwithinss,
                          datos = datos_tp)
total_withinss
#=======================================================
```

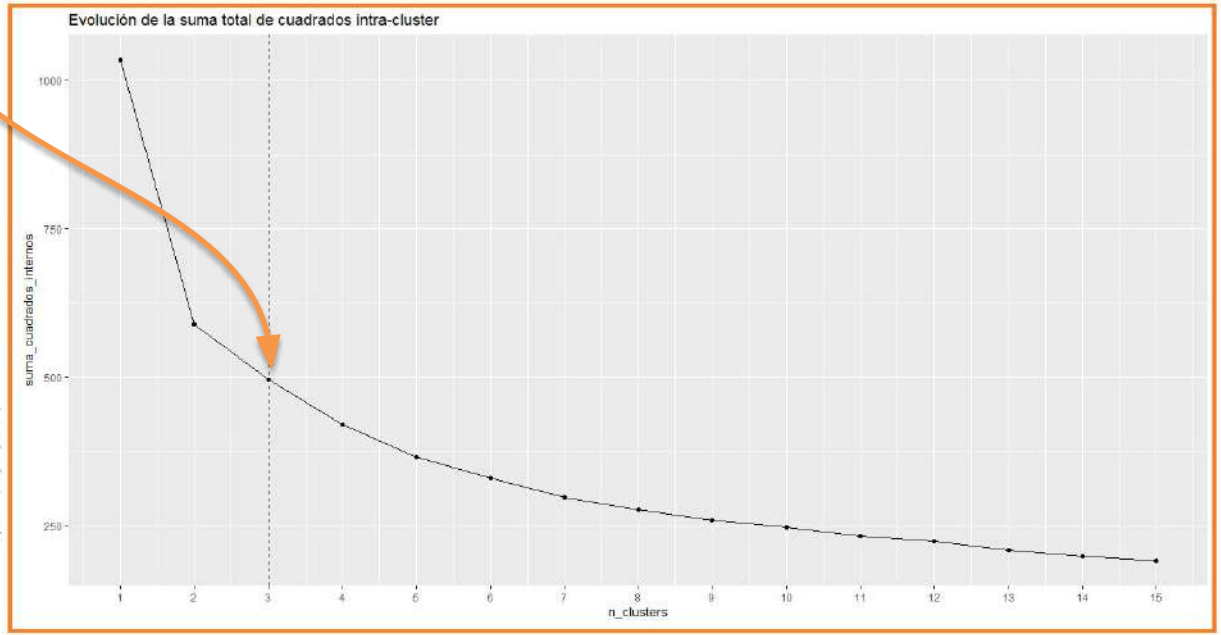
```
> total_withinss
 [1] 1035.0000  589.5185  495.5168  420.7814  365.5961  330.8595  298.0038  277.2757  258.9469  247.2791  229.8616
[12]  222.0422  211.2453  196.8524  187.9566
>
```

Clústering método k-means para medir nivel de gestión

Calculamos el número óptimo de cluster

```
#=====================================================
data.frame(n_clusters = 1:15, suma_cuadrados_internos = total_withinss) %>%
  ggplot(aes(x = n_clusters, y = suma_cuadrados_internos)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = 1:15) +
  labs(title = "Evolución de la suma total de cuadrados intra-cluster") +
  geom_vline(xintercept = 3, linetype = 2)
#=====================================================
```

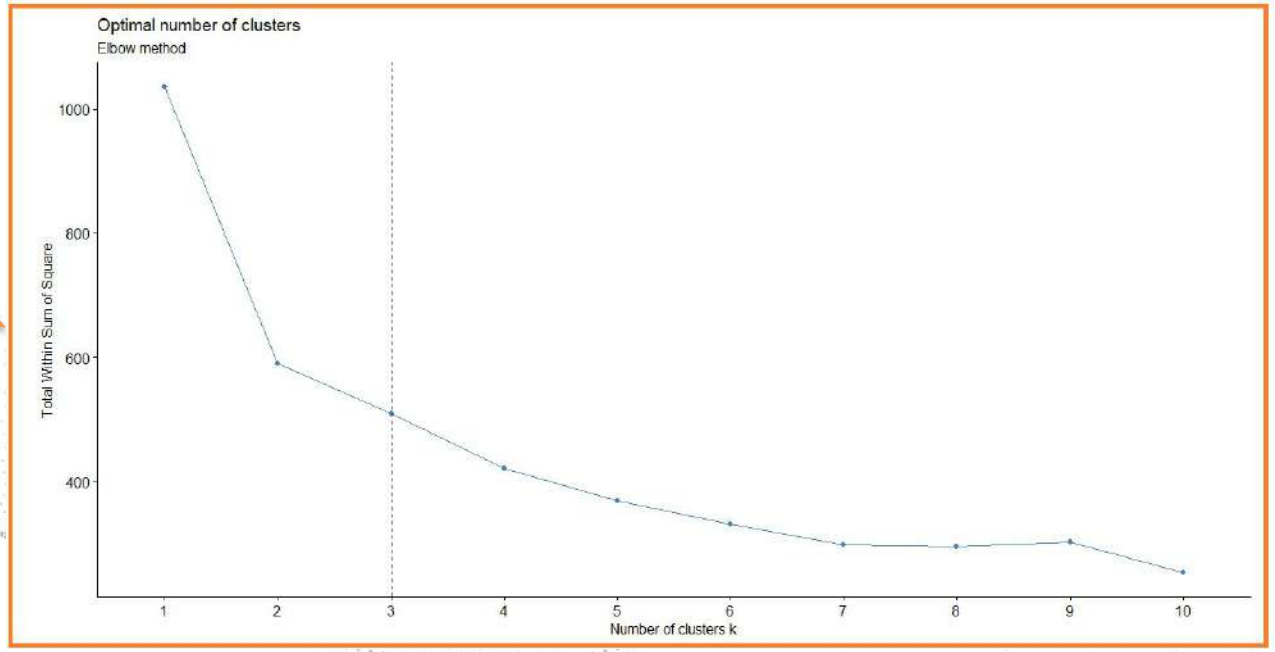
Por lo tanto, para k = 3, la relación suma de cuadrados entre grupos (wss) /total de la suma de cuadrados tiende a cambiar lentamente y permanece menos cambiante en comparación con otras k. ¿cómo decidimos cuál será la opción óptima? Así que miramos el segundo enfoque que viene con un nuevo paquete.



Clústering método k-means para medir nivel de gestión

Calculamos el k óptimo de por el método de Elbow

```
#=============================================================================
#Método de Elbow
#=============================================================================
fviz_nbclust(datos_tp, kmeans, method = "wss") +
  geom_vline(xintercept = 3, linetype = 2)+
  labs(subtitle = "Elbow method")
#=============================================================================
```



Clústering método k-means para medir nivel de gestión

Método de Jambú

Entonces, a fin de implementar el modelo de **K-Means**, comencemos por determinar la cantidad óptima de centroides a utilizar. Para ello, aplicaremos la función k- means al conjunto de datos, **variando en cada caso el valor de k**, y acumulando los valores de **WCSS** obtenidos, esta es una forma aproximada de obtener **el k-óptimo por el método de Jambú**.

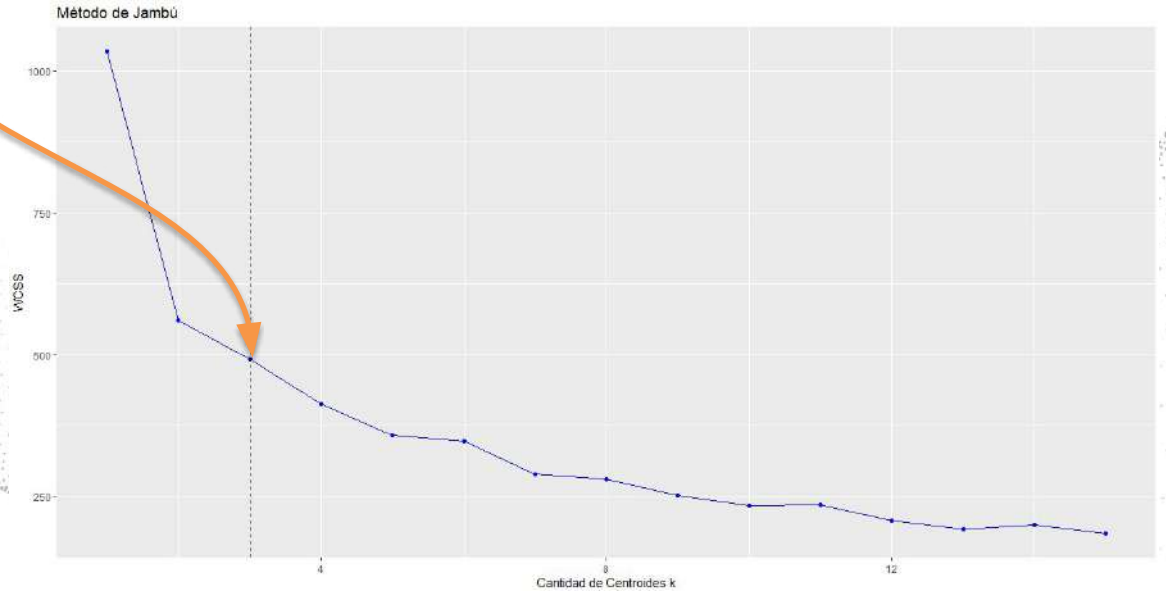
```
#=============================================================================
set.seed(1984)
wcss <- vector()
for(i in 1:15){
  wcss[i] <- sum(kmeans(datos_tp, i)$withinss)
}

ggplot() + geom_point(aes(x = 1:15, y = wcss), color = 'blue') +
  geom_line(aes(x = 1:15, y = wcss), color = 'blue') +
  ggtitle("Método de Jambú") +
  geom_vline(xintercept = 3, linetype = 2) +
  xlab('Cantidad de Centroides k') +
  ylab('WCSS')
#=============================================================================
```

Clústering método k-means para medir nivel de gestión

Calculamos la cantidad de centroides por método de Jambú

```
#=========================================================================
set.seed(1984)
wcss <- vector()
for(i in 1:15){
  wcss[i] <- sum(kmeans(datos_tp, i)$withinss)
}

ggplot() + geom_point(aes(x = 1:15, y = wcss), color = 'blue') +
  geom_line(aes(x = 1:15, y = wcss), color = 'blue') +
  ggtitle("Método de Jambú") +
  geom_vline(xintercept = 3, linetype = 2) +
  xlab('Cantidad de Centroides k') +
  ylab('WCSS')
#=========================================================================
```

Se observa que a partir de k= 3 en el método del Elbow y en el método de Jambú la suma total de cuadrados entre clúster se reduce, lo que equivale a decir reducción en la suma total de cuadrados internos dando como resultado que se estabiliza indicando que k = 3 es una buena opción.



Clústering método k-means para medir nivel de gestión

Implementación del método K-means, con k=3

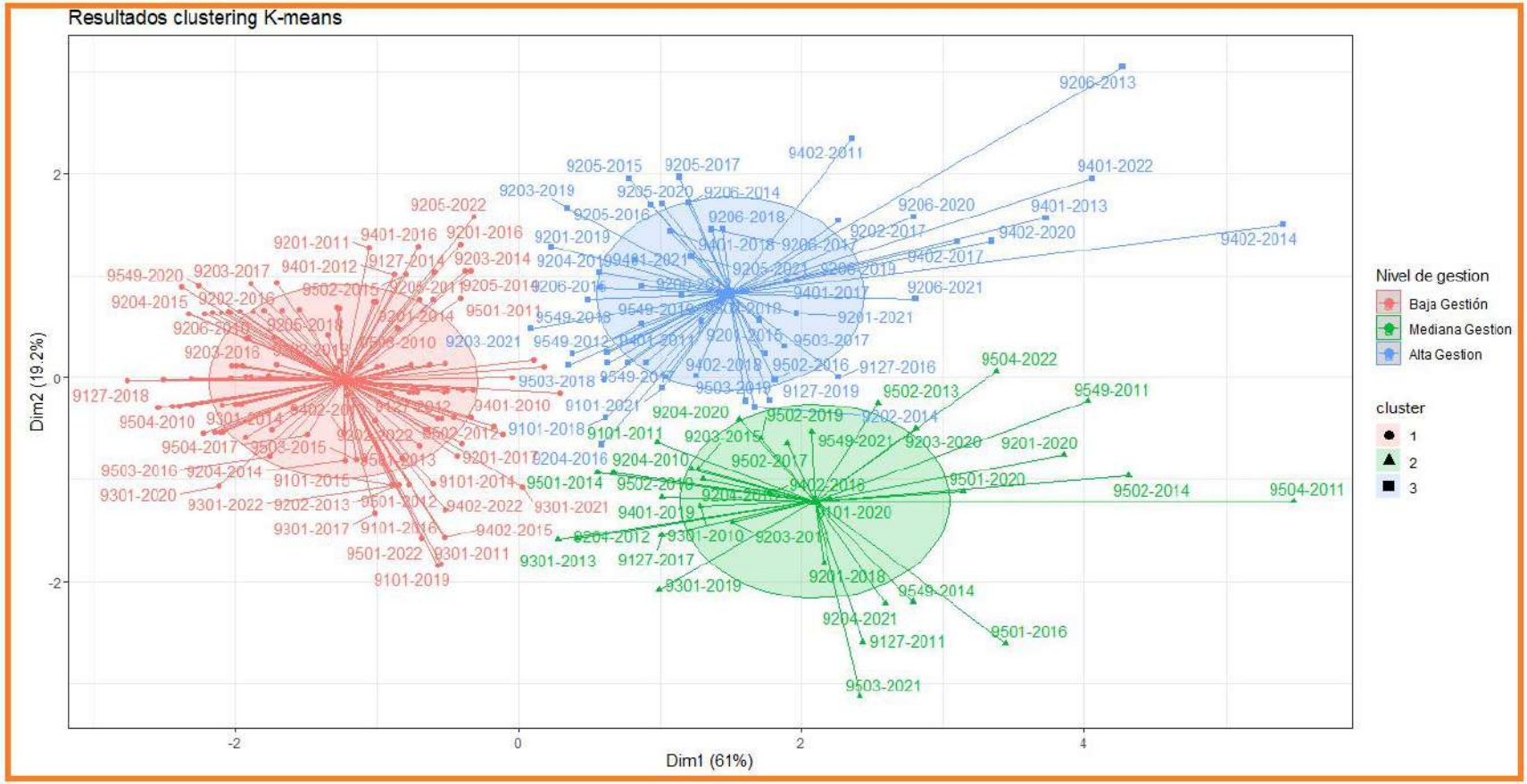
```
#=============================================================================
set.seed(123) #Establecimiento de semilla
km_clusters <- kmeans(x = datos_tp, centers = 3, iter.max = 16 , nstart = 100)
#Tamaño de observaciones en cada cluster
km_clusters$size
# [1]  121  33  54
#=============================================================================
```

El paquete **factoextra** también permite obtener visualizaciones de las agrupaciones resultantes. Si el número de variables (**dimensionalidad**) es mayor de 2, **automáticamente realiza un PCA** y **representa las dos primeras componentes principales**

```
#=============================================================================
#Clustering mediante el método de k-means
#=============================================================================
nCluster <- c(1,2,3)
nColor <- c("red", "blue", "forestgreen")
fviz_cluster(object = km_clusters, data = datos_tp,
             show.clust.cent = TRUE,
             #ellipse.type = "convex",  # Dibuja Poligono
             ellipse.type = "euclid", # Dibuja Elipse
             star.plot = TRUE,
             repel = TRUE, # Evita solapar etiquetas
             cex = 0.1) +
  labs(title = "Resultados clustering K-means")
#=============================================================================
```

Clústering método k-means para medir nivel de gestión

Clúster formados por el método K-means



Análisis de componentes principales

Mediante el análisis de componentes principales aplicado al método k-means es posible evidenciar la representatividad de las variables de interés que permiten medir el nivel de gestión por clúster, como se evidencia a continuación parte del código:

```
#=============================================================================
set.seed(123)
pca_1 <- prcomp(datos_tp)
df.pca <- pca_1$x
view(df.pca)
kc <- kmeans(df.pca[,1:2], centers = 3, iter.max = 16 , nstart = 100)
kc$size
# [1] 122  39  47
view(kc$cluster)
#=============================================================================
```

Filter

	Centro_Año	PC1	PC2	Cluster
9101-2010	9101-2010	-1.95918713	-0.259175836	1
9127-2010	9127-2010	-0.94916125	-0.155671705	1
9201-2010	9201-2010	-1.28280813	0.688653291	1
9202-2010	9202-2010	0.62139168	0.245540635	3
9203-2010	9203-2010	-1.12114712	0.105906055	1
9204-2010	9204-2010	1.22120081	-0.896759510	2
9205-2010	9205-2010	-1.26485059	0.687010409	1
9206-2010	9206-2010	-1.91377449	0.389999355	1
9301-2010	9301-2010	1.28614513	-1.267902051	2
9401-2010	9401-2010	0.28859707	-0.155334450	1
9402-2010	9402-2010	-0.03940292	0.001072117	1
9501-2010	9501-2010	-0.52320899	-0.137190666	1
9502-2010	9502-2010	0.67402428	-0.928805671	2
9503-2010	9503-2010	-0.85252036	0.468444408	1
9504-2010	9504-2010	-2.22578474	-0.537329038	1

Showing 1 to 16 of 208 entries, 4 total columns

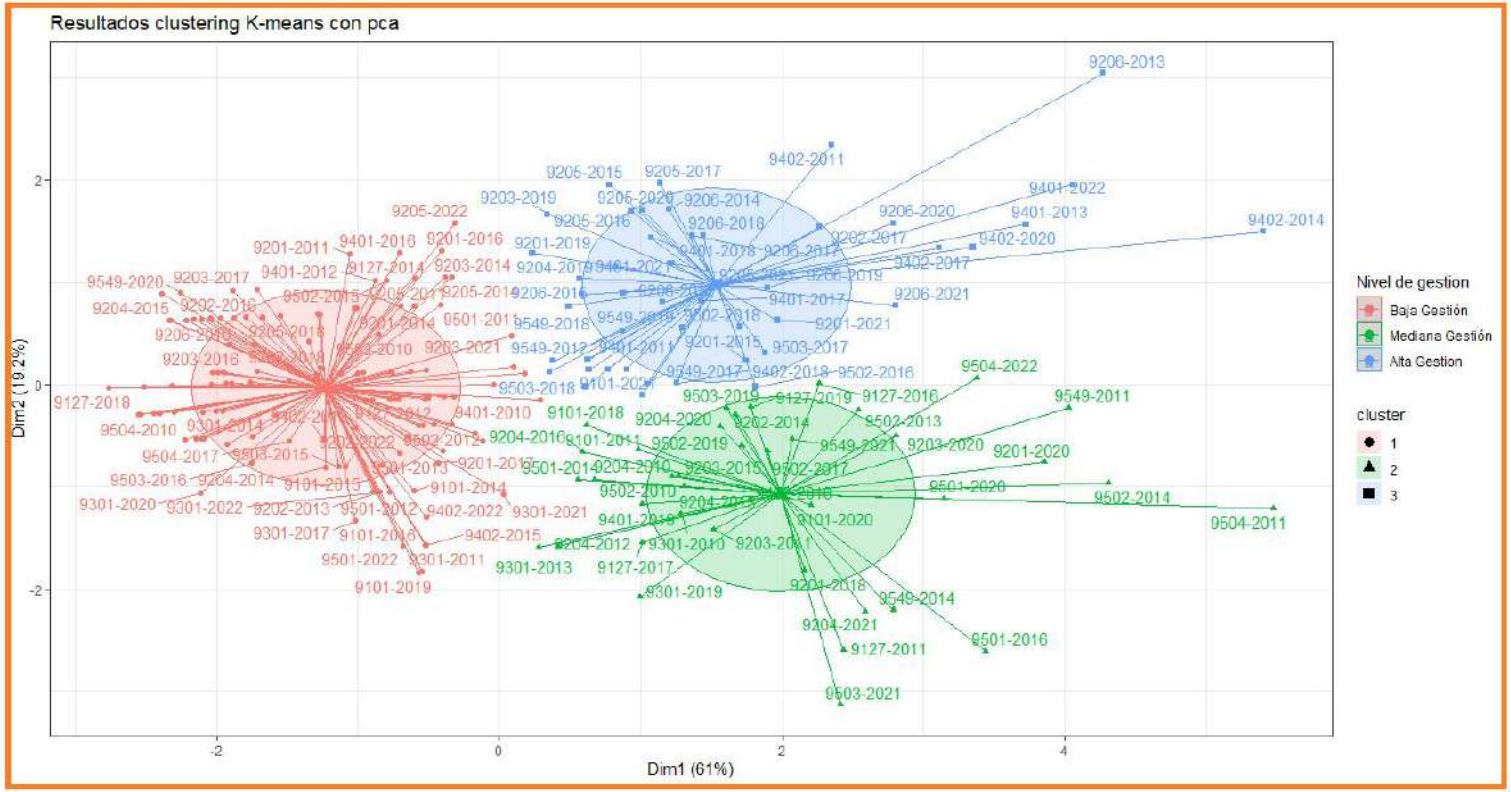
Aplicación de clustering mediante PCA

A continuación se evidencia la aplicación de clustering mediante el método de k-means y PCA

```
#=============================================================================
#Clustering mediante el método de k-means y pca
#=============================================================================

#=============================================================================
nCluster <- c(1,2,3)
nColor <- c("red", "blue", "forestgreen")
fviz_cluster(object = kc, data = datos_tp,
             show.clust.cent = TRUE,
             #ellipse.type = "convex",  # Dibuja Poligono
             ellipse.type = "euclid", # Dibuja Elipse
             star.plot = TRUE,
             repel = TRUE, # Evita solapar etiquetas
             cex = 0.1) +
  labs(title = "Resultados clustering K-means con pca") +
  scale_color_discrete("Nivel de gestion",labels = c("Baja Gestión", "Mediana Gestión", "Alta Gestion")) +
  theme_bw()
#=============================================================================
```

Aplicación de clustering mediante PCA

A continuación se evidencia la aplicación de clustering mediante el método de k-means y PCA



Plano factorial de pca para generar la gráfica de nivel de gestión por centros

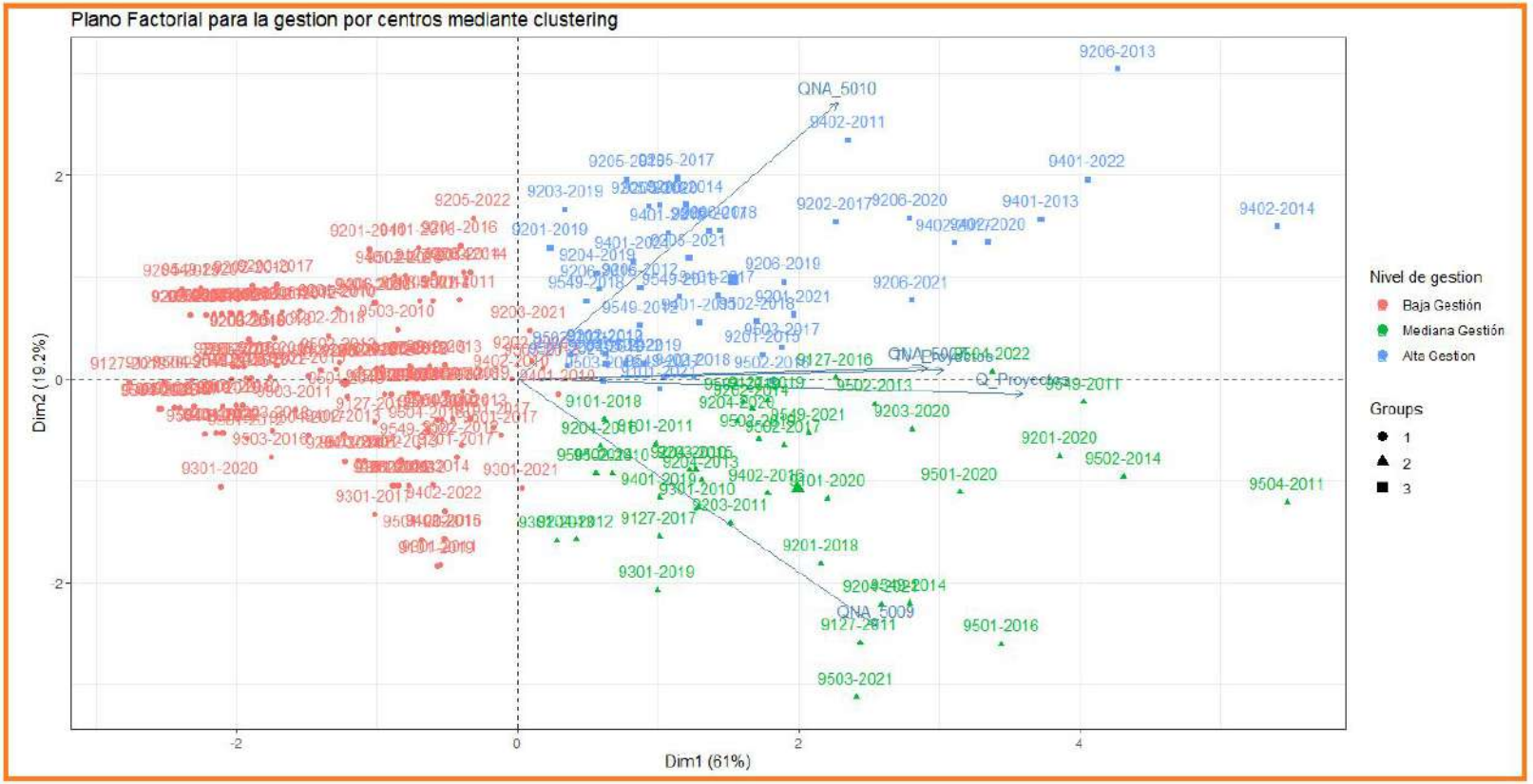
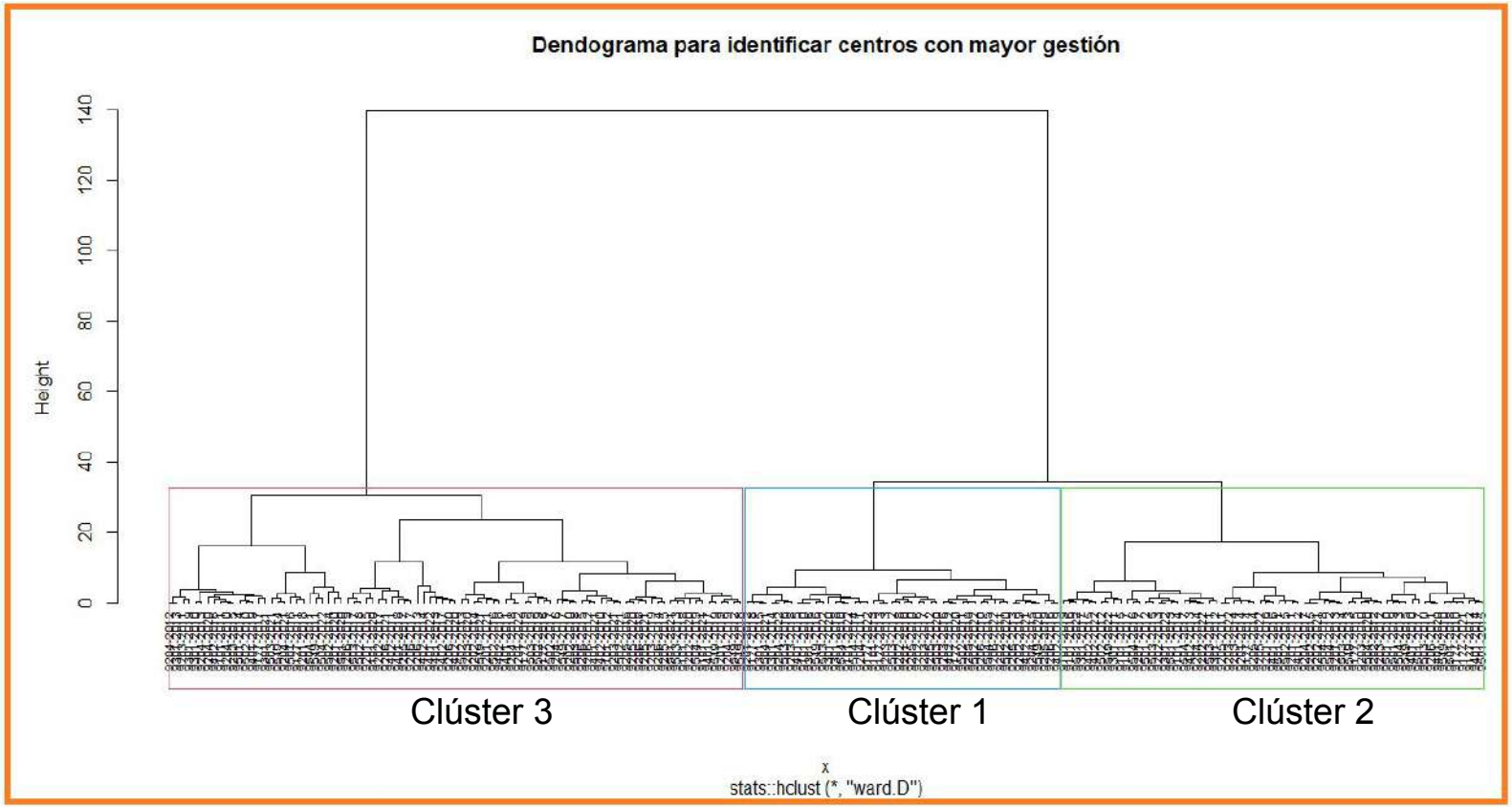


Gráfico de dos componentes principales para identificar comportamiento del nivel de gestión por centros

Implementación de clustering jerárquico




```
> hc.cut$size
[1] 50 67 91
>
```

Visualización de la tabla de procesos clasificada por nivel de gestión

Filter

	Centro_Año	Q_Proyectos	QNA_5007	QNA_5009	QNA_5010	TV_Proyectos	Cluster	Gestion
9101-2010	9101-2010	1	1	2	0	465292449	1	Baja Gestion
9127-2010	9127-2010	2	2	4	1	392005181	1	Baja Gestion
9201-2010	9201-2010	1	4	1	1	455662749	1	Baja Gestion
9202-2010	9202-2010	3	5	5	2	407185228	3	Alta Gestion
9203-2010	9203-2010	2	4	3	1	290617466	1	Baja Gestion
9204-2010	9204-2010	3	5	7	1	564001653	2	Media Gestion
9205-2010	9205-2010	1	3	1	1	508395371	1	Baja Gestion
9206-2010	9206-2010	1	3	2	1	300860419	1	Baja Gestion
9301-2010	9301-2010	3	6	6	0	652146153	2	Media Gestion
9401-2010	9401-2010	3	2	4	1	584555799	1	Baja Gestion
9402-2010	9402-2010	2	2	6	2	469064097	1	Baja Gestion
9501-2010	9501-2010	2	1	4	1	550483170	1	Baja Gestion
9502-2010	9502-2010	3	3	7	1	518339355	2	Media Gestion
9503-2010	9503-2010	2	1	4	2	385085638	1	Baja Gestion
9504-2010	9504-2010	1	1	3	0	356850041	1	Baja Gestion
9549-2010	9549-2010	2	1	3	1	362870121	1	Baja Gestion
9101-2011	9101-2011	3	5	6	1	541412297	2	Media Gestion

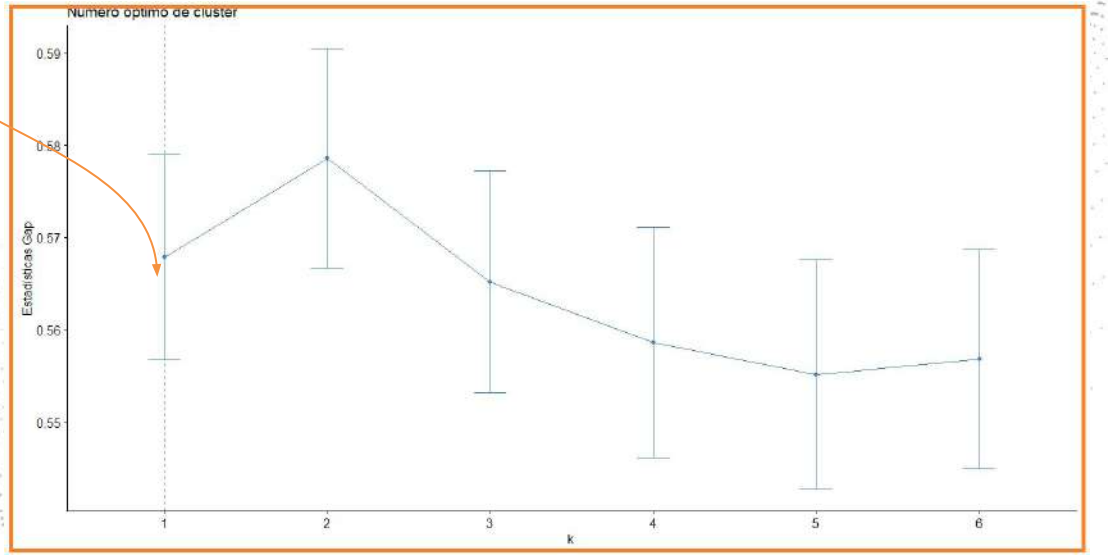
Showing 1 to 17 of 208 entries, 8 total columns

GC-F-004v.05

4. Modelo de analítica de datos para medición de productividad por centros

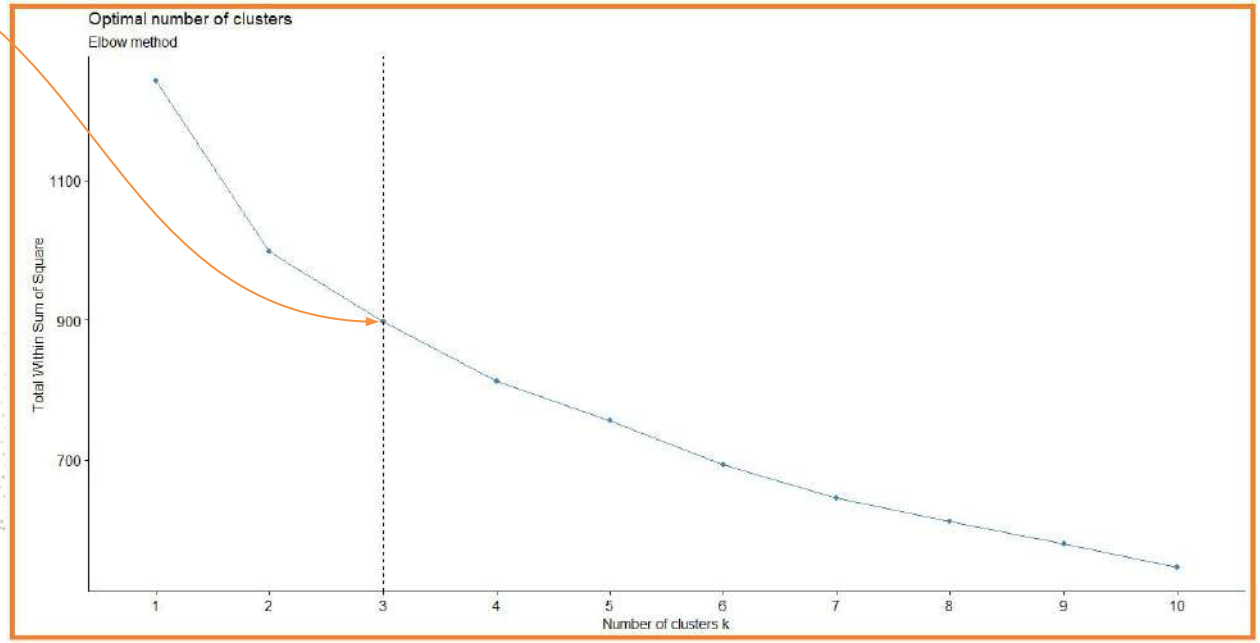
Calculo del número óptimo de clúster por método de Bootstrapping

```
#=============================================================================
#Calculo del óptimo método de Bootsrapping
#=============================================================================
fviz_nbclust(datos_tppd, kmeans, method = "gap_stat", k.max = 6,
             diss = get_dist(datos_tppd, method = "euclidean"), nstart = 100)+
  labs(title= "Número optimo de cluster") +
  xlab("k ") +
  ylab("Estadísticas Gap")
#=============================================================================
```



Calculo del número óptimo de clúster por método de Elbow

```
#=============================================================================
#Método de Elbow
#=============================================================================
fviz_nbclust(datos_tppd, kmeans, method = "wss") +
  geom_vline(xintercept = 3, linetype = 2)+
  labs(subtitle = "Elbow method")
#=============================================================================
```



Estimación del número óptimo de cluster

La siguiente función aplica el algoritmo k - means y devuelve la suma total de cuadrados internos

```
#=========================================================
calcular_totwithinss <- function(n_clusters, datos_tppd, iter.max=1000, nstart=50){
  cluster_kmeans <- kmeans(centers = n_clusters, x = datos_tppd, iter.max = iter.max,
                           nstart = nstart)
  return(cluster_kmeans$tot.withinss)
}
#=========================================================
```

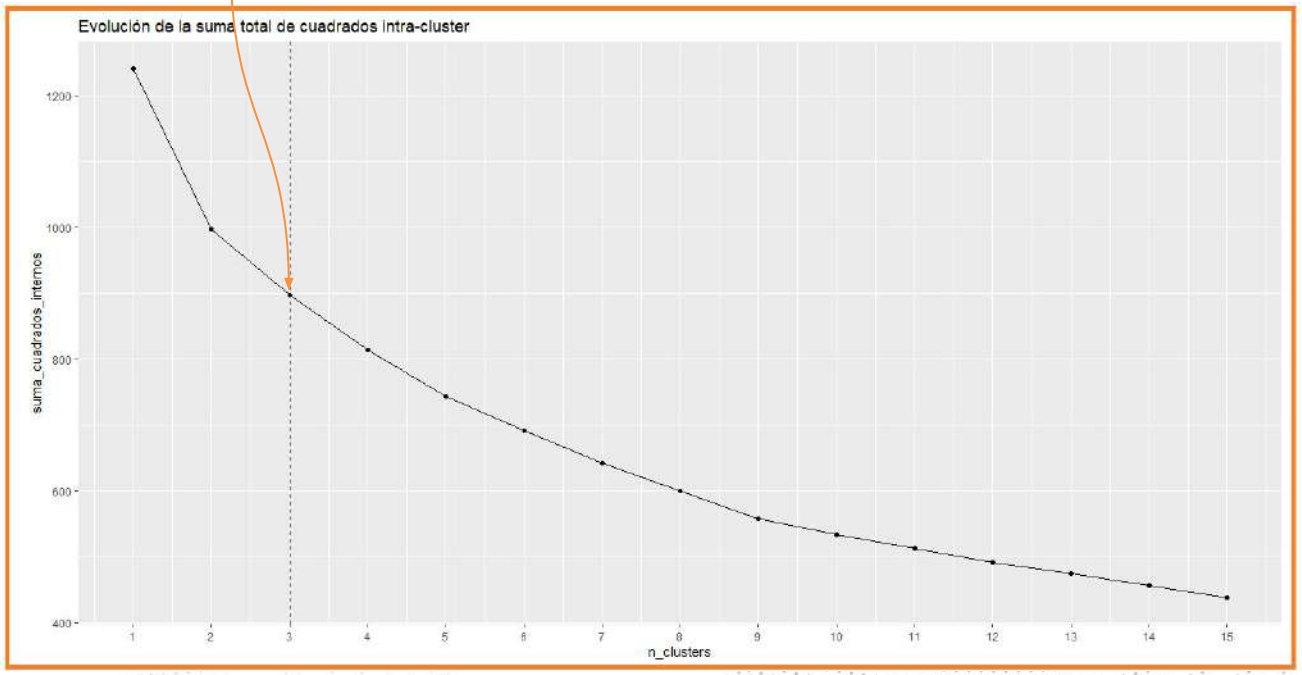
Se aplica la función para diferentes valores de k

```
#=========================================================
library(purrr)
total_withinss <- map_dbl(.x = 1:15,
                          .f = calcular_totwithinss,
                          datos = datos_tppd)
total_withinss
#=========================================================
```

```
> total_withinss
 [1] 1242.0000  998.1499  897.6900  813.7018  744.4412  691.0692  642.3362  600.4098  558.2706  534.3194  513.1994
[12]  491.6664  473.9465  456.0774  437.2828
>
```

Estimación del número óptimo de clúster

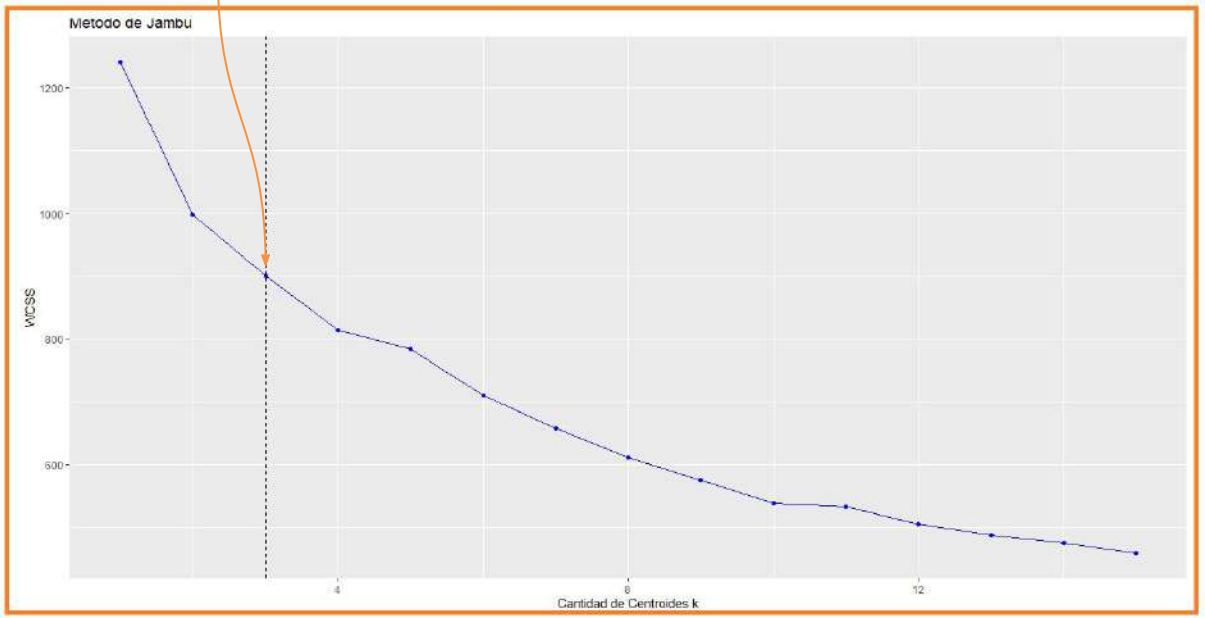
```r
#=====================================================
data.frame(n_clusters = 1:15, suma_cuadrados_internos = total_withinss) %>%
  ggplot(aes(x = n_clusters, y = suma_cuadrados_internos)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = 1:15) +
  labs(title = "Evolución de la suma total de cuadrados intra-cluster") +
  geom_vline(xintercept = 3, linetype = 2)
#=====================================================
```



Calculo del número óptimo de clúster por método de Jambú

```
#=========================================================================
set.seed(1984)
wcss <- vector()
for(i in 1:15){
  wcss[i] <- sum(kmeans(datos_tppd, i)$withinss)
}

ggplot() + geom_point(aes(x = 1:15, y = wcss), color = 'blue') +
  geom_line(aes(x = 1:15, y = wcss), color = 'blue') +
  ggtitle("Método de Jambú") +
  geom_vline(xintercept = 3, linetype = 2) +
  xlab('cantidad de centroides k') +
  ylab('WCSS')
#=========================================================================
```

Se observa que a partir de 3 en el método del Elbow y en el método de Jambú la suma total de cuadrados entre clúster se reduce a partir de 3, lo que equivale a decir reducción en la suma total de cuadrados internos se estabiliza indicando que k = 3

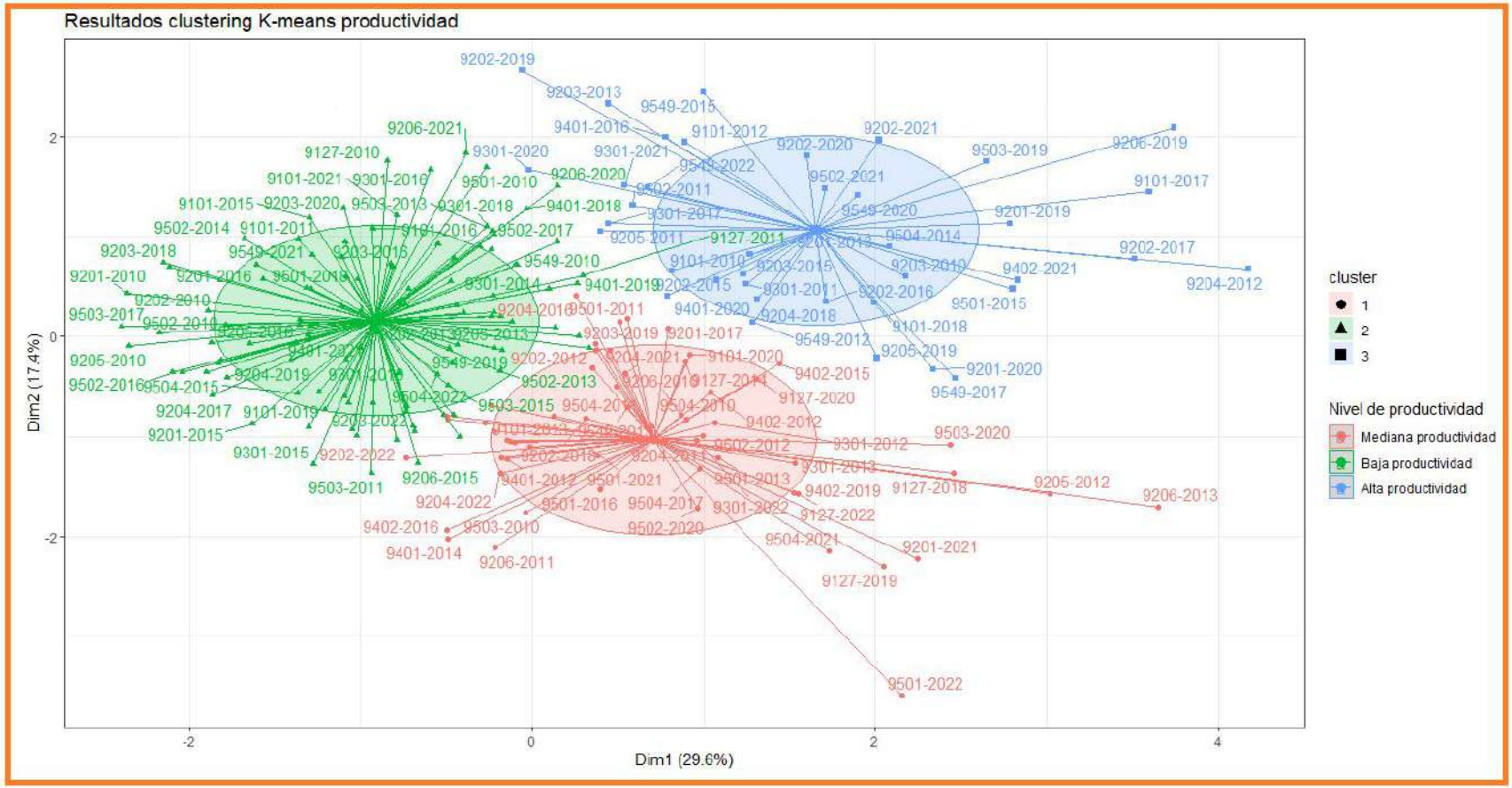


Implementación de la función k - means en R, con k=3

```
#=============================================================================
set.seed(123) #Establecimiento de semilla
km_clusters <- kmeans(x = datos_tppd, centers = 3, iter.max = 16 , nstart = 100)
#Tamaño de observaciones en cada cluster
km_clusters$size
# [1]  57 113  38
#=============================================================================
```

```
#=============================================================================
#Clustering mediante el método de k-means
#=============================================================================
nCluster <- c(1,2,3)
nColor <- c("red", "blue", "forestgreen")
fviz_cluster(object = km_clusters, data = datos_tppd,
             show.clust.cent = TRUE,
             #ellipse.type = "convex",  # Dibuja Poligono
             ellipse.type = "euclid", # Dibuja Elipse
             star.plot = TRUE,
             repel = TRUE, # Evita solapar etiquetas
             cex = 0.1) +
  labs(title = "Resultados clustering K-means productividad") +
  scale_color_discrete("Nivel de productividad",labels = c("Mediana productividad", "Baja productividad", "Alta
productividad")) +
  theme_bw()
#=============================================================================
```

Implementación de la función k - means en R, con k=3



Análisis de componentes principales para medir representatividad

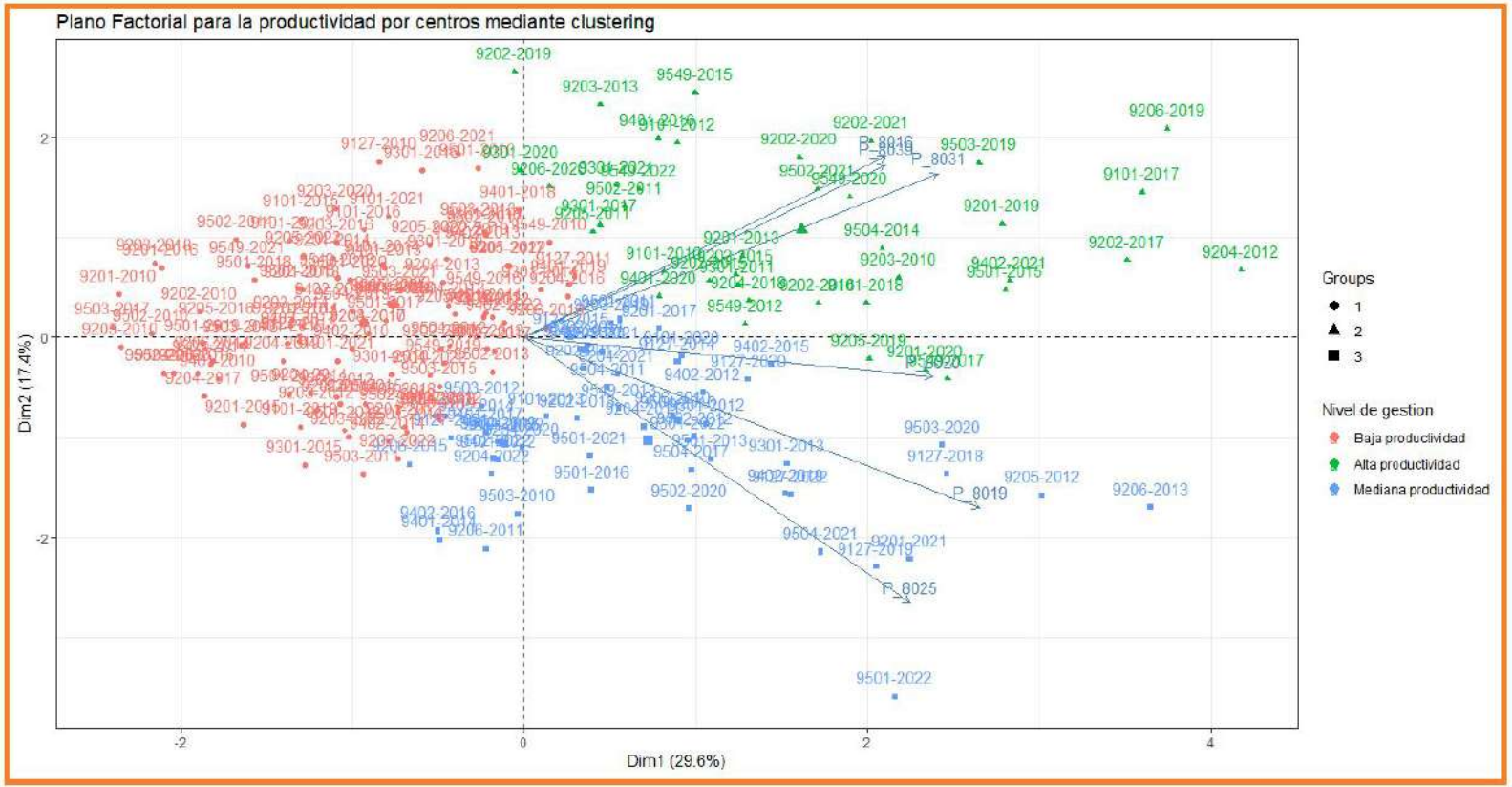
```
#=============================================================================
#Análisis de componentes principales para medir representatividad
#=============================================================================


#=============================================================================
set.seed(123)
pca_1 <- prcomp(datos_tppd)
df.pca <- pca_1$x
view(df.pca)
kc <- kmeans(df.pca[,1:2], centers = 3, iter.max = 16 , nstart = 100)
kc$size
# [1] 112  39  57
view(kc$cluster)
#=============================================================================
```

Filter

	Centro_Año	PC1	PC2	Cluster
9101-2010	9101-2010	0.81427935	0.660407369	2
9127-2010	9127-2010	-0.84504828	1.755171158	1
9201-2010	9201-2010	-2.36025093	0.430049256	1
9202-2010	9202-2010	-1.89369836	0.255678095	1
9203-2010	9203-2010	2.18003828	0.599621826	2
9204-2010	9204-2010	-0.90663781	0.034679520	1
9205-2010	9205-2010	-2.35187370	-0.096050084	1
9206-2010	9206-2010	0.86524729	-0.784240926	3
9301-2010	9301-2010	-0.77547270	-0.363745343	1
9401-2010	9401-2010	-1.77994793	-0.411196868	1
9402-2010	9402-2010	-1.00201316	-0.110213480	1
9501-2010	9501-2010	-0.26864288	1.690543796	1
9502-2010	9502-2010	-2.17213824	0.037081642	1
9503-2010	9503-2010	-0.03824655	-1.763277201	3

Showing 1 to 14 of 208 entries, 4 total columns

Plano factorial de pca para generar la gráfica de nivel de productividad por centros



Clustering mediante el método de k-means y pca

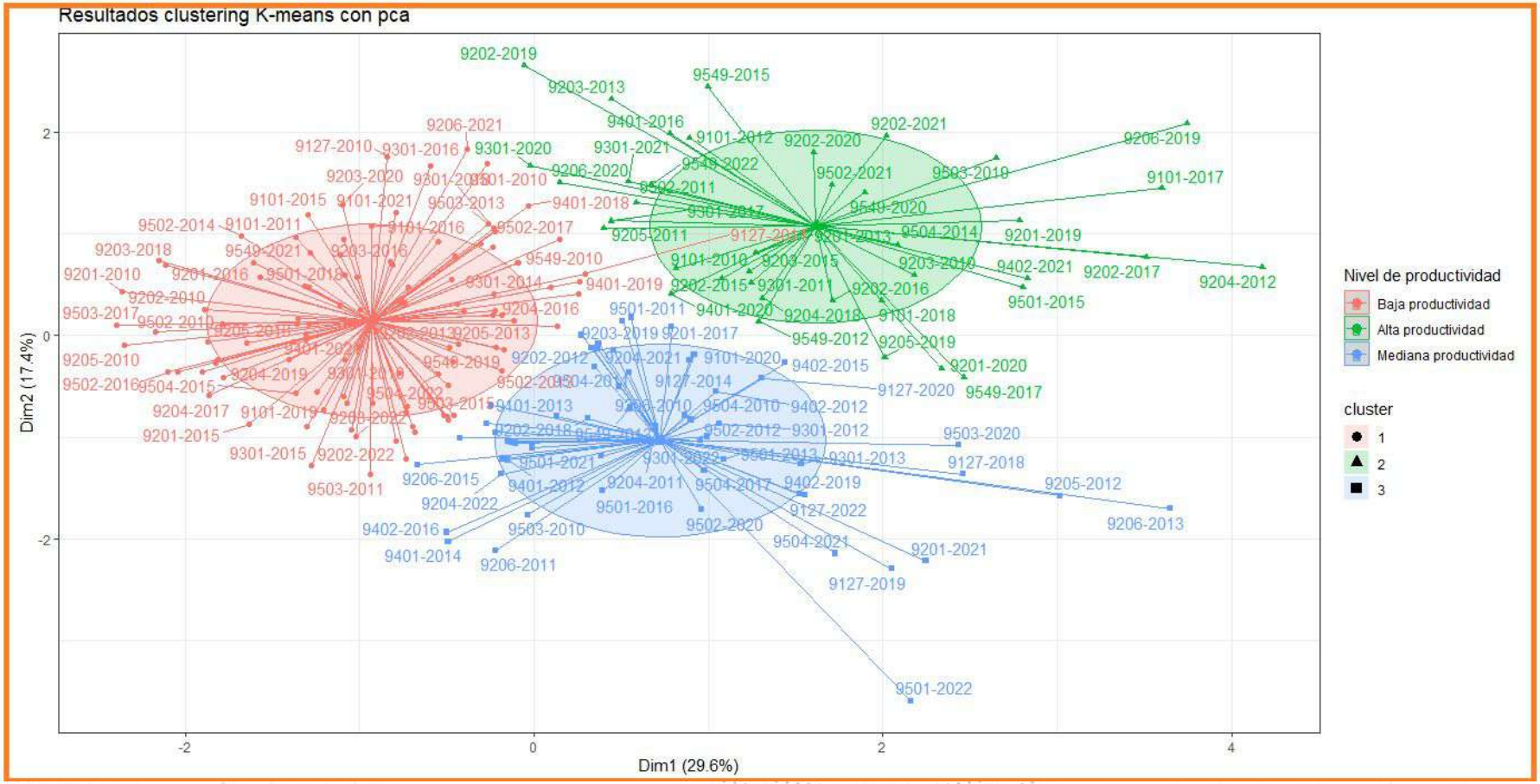
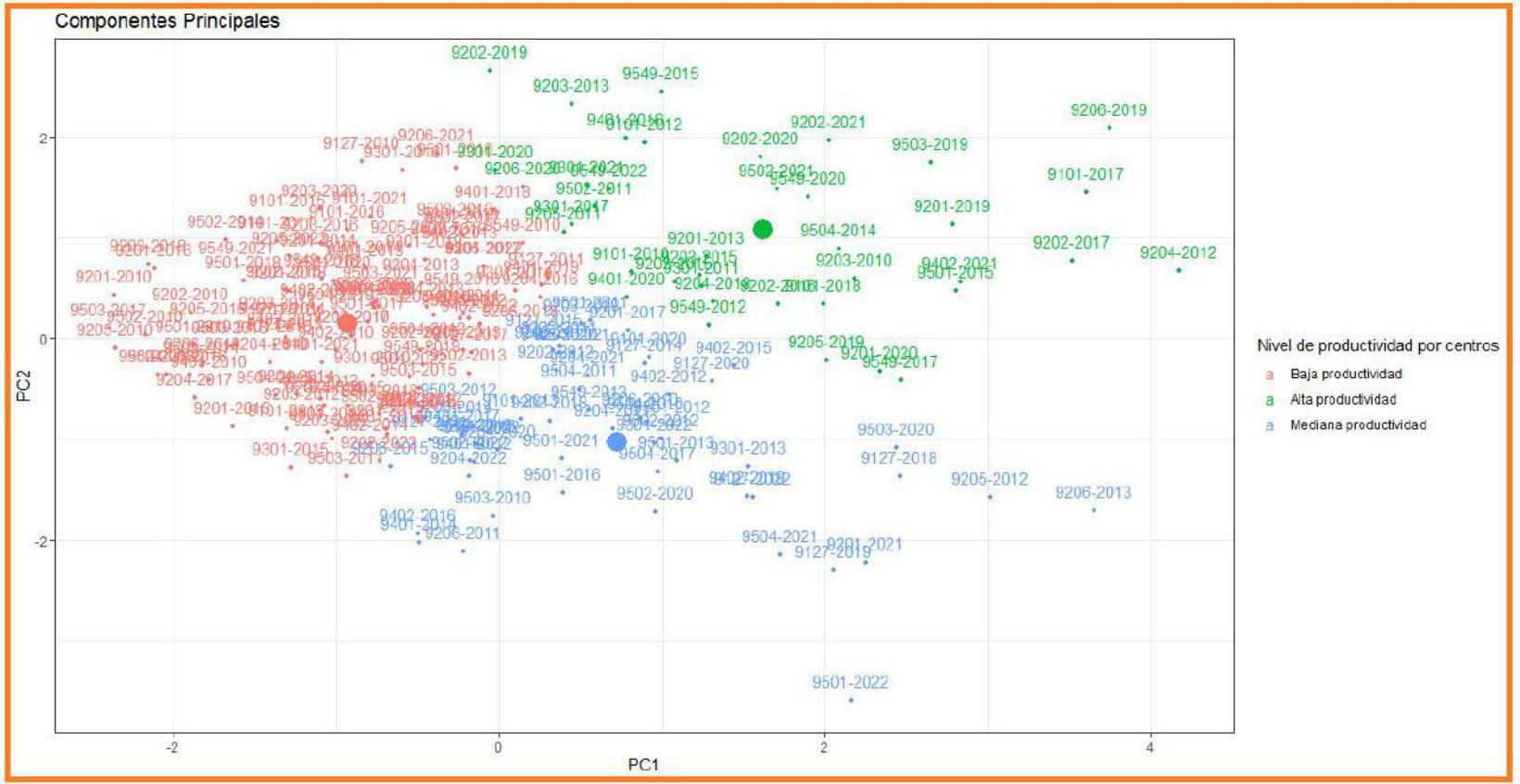
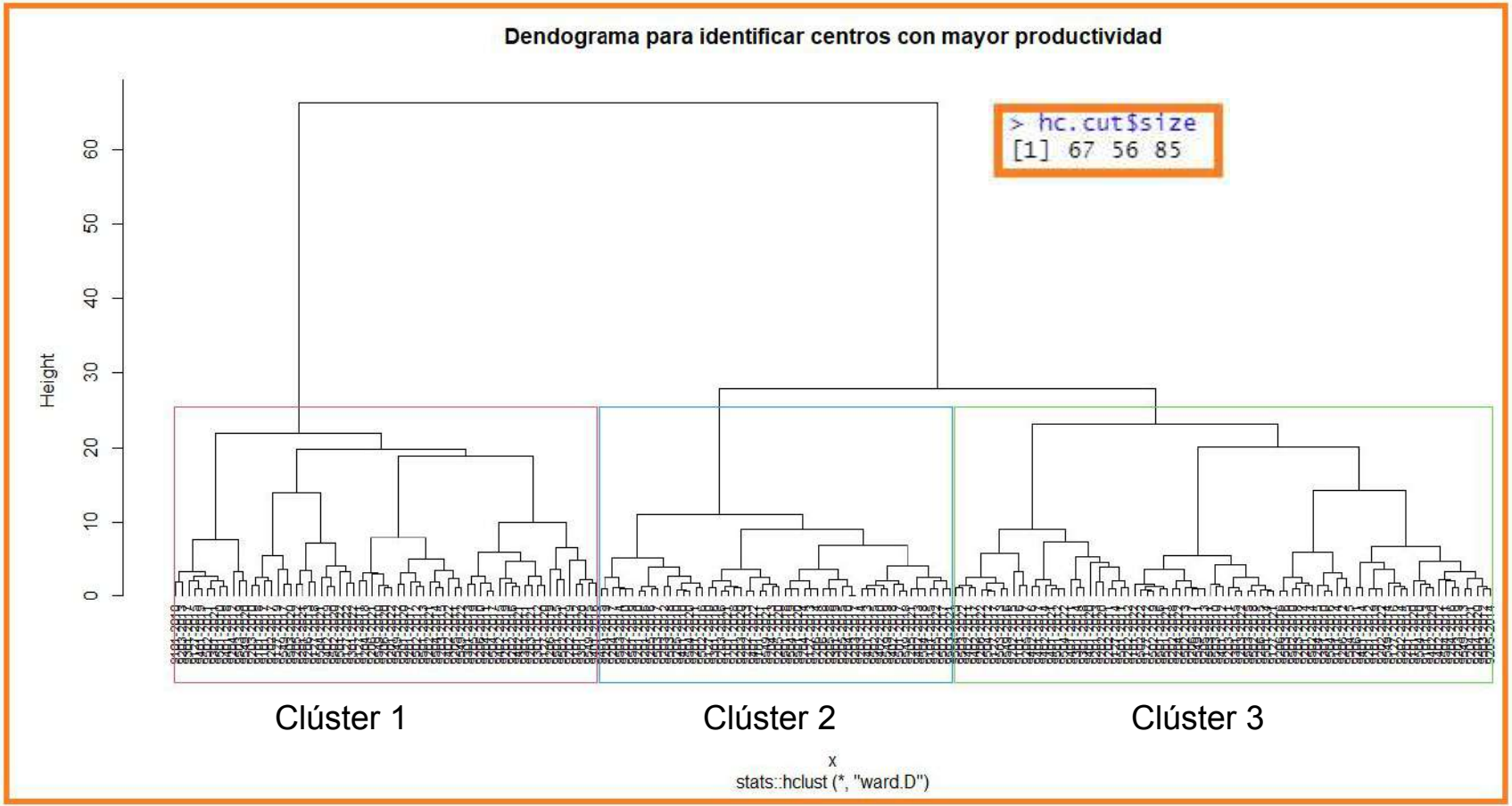


Gráfico de dos componentes principales para identificar comportamiento del nivel de productividad por centros



Implementación de clustering jerárquico



GC-F-004v.05

5. Modelo de analítica de datos para medición de gestión por centros – Algoritmo KNN

Contenido

1. Base de datos procesos
2. Base de datos productividad
3. Modelo de analítica de datos para medición de gestión por centros
   - Que es el algoritmo K-NN
   - Implementación del algoritmo K-NN en R statistics

GC-F-004v.05

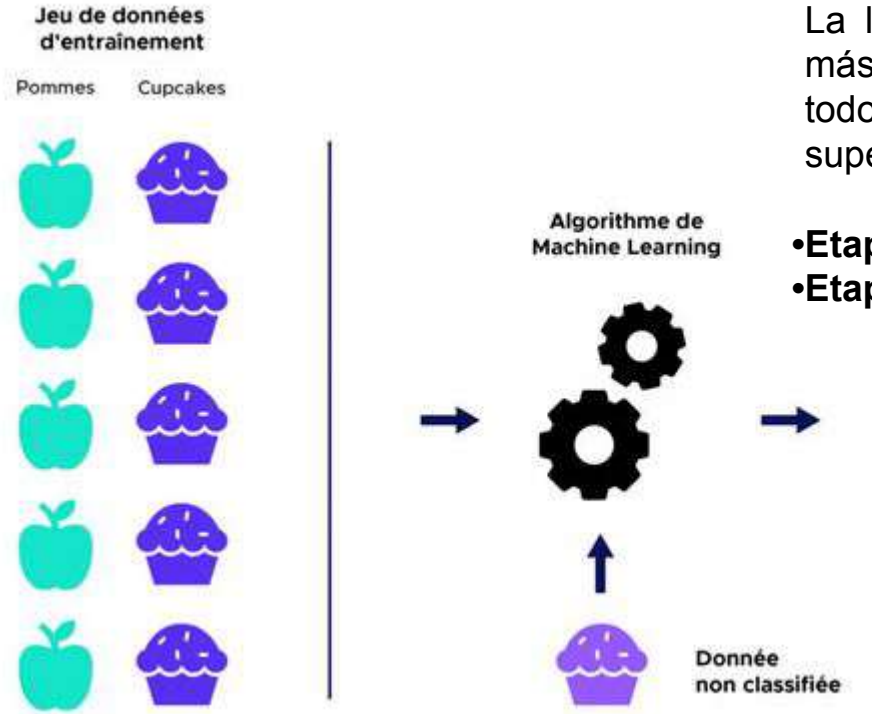
Que es el algoritmo K-NN

¿Qué es un algoritmo de aprendizaje supervisado?

En aprendizaje supervisado, un algoritmo recibe un conjunto de datos que están etiquetados con los valores de salida correspondientes sobre los que puede entrenarse y definir un modelo de predicción. Este algoritmo podrá luego ser utilizado sobre datos nuevos para predecir sus valores de salida correspondientes.

El algoritmo de las K vecinas más cercanas o K-nearest neighbors (kNN) es un algoritmo de Machine Learning que pertenece a los algoritmos de aprendizaje supervisado simples y fáciles de aplicar que pueden ser utilizados para resolver problemas de clasificación y de regresión. En este artículo veremos la definición de este algoritmo, su funcionamiento así como también una aplicación directa en programación.

Que es el algoritmo K-NN



La lógica detrás del algoritmo de las K vecinas más cercanas es una de las más sencillas de todos los algoritmos de Machine Learning supervisados:

- **Etapa 1:** Seleccionar el número de K vecinas
- **Etapa 2:** Calcular la distancia

$$\sum_{i=1}^{n} |x_i - y_i|$$

Euclidienne

$$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

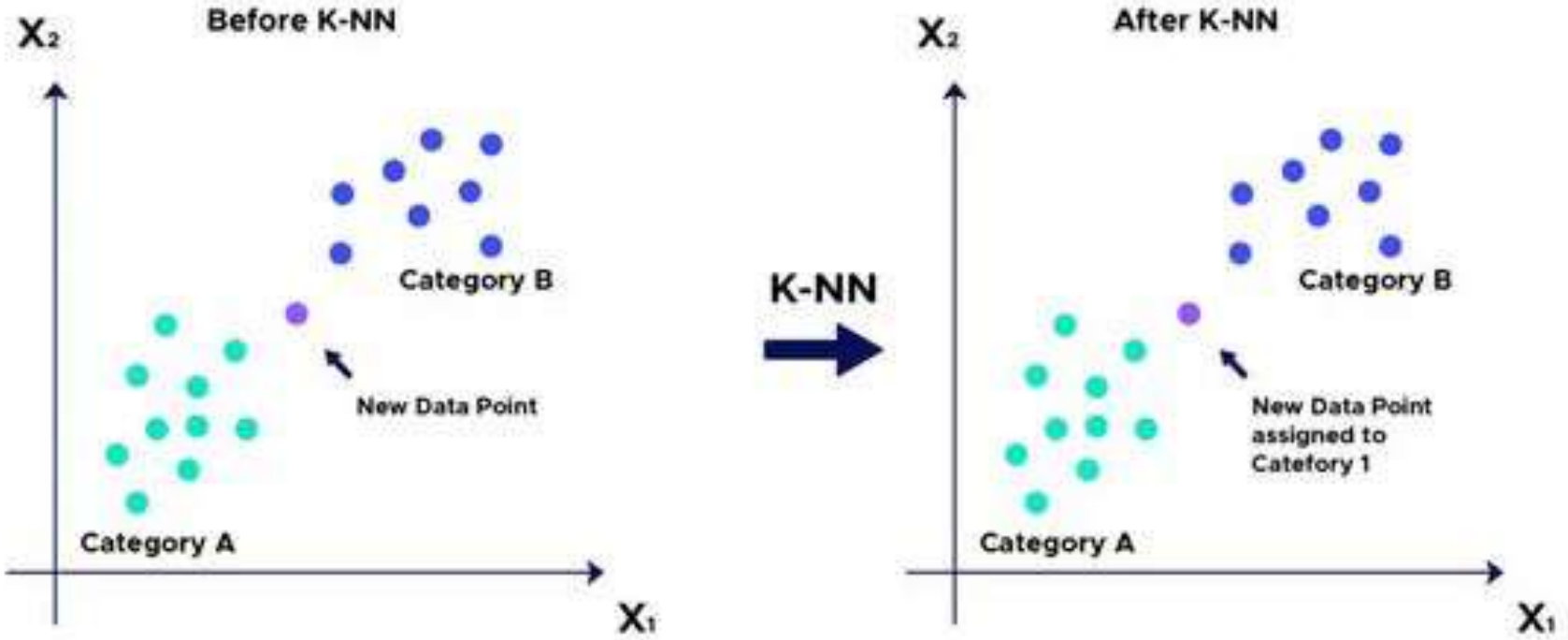
Manhattan

Que es el algoritmo K-NN

Desde un punto no clasificado a otros puntos:

- **Etapa 3:** tomar las K vecinas más cercanas según la distancia calculada
- **Etapa 4:** entre las K vecinas, contar el número de puntos en cada categoría.
- **Etapa 5:** atribuir un nuevo punto a la categoría más presente entre las K vecinas
- **Etapa 6:** El modelo está listo.

Que es el algoritmo K-NN



Implementación del algoritmo K-NN en R statistics

Lectura de base de datos Tabla_Pprocesos.xlsx

```
#=============================================================================
# Base de datos procesos: Tabla_PProceso.xlsx
#=============================================================================
setwd("E:/Documents/Codigo en R/SENA paola")
data_procesos <- read_xlsx("Tabla_PProceso.xlsx")
View(data_procesos)
#=============================================================================
```

Filter

	Centro_Año	Q_Proyectos	QNA_5007	QNA_5009	QNA_5010	TV_Proyectos	Cluster	Gestion
1	9101-2010	1	1	2	0	465292449	1	Baja Gestion
2	9127-2010	2	2	4	1	392005181	1	Baja Gestion
3	9201-2010	1	4	1	1	455662749	1	Baja Gestion
4	9202-2010	3	5	5	2	407185228	3	Alta Gestion
5	9203-2010	2	4	3	1	290617466	1	Baja Gestion
6	9204-2010	3	5	7	1	564001653	2	Media Gestion
7	9205-2010	1	3	1	1	508395371	1	Baja Gestion
8	9206-2010	1	3	2	1	300860419	1	Baja Gestion
9	9301-2010	3	6	6	0	652146153	2	Media Gestion
10	9401-2010	3	2	4	1	584555799	1	Baja Gestion
11	9402-2010	2	2	6	2	469064097	1	Baja Gestion
12	9501-2010	2	1	4	1	550483170	1	Baja Gestion
13	9502-2010	3	3	7	1	518339355	2	Media Gestion
14	9503-2010	2	1	4	2	385085638	1	Baja Gestion
15	9504-2010	1	1	3	0	356850041	1	Baja Gestion
16	9549-2010	2	1	3	1	362870121	1	Baja Gestion
17	9101-2011	3	5	6	1	541412297	2	Media Gestion

Showing 1 to 17 of 208 entries, 8 total columns

Implementación del algoritmo K-NN en R statistics

Exploración de los datos

```
#=============================================================================
# Primer paso: Exploración de los datos
#=============================================================================
summary(data_procesos)
str(data_procesos)
#=============================================================================
```

```
> summary(data_procesos)
  Centro_Año         Q_Proyectos       QNA_5007        QNA_5009        QNA_5010      TV_Proyectos          Cluster
 Length:208         Min.   :1.000   Min.   : 0.000   Min.   : 0.00   Min.   :0.000   Min.   :290617466   Min.   :1.000
 Class :character   1st Qu.:2.000   1st Qu.: 2.000   1st Qu.: 2.00   1st Qu.:0.000   1st Qu.:394601862   1st Qu.:1.000
 Mode  :character   Median :2.000   Median : 3.000   Median : 4.00   Median :1.000   Median :456991575   Median :1.000
                    Mean   :2.404   Mean   : 3.668   Mean   : 4.13   Mean   :1.255   Mean   :480505237   Mean   :1.639
                    3rd Qu.:3.000   3rd Qu.: 5.000   3rd Qu.: 6.00   3rd Qu.:2.000   3rd Qu.:541300336   3rd Qu.:2.000
                    Max.   :6.000   Max.   :14.000   Max.   :13.00   Max.   :5.000   Max.   :932664228   Max.   :3.000
   Gestion
 Length:208
 Class :character
 Mode  :character


> str(data_procesos)
tibble [208 × 8] (S3: tbl_df/tbl/data.frame)
 $ Centro_Año  : chr [1:208] "9101-2010" "9127-2010" "9201-2010" "9202-2010" ...
 $ Q_Proyectos : num [1:208] 1 2 1 3 2 3 1 1 3 3 ...
 $ QNA_5007    : num [1:208] 1 2 4 5 4 5 3 3 6 2 ...
 $ QNA_5009    : num [1:208] 2 4 1 5 3 7 1 2 6 4 ...
 $ QNA_5010    : num [1:208] 0 1 1 2 1 1 1 1 0 1 ...
 $ TV_Proyectos: num [1:208] 4.65e+08 3.92e+08 4.56e+08 4.07e+08 2.91e+08 ...
 $ Cluster     : num [1:208] 1 1 1 3 1 2 1 1 2 1 ...
 $ Gestion     : chr [1:208] "Baja Gestion" "Baja Gestion" "Baja Gestion" "Alta Gestion" ...
```

Implementación del algoritmo K-NN en R statistics

Se convierten los datos en un data.frame

```
#=============================================================================
# Segundo paso: convertimos la variable de clasificación Gestión en factor
# Con el propósito de identificarla por niveles, pero primero convertimos los datos
# en data.frame
#=============================================================================
data_procesos <- as.data.frame(data_procesos)
str(data_procesos)
#=============================================================================
```

```
> str(data_procesos)
'data.frame':   208 obs. of  8 variables:
 $ Centro_Año  : chr  "9101-2010" "9127-2010" "9201-2010" "9202-2010" ...
 $ Q_Proyectos : num  1 2 1 3 2 3 1 1 3 3 ...
 $ QNA_5007    : num  1 2 4 5 4 5 3 3 6 2 ...
 $ QNA_5009    : num  2 4 1 5 3 7 1 2 6 4 ...
 $ QNA_5010    : num  0 1 1 2 1 1 1 1 0 1 ...
 $ TV_Proyectos: num  4.65e+08 3.92e+08 4.56e+08 4.07e+08 2.91e+08 ...
 $ Cluster     : num  1 1 1 3 1 2 1 1 2 1 ...
 $ Gestion     : chr  "Baja Gestion" "Baja Gestion" "Baja Gestion" "Alta Gestion" ...
```

Implementación del algoritmo K-NN en R statistics

Se realiza el escalamiento de los datos

```
#=============================================================================
#Escalamiento de datos
#=============================================================================
datos_TPProceso <- scale(data_procesos[,c(2:6)])
View(datos_TPProceso)
#Se adiciona nuevamente la columna de etiqueta Centro-Año, Cluster y Gestion
datos_scale_TPProceso <- as.data.frame(cbind(data_procesos$Centro_Año, datos_TPProceso, data_procesos$Cluster,
data_procesos$Gestion))
View(datos_scale_TPProceso)
str(datos_scale_TPProceso)
dim(datos_scale_TPProceso)
datos_scale_TPProceso$Q_Proyectos <- as.numeric(datos_scale_TPProceso$Q_Proyectos)
datos_scale_TPProceso$QNA_5007 <- as.numeric(datos_scale_TPProceso$QNA_5007)
datos_scale_TPProceso$QNA_5009 <- as.numeric(datos_scale_TPProceso$QNA_5009)
datos_scale_TPProceso$QNA_5010 <- as.numeric(datos_scale_TPProceso$QNA_5010)
datos_scale_TPProceso$TV_Proyectos <- as.numeric(datos_scale_TPProceso$TV_Proyectos)
str(datos_scale_TPProceso)
names(datos_scale_TPProceso) <- names(data_procesos)
View(datos_scale_TPProceso)
row.names(datos_scale_TPProceso) <- datos_scale_TPProceso$Centro_Año
View(datos_scale_TPProceso)
proceso <- datos_scale_TPProceso[, c(-1,-7)]
View(proceso)
#=============================================================================
```

Implementación del algoritmo K-NN en R statistics

Se transforma la variable Gestion en una variable factor de tres niveles

```
#=============================================================================
# La variable factor está indexada en la 6ª posición
#=============================================================================

proceso$Gestion <- as.factor(proceso$Gestion)
View(proceso)
str(proceso)

#=============================================================================
```

```
> str(proceso)
'data.frame':   208 obs. of  6 variables:
 $ Q_Proyectos : num  -1.257 -0.362 -1.257 0.534 -0.362 ...
 $ QNA_5007    : num  -1.103 -0.689 0.137 0.55 0.137 ...
 $ QNA_5009    : num  -0.8238 -0.0502 -1.2106 0.3366 -0.437 ...
 $ QNA_5010    : num  -1.108 -0.225 -0.225 0.658 -0.225 ...
 $ TV_Proyectos: num  -0.126 -0.733 -0.206 -0.607 -1.573 ...
 $ Gestion     : Factor w/ 3 levels "Alta Gestion",..: 2 2 2 1 2 3 2 2 3 2 ...
```

Implementación del algoritmo K-NN en R statistics

Se obtiene una tabla, donde se evidencia la cantidad de observaciones por nivel de gestión

```
#=============================================================================
table(proceso$Gestion)
# Alta Gestion  Baja Gestion Media Gestion
#           47           122            39
#=============================================================================
```

```
> table(proceso$Gestion)

 Alta Gestion  Baja Gestion Media Gestion
           47           122            39
```

Porcentaje de centros por nivel de gestion

```
#=============================================================================
# Porcentaje de centros
#=============================================================================
round(prop.table(table(proceso$Gestion)) * 100,digits = 1)
# Alta Gestion  Baja Gestion Media Gestion
#         22.6          58.7          18.8
#=============================================================================
```

Implementación del algoritmo K-NN en R statistics

Analítica de datos

```
#=============================================================================
# Analitica de datos
#=============================================================================

#=============================================================================
# Paso 1: Preparación de datos: Escalado de datos y creación de conjuntos de
# datos de prueba y entrenamiento
#=============================================================================

#=============================================================================
# División aleatoria de los datos del iris como 70 % de tren y
# 30 % de conjuntos de datos de prueba
#=============================================================================
```

```
#=============================================================================
# División aleatoria de los datos como 70 % de entrenamiento y
# 30 % de conjuntos de datos de prueba
#=============================================================================

set.seed(123)
ind <- sample(2, nrow(proceso), replace=TRUE, prob=c(0.7, 0.3))
trainData <- proceso[ind==1,]
testData <-  proceso[ind==2,]
#=============================================================================
```

Implementación del algoritmo K-NN en R statistics

Datos de entrenamiento

```
> trainData
          Q_Proyectos   QNA_5007    QNA_5009   QNA_5010 TV_Proyectos       Gestion
9101-2010  -1.2572745 -1.1026653 -0.82382836 -1.1079853  -0.12603990  Baja Gestion
9127-2010  -0.3616817 -0.6894141 -0.05021076 -0.2249932  -0.73323429  Baja Gestion
9201-2010  -1.2572745  0.1370881 -1.21063716 -0.2249932  -0.20582319  Baja Gestion
9202-2010   0.5339111  0.5503392  0.33659804  0.6579990  -0.60746565  Alta Gestion
9203-2010  -0.3616817  0.1370881 -0.43701956 -0.2249932  -1.57324448  Baja Gestion
9204-2010   0.5339111  0.5503392  1.11021564 -0.2249932   0.69177849 Media Gestion
9205-2010  -1.2572745 -0.2761630 -1.21063716 -0.2249932   0.23107333  Baja Gestion
9301-2010   0.5339111  0.9635904  0.72340684 -1.1079853   1.42206695 Media Gestion
9401-2010   0.5339111 -0.6894141 -0.05021076 -0.2249932   0.86207222  Baja Gestion
9502-2010   0.5339111 -0.2761630  1.11021564 -0.2249932   0.31346051 Media Gestion
9549-2010  -0.3616817 -1.1026653 -0.43701956 -0.2249932  -0.97462198  Baja Gestion
9101-2011   0.5339111  0.5503392  0.72340684 -0.2249932   0.50462279 Media Gestion
9127-2011   1.4295039  0.9635904  2.65745084 -1.1079853   1.19785478 Media Gestion
9201-2011  -0.3616817 -1.1026653 -1.21063716  0.6579990  -0.24617717  Baja Gestion
9202-2011  -1.2572745 -1.1026653 -1.21063716 -0.2249932  -1.25254106  Baja Gestion
9203-2011   0.5339111 -0.2761630  1.88383324 -0.2249932   1.46865200 Media Gestion
9205-2011  -0.3616817  0.1370881 -0.43701956  0.6579990  -1.12733049  Baja Gestion
9206-2011  -0.3616817 -1.1026653 -0.43701956  0.6579990  -0.80493450  Baja Gestion
9301-2011  -0.3616817  0.1370881  1.49702444 -1.1079853  -1.31768069  Baja Gestion
9401-2011   0.5339111  1.7900926 -0.05021076  0.6579990  -0.02973318  Alta Gestion
9402-2011   1.4295039  0.1370881 -0.82382836  2.4239833   2.09936667  Alta Gestion
9501-2011  -0.3616817 -0.6894141 -0.43701956  0.6579990   0.08207881  Baja Gestion
9502-2011  -1.2572745 -0.2761630 -1.21063716 -0.2249932  -0.90577489  Baja Gestion
```

```
> dim(trainData)
[1] 146   6
```

Implementación del algoritmo K-NN en R statistics

Datos de prueba

```
> testData
         Q_Proyectos   QNA_5007    QNA_5009   QNA_5010 TV_Proyectos       Gestion
9206-2010  -1.2572745 -0.2761630 -0.82382836 -0.2249932  -1.48838030  Baja Gestion
9402-2010  -0.3616817 -0.6894141  0.72340684  0.6579990  -0.09479131  Baja Gestion
9501-2010  -0.3616817 -1.1026653 -0.05021076 -0.2249932   0.57977613  Baja Gestion
9503-2010  -0.3616817 -1.1026653 -0.05021076  0.6579990  -0.79056358  Baja Gestion
9504-2010  -1.2572745 -1.1026653 -0.43701956 -1.1079853  -1.02449911  Baja Gestion
9204-2011  -1.2572745 -1.5159164 -0.82382836 -1.1079853  -0.67824374  Baja Gestion
9503-2011  -0.3616817 -0.6894141 -0.82382836 -1.1079853  -0.77448962  Baja Gestion
9549-2011   2.3250967  1.3768415  1.11021564  0.6579990   3.13047861 Media Gestion
9204-2012   0.5339111 -0.6894141  1.11021564 -1.1079853   0.81676413 Media Gestion
9206-2012   0.5339111 -0.2761630  0.33659804  1.5409911   0.07111701  Alta Gestion
9301-2012  -0.3616817 -1.1026653 -0.43701956 -1.1079853  -1.44804700  Baja Gestion
9401-2012  -0.3616817 -1.1026653 -0.82382836  0.6579990  -0.19126295  Baja Gestion
9549-2012   0.5339111  0.5503392 -0.05021076  0.6579990   0.25909397  Alta Gestion
9202-2013  -0.3616817 -0.2761630  0.33659804 -1.1079853  -0.57919012  Baja Gestion
9203-2013  -1.2572745 -0.2761630 -1.21063716 -1.1079853  -0.54093683  Baja Gestion
9204-2013   0.5339111  0.9635904  0.33659804 -1.1079853   1.79306758 Media Gestion
9206-2013   2.3250967  4.2695994 -0.82382836  3.3069754   0.50338597  Alta Gestion
9504-2013  -0.3616817 -0.6894141 -0.05021076 -0.2249932   0.22382351  Baja Gestion
9201-2014  -0.3616817 -0.2761630 -0.43701956  0.6579990  -0.73354560  Baja Gestion
9202-2014   1.4295039  0.5503392  1.11021564  0.6579990  -0.07591645 Media Gestion
9203-2014  -0.3616817  0.5503392 -0.82382836  0.6579990  -0.63227742  Baja Gestion
9301-2014  -1.2572745 -0.6894141 -0.43701956 -1.1079853  -1.25243402  Baja Gestion
9204-2015  -1.2572745 -1.5159164 -1.59744596 -0.2249932  -0.65486134  Baja Gestion
9205-2015   0.5339111  0.9635904 -1.21063716  1.5409911  -0.02895020  Alta Gestion
9206-2015   0.5339111 -0.2761630  0.33659804  1.5409911  -0.56492677  Alta Gestion
9301-2015  -1.2572745 -1.1026653 -0.82382836 -1.1079853  -1.36962568  Baja Gestion
9501-2015  -0.3616817 -1.1026653 -0.05021076 -0.2249932  -0.23055609  Baja Gestion
9502-2015  -0.3616817 -1.5159164 -0.82382836  0.6579990   0.38050771  Baja Gestion
9504-2015  -1.2572745 -0.2761630 -1.21063716 -0.2249932  -1.44587320  Baja Gestion
9549-2015  -0.3616817 -0.6894141 -0.43701956 -0.2249932   0.27964318  Baja Gestion
9301-2016  -1.2572745 -1.1026653 -0.82382836 -1.1079853  -0.21917381  Baja Gestion
```

```
> dim(testData)
[1] 62  6
```

Implementación del algoritmo K-NN en R statistics

Verificación de dimensiones de datos de entrenamiento y prueba

```
#=============================================================================
# Verificamos las dimensiones de los conjuntos de datos de entrenamiento y prueba
#=============================================================================

dim(trainData)
dim(trainData1)
dim(testData)
dim(testData1)

#=============================================================================
```

```
> dim(testData)
[1] 57  6
> dim(trainData)
[1] 151   6
> dim(trainData1)
[1] 151   5
> dim(testData)
[1] 57  6
> dim(testData1)
[1] 57  5
```

Implementación del algoritmo K-NN en R statistics

Paso en el que se asegura que las etiquetas de entrenamiento y prueba sean tipo factor

```
#=============================================================================
# Paso 2: almacenar la variable objetivo para probar y entrenar datos:
# En este paso se debe asegurar que las etiquetas de entrenamiento y prueba
# sean datos de tipo de factor
#=============================================================================

data_procesos_train_labels <- trainData$Gestion
dim(data_procesos_train_labels)
class(data_procesos_train_labels)

data_procesos_test_labels <- testData$Gestion
dim(data_procesos_test_labels)
class(data_procesos_test_labels)

#=============================================================================
```

```
> data_procesos_train_labels <- trainData$Gestion
> dim(data_procesos_train_labels)
NULL
> class(data_procesos_train_labels)
[1] "factor"
> data_procesos_test_labels <- testData$Gestion
> dim(data_procesos_test_labels)
NULL
> class(data_procesos_test_labels)
[1] "factor"
```

Implementación del algoritmo K-NN en R statistics

Se implementa el algoritmo K-NN en R, mediante la función knn

```
#=====================================================================================
# Paso 3: entrenar un modelo con los datos
#=====================================================================================

install.packages("class") #Instalación de librerias
library(class)
data_procesos_test_pred1 <- knn(train = trainData1, test = testData1,
                                cl=data_procesos_train_labels, k = 3, prob=TRUE)
#KNN devuelve las etiquetas previstas para el conjunto de datos de prueba
#=====================================================================================
```

```
> data_procesos_test_pred1
 [1] Baja Gestion  Baja Gestion  Media Gestion Baja Gestion  Media Gestion Baja Gestion  Baja Gestion  Alta Gestion
 [9] Media Gestion Baja Gestion  Baja Gestion  Baja Gestion  Baja Gestion  Baja Gestion  Media Gestion Alta Gestion
[17] Baja Gestion  Alta Gestion  Baja Gestion  Alta Gestion  Alta Gestion  Baja Gestion  Baja Gestion  Baja Gestion
[25] Baja Gestion  Baja Gestion  Alta Gestion  Baja Gestion  Baja Gestion  Baja Gestion  Alta Gestion  Alta Gestion
[33] Baja Gestion  Baja Gestion  Media Gestion Baja Gestion  Alta Gestion  Alta Gestion  Baja Gestion  Alta Gestion
[41] Alta Gestion  Media Gestion Media Gestion Media Gestion Alta Gestion  Baja Gestion  Alta Gestion  Alta Gestion
[49] Media Gestion Baja Gestion  Alta Gestion  Alta Gestion  Alta Gestion  Baja Gestion  Baja Gestion  Baja Gestion
[57] Baja Gestion
attr(,"prob")
 [1] 1.0000000 1.0000000 1.0000000 0.6666667 1.0000000 1.0000000 1.0000000 0.6666667 1.0000000 1.0000000 1.0000000
[12] 1.0000000 1.0000000 1.0000000 0.6666667 1.0000000 1.0000000 0.6666667 1.0000000 1.0000000 0.6666667 1.0000000
[23] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.6666667 1.0000000 1.0000000 0.6666667 0.3333333
[34] 1.0000000 0.3333333 1.0000000 1.0000000 0.6666667 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.6666667
[45] 1.0000000 0.6666667 1.0000000 1.0000000 0.6666667 1.0000000 1.0000000 0.6666667 0.6666667 1.0000000 1.0000000
[56] 1.0000000 1.0000000
Levels: Alta Gestion Baja Gestion Media Gestion
```

Implementación del algoritmo K-NN en R statistics

Matriz de confusión

Matriz de confusión		Estimado por el modelo			
		Negativo (N)	*Positivo (P)*		
Real	*Negativo*	a: (TN)	b: (FP)		
	Positivo	c: (FN)	d: (TP)	**Precisión** ***("precision")*** Porcentaje predicciones positivas correctas:	d/(b+d)
		Sensibilidad, exhaustividad ***("Recall")*** Porcentaje casos positivos detectados	**Especifidad** ***(Specifity)*** Porcentaje casos negativos detectados	**Exactitud ("accuracy")** Porcentaje de predicciones correctas *(No sirve en datasets poco equilibrados)*	
		d/(d+c)	a/(a+b)	(a+d)/(a+b+c+d)	

Figura 4: Matriz de confusión y métricas asociadas

Implementación del algoritmo K-NN en R statistics

Métricas asociadas a la matriz de confusión

Exactitud

Si recordamos, la exactitud (o «accuracy«) representa el porcentaje de predicciones correctas frente al total. Por tanto, es el cociente entre los casos bien clasificados por el modelo (verdaderos positivos y verdaderos negativos, es decir, los valores en la diagonal de la matriz de confusión), y la suma de todos los casos.

Sin embargo, cuando un conjunto de datos es poco equilibrado, no es una métrica útil. Por ejemplo, si lo que intentamos predecir es una enfermedad rara, y nuestro algoritmo clasifica a todos los individuos como sanos, podría ser muy preciso (incluso un 99%), pero también, totalmente inútil.

Precisión

La **precisión,** (o"*precision*") se refiere a lo cerca que está el resultado de una predicción del valor verdadero. Por tanto, es el cociente entre los casos positivos bien clasificados por el modelo y el total de predicciones positivas.

Implementación del algoritmo K-NN en R statistics

Métricas asociadas a la matriz de confusión

Sensibilidad, exhaustividad

La sensibilidad (o recall) representa la tasa de verdaderos positivos (True Positive Rate) ó TP. Es la proporción entre los casos positivos bien clasificados por el modelo, respecto al total de positivos.

Especifidad

La especificidad, por su parte, es la tasa de verdaderos negativos, (“true negative rate”)o TN. Es la proporción entre los casos negativos bien clasificados por el modelo, respecto al total de negativos.

Matriz de confusión obtenida en R statistics

Matriz de confusión para el algoritmo K-NN en R, con k=3

```
> Validacion_modelo_procesos <- CrossTable(x = data_procesos_test_labels, y = data_procesos_test_pred1,prop.chisq=TRUE)


   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  62


                         | data_procesos_test_pred1
data_procesos_test_labels |  Alta Gestion |  Baja Gestion | Media Gestion |     Row Total |
--------------------------|---------------|---------------|---------------|---------------|
             Alta Gestion |            16 |             0 |             1 |            17 |
                          |        24.805 |         9.323 |         1.107 |               |
                          |         0.941 |         0.000 |         0.059 |         0.274 |
                          |         0.889 |         0.000 |         0.100 |               |
                          |         0.258 |         0.000 |         0.016 |               |
--------------------------|---------------|---------------|---------------|---------------|
             Baja Gestion |             1 |            34 |             0 |            35 |
                          |         8.260 |        11.422 |         5.645 |               |
                          |         0.029 |         0.971 |         0.000 |         0.565 |
                          |         0.056 |         1.000 |         0.000 |               |
                          |         0.016 |         0.548 |         0.000 |               |
--------------------------|---------------|---------------|---------------|---------------|
            Media Gestion |             1 |             0 |             9 |            10 |
                          |         1.248 |         5.484 |        33.833 |               |
                          |         0.100 |         0.000 |         0.900 |         0.161 |
                          |         0.056 |         0.000 |         0.900 |               |
                          |         0.016 |         0.000 |         0.145 |               |
--------------------------|---------------|---------------|---------------|---------------|
            Column Total |            18 |            34 |            10 |            62 |
                          |         0.290 |         0.548 |         0.161 |               |
--------------------------|---------------|---------------|---------------|---------------|
```

Matriz de confusión obtenida en R statistics

Matriz de confusión para el algoritmo K-NN en R, con k=3, y sus respectivas métricas

```
> con=confusionMatrix(testData[,ncol(testData)], data_procesos_test_pred1, mode =  "everything")
> con
Confusion Matrix and Statistics

               Reference
Prediction      Alta Gestion Baja Gestion Media Gestion
  Alta Gestion            16            0             1
  Baja Gestion             1           34             0
  Media Gestion            1            0             9

Overall Statistics

               Accuracy : 0.9516
                 95% CI : (0.865, 0.9899)
    No Information Rate : 0.5484
    P-Value [Acc > NIR] : 1.495e-12

                  Kappa : 0.9173

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Alta Gestion Class: Baja Gestion Class: Media Gestion
Sensitivity                       0.8889              1.0000                0.9000
Specificity                       0.9773              0.9643                0.9808
Pos Pred Value                    0.9412              0.9714                0.9000
Neg Pred Value                    0.9556              1.0000                0.9808
Precision                         0.9412              0.9714                0.9000
Recall                            0.8889              1.0000                0.9000
F1                                0.9143              0.9855                0.9000
Prevalence                        0.2903              0.5484                0.1613
Detection Rate                    0.2581              0.5484                0.1452
Detection Prevalence              0.2742              0.5645                0.1613
Balanced Accuracy                 0.9331              0.9821                0.9404
```

Métricas del clasificador K-NN con k=3 en R statistics

```
#=============================================================================
# se almacena la matriz de confusion en una tabla
#=============================================================================

Tabla_matrix<-as.data.frame(matrix(data = con$table, nrow=3,ncol=3))
names(Tabla_matrix) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")
row.names(Tabla_matrix) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")
Tabla_matrix$Total <- 0
Tabla_matrix <- rbind(Tabla_matrix,c(0,0,0,0))
row.names(Tabla_matrix) <-c("Alta Gestion", "Baja Gestion", "Media Gestion", "Total")
Tabla_matrix$Total <- c(sum(Tabla_matrix[1,]), sum(Tabla_matrix[2,]), sum(Tabla_matrix[3,]), 0)
Tabla_matrix[4,] <- c(sum(Tabla_matrix$`Alta Gestion`), sum(Tabla_matrix$`Baja Gestion`), sum
(Tabla_matrix$`Media Gestion`), sum(Tabla_matrix$Total))
#=============================================================================
```

```
#=============================================================================
# Tabla de metricas para las tres clases
#=============================================================================

Tabla_metricas_clase <- as.data.frame(matrix(data=t(con$byClass), nrow=11, ncol=3))
row.names(Tabla_metricas_clase) <- c("Sensitivity", "Specificity", "Pos Pred Value",
                                     "Neg Pred Value","Precision", "Recall", "F1",
                                     "Prevalence", "Detection Rate", "Detection Prevalence",
                                     "Balanced Accuracy" )
names(Tabla_metricas_clase) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")

#=============================================================================
```

Métricas del clasificador K-NN con k=3 en R statistics

```
#=============================================================================
# Calculo de las métricas del clasificador
#=============================================================================

Accuracy <- con$overall[1]
clasificador_metricas <- data.frame("Metricas_del_Clasificador"=c(0,0,0,0,0,0))
k <- c(1,2,5,6,7)
for (i in 1:5)
{
  for (j in 1:3)
  {
    clasificador_metricas[i,1] <- clasificador_metricas[i,1] +
      (Tabla_matrix[j,4]/Tabla_matrix[4,4])*Tabla_metricas_clase[k[i],j]
  }

}
clasificador_metricas[6,1] <- Accuracy
row.names(clasificador_metricas) <- c("Sensitivity", "Specificity",
                                      "Precision", "Recall", "F1", "Accuracy" )
clasificador_metricas <- t(clasificador_metricas)

#=============================================================================
```

```
> clasificador_metricas
                        Sensitivity Specificity Precision   Recall        F1  Accuracy
Metricas_del_Clasificador    0.953405   0.9705053 0.9516129 0.953405 0.9521873 0.9516129
```

Implementación del algoritmo K-NN en R statistics

Validación del modelo a través de la matriz de confusión

```
#=====================================================================================
# Matriz de Confusión

# Una vez obtenido el modelo podemos contrastar con el dataset de test su porcentaje
# de acierto, del 95,16%
#=====================================================================================

con=confusionMatrix(testData[,ncol(testData)], data_procesos_test_pred1, mode =  "everything")

#=====================================================================================
```

Implementación del algoritmo K-NN en R statistics

Para nuestro caso la tabla de contingencia arrojada por la matriz de confusión consta de $N = 62$ observaciones clasificadas en forma cruzada en las variables de fila y columna, que normalmente tendrían los mismos niveles, sin embargo la estadístico de McNemar se usa para **probar la hipótesis nula de simetría**, es decir, **que la probabilidad de que una observación se clasifique en celda [i,j] es igual a la probabilidad de ser clasificado en la celda [j,i]**, (mcnemar.test)

Para nuestro caso al obtener un P-Value: NA, es un indicio de que la clasificación esta fallando, por lo tanto debemos calcular cual es el óptimo k-vécino cercano en nuestra base de datos, como se describe a continuación.

References

[1] mcnemar.test. McNemar's Chi-Square Test for Count Data. https://docs.tibco.com/pub/enterprise-runtime-for-R/3.1.1/doc/html/Language_Reference/stats/mcnemar.test.html

[2] Fleiss, J. L. (1981). Statistical Methods for Rates and Proportions, 2nd ed. New York: Wiley.

[3] Zar, J. H. (1984). Biostatistical Analysis, 2nd ed. Englewood Cliffs: Prentice-Hall.

Alan Agresti (1990). Categorical data analysis. New York: Wiley. Pages 350--354.

Implementación del algoritmo K-NN en R statistics – k = 3

Validación del modelo a través de la matriz de confusión

```
> con=confusionMatrix(testData[,ncol(testData)], data_procesos_test_pred1, mode =  "everything")
> con
Confusion Matrix and Statistics

                Reference
Prediction       Alta Gestion Baja Gestion Media Gestion
  Alta Gestion             16            0             1
  Baja Gestion              1           34             0
  Media Gestion             1            0             9

Overall Statistics

               Accuracy : 0.9516
                 95% CI : (0.865, 0.9899)
    No Information Rate : 0.5484
    P-Value [Acc > NIR] : 1.495e-12

                  Kappa : 0.9173

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Alta Gestion Class: Baja Gestion Class: Media Gestion
Sensitivity                       0.8889              1.0000                0.9000
Specificity                       0.9773              0.9643                0.9808
Pos Pred Value                    0.9412              0.9714                0.9000
Neg Pred Value                    0.9556              1.0000                0.9808
Precision                         0.9412              0.9714                0.9000
Recall                            0.8889              1.0000                0.9000
F1                                0.9143              0.9855                0.9000
Prevalence                        0.2903              0.5484                0.1613
Detection Rate                    0.2581              0.5484                0.1452
Detection Prevalence              0.2742              0.5645                0.1613
Balanced Accuracy                 0.9331              0.9821                0.9404
```

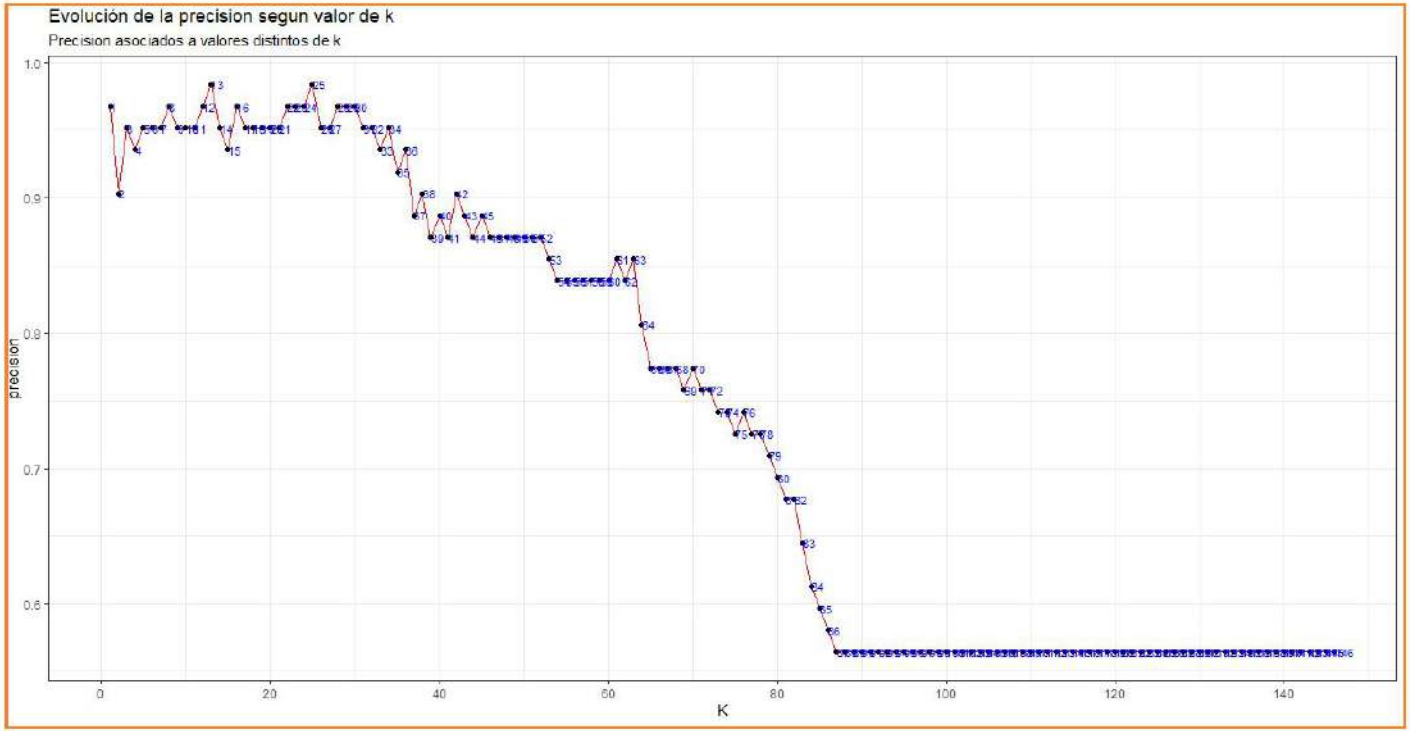
Calculo del valor de K óptimo para el K-NN en R statistics

Hasta el momento solo hemos utilizado un solo valor para k, 3 vecinos más cercanos. Sin embargo, podríamos ver la exactitud o *accuracy* para diferentes valores de k. Como tenemos 146 valores en nuestra data de entrenamiento veremos los aciertos tomando máximo 146 vecinos más cercanos.

```
#=============================================================================

k <- 1:146
resultado <- data.frame(k, precision = 0)

for(n in k){
  proceso_test_output_kNN <- knn(train =  trainData1,
                                 cl = data_procesos_train_labels,
                                 test = testData1,
                                 k = n)

  resultado$precision[n] <- mean(proceso_test_output_kNN == testData[,6])
}
resultado
library(extrafont)
#resultado %>%
ggplot(resultado,aes(x=k, y=precision))+
  labs(title = 'Evolución de la precision segun valor de k')+
  labs(subtitle = 'Precision asociados a valores distintos de k')+
  geom_line(color="red")+ geom_point()+
  scale_x_continuous(breaks=pretty(resultado$k, n = 10), name="K")+
  theme(legend.position="top", legend.text = element_text(colour="blue", size=6)) +
  geom_text(aes(label=k),hjust = 0, nudge_x = 0.05, color="blue",size=3)+
  theme(text=element_text(size=12, family="Calibri"))+
  theme_bw()
#=============================================================================
```

Calculo del valor de K óptimo para el K-NN en R statistics

Precisión del modelo para diferentes valores de k



Como vemos, para este caso, a partir de cierto número de vecinos más cercanos el nivel de aciertos de nuestro algoritmo empieza a reducirse. Dependerá de cada caso el escoger el mejor “k” para nuestro modelo, para el caso se elige $k = 2$

Matriz de confusión obtenida en R statistics

Matriz de confusión para el algoritmo K-NN en R, con k=2, y sus respectivas métricas

```
> con
Confusion Matrix and Statistics

                Reference
Prediction       Alta Gestion Baja Gestion Media Gestion
  Alta Gestion             15            0             2
  Baja Gestion              1           33             1
  Media Gestion             2            0             8

Overall Statistics

               Accuracy : 0.9032
                 95% CI : (0.8012, 0.9637)
    No Information Rate : 0.5323
    P-Value [Acc > NIR] : 3.354e-10

                  Kappa : 0.8363

 Mcnemar's Test P-Value : 0.5724

Statistics by Class:

                     Class: Alta Gestion Class: Baja Gestion Class: Media Gestion
Sensitivity                       0.8333              1.0000                0.7273
Specificity                       0.9545              0.9310                0.9608
Pos Pred Value                    0.8824              0.9429                0.8000
Neg Pred Value                    0.9333              1.0000                0.9423
Precision                         0.8824              0.9429                0.8000
Recall                            0.8333              1.0000                0.7273
F1                                0.8571              0.9706                0.7619
Prevalence                        0.2903              0.5323                0.1774
Detection Rate                    0.2419              0.5323                0.1290
Detection Prevalence              0.2742              0.5645                0.1613
Balanced Accuracy                 0.8939              0.9655                0.8440
```

Implementación del algoritmo K-NN en R statistics

Evaluación del rendimiento del modelo: métricas asociadas al clasificador K-NN, con k=2

```
> clasificador_metricas
                          Sensitivity Specificity Precision    Recall        F1  Accuracy
Metricas_del_Clasificador   0.9103128   0.9422794 0.9032258 0.9103128 0.9058236 0.9032258
```

Implementación del algoritmo K-NN en R statistics

Análisis de resultados

Hemos clasificado 15 centros de Alta Gestion como Alta Gestion, lo que indica que está bien clasificado y 2 centros de Alta Gestion como de Media Gestion, lo que indica una incorrecta clasificación. A su vez analizando los centros en Baja Gestion la predicción en la clasificación arroja que 33 centros de Baja Gestion quedaron clasificados como de Baja Gestion lo cual indica que quedaron bien clasificados. Por otro lado, 1 centro siendo de Baja Gestion quedo clasificado como que fuera de Alta Gestion, asimismo otro centro siendo de Baja Gestion quedo clasificado en Media Gestion. Por último 8 centros de Media Gestion fueron clasificados como de Media Gestion, lo cual indica una correcta clasificación y 2 centros de pertenecientes a Media Gestion quedaron clasificados como si fueran de Alta Gestion, lo cual indica que esta incorrectamente clasificado.

Implementación del algoritmo K-NN en R statistics

Conclusiones

Sensibilidad de acuerdo al valor obtenido del 0.9103128, esto me indica **¿qué porcentaje de los centros de acuerdo a su nivel de gestión somos capaces de identificar?** Para lo cuál diríamos que el 91,03% de los centros se identifican acertadamente de acuerdo al nivel de gestión.

El valor **F1** se utiliza para combinar las medidas de **precision** y **recall (sensibilidad)** en un sólo valor. Asume que nos importa de igual forma la precisión y la exhaustividad. Esto no tiene que ser así en todos los problemas. Ya que necesitamos predecir si el nivel de gestión del centro se detecto adecuadamente, por lo tanto podemos valorar más la sensibilidad a riesgo de tener una peor precisión. Sin embargo para nuestro caso el F1 es muy bueno igual 0,9058236 equivalente al 90,58% lo que es muy aproximado al 91,03% de sensibilidad del modelo y a la precisión del modelo equivalente a 90,32%.

Implementación del algoritmo K-NN en R statistics

Conclusiones

Exactitud (Accuracy): La exactitud (accuracy) mide el porcentaje de casos que el modelo ha acertado. Es decir, el modelo acierta el 90,32% de las veces. Como se observa, la exactitud es una métrica muy engañosa. De hecho, si tuviésemos un modelo que siempre predijera que el centro siempre va a estar en un nivel de gestion determinado, su accuracy podría ser del 80% o más y esto normalmente se observa en problemas con clases desbalanceadas por lo tanto es mucho mejor usar precision, recall y F1. Estas métricas dan una mejor idea de la calidad del modelo. IArtificial.net. (2020).

Referencias

IArtificial.net. (2020). https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/

Especificidad: el modelo tiene una muy alta capacidad para identificar los centros que no fueron clasificados correctamente según su nivel de gestión entre todos los centros cuyos niveles de gestión fueron clasificados correctamente. Para nuestro caso la especificidad del clasificador es de 94,23%.

Implementación del algoritmo K-NN en R statistics

Conclusiones

Por lo tanto, este modelo es mucho más específico que sensible en la clasificación esto se traduce en que si un falso positivo significa que el modelo tiene una muy alta capacidad para identificar los centros que no fueron clasificados correctamente según su nivel de gestión evitando los falsos positivos.

GC-F-004v.05

6. Modelo de analítica de datos para medición de Productividad por centros – Algoritmo KNN

Implementación del algoritmo K-NN en R statistics

Lectura de base de datos Tabla_Nivel_Productividad.xlsx

```
#=============================================================================
# Base de datos procesos: Tabla_Nivel_Productividad.xlsx
#=============================================================================
setwd("E:/Documents/Codigo en R/SENA paola")
data_productividad <- read_xlsx("Tabla_Nivel_Productividad.xlsx")
View(data_productividad)
#=============================================================================
```

Filter

	Centro_Año	Cantidad_Proyectos	Fecha_inicio_Proyecto	Cantidad_Productos	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031	Cluster	Nivel_Productividad
1	9101-2010	4	2010	14	1	3	1	6	0	3	2	Alta Productividad
2	9127-2010	2	2010	8	1	0	0	2	3	2	1	Baja Productividad
3	9201-2010	1	2010	2	0	0	0	0	0	2	1	Baja Productividad
4	9202-2010	1	2010	4	0	1	0	2	0	1	1	Baja Productividad
5	9203-2010	5	2010	19	6	2	2	4	2	3	2	Alta Productividad
6	9204-2010	2	2010	7	0	1	2	0	1	3	1	Baja Productividad
7	9205-2010	1	2010	2	0	0	1	0	0	1	1	Baja Productividad
8	9206-2010	4	2010	14	2	4	2	1	4	1	3	Mediana Productividad
9	9301-2010	2	2010	8	0	2	2	1	3	0	1	Baja Productividad
10	9401-2010	1	2010	4	1	0	2	0	0	1	1	Baja Productividad
11	9402-2010	2	2010	7	2	1	1	1	0	2	1	Baja Productividad
12	9501-2010	3	2010	10	1	0	1	5	0	3	1	Baja Productividad
13	9502-2010	1	2010	3	0	1	0	0	2	0	1	Baja Productividad
14	9503-2010	3	2010	10	1	2	5	0	1	1	3	Mediana Productividad

Showing 1 to 14 of 208 entries, 12 total columns

Implementación del algoritmo K-NN en R statistics

Exploración de los datos

```
#=============================================================================
# Primer paso: Exploración de los datos
#=============================================================================
summary(data_productividad)
str(data_productividad)
#=============================================================================
```

```
> summary(data_productividad)
  Centro_Año        Cantidad_Proyectos Fecha_inicio_Proyecto Cantidad_Productos
 Length:208         Min.   :1.000      Min.   :2010          Min.   : 2.00
 Class :character   1st Qu.:1.000      1st Qu.:2013          1st Qu.: 7.00
 Mode  :character   Median :2.000      Median :2016          Median :10.00
                    Mean   :2.404      Mean   :2016          Mean   :10.75
                    3rd Qu.:3.000      3rd Qu.:2019          3rd Qu.:14.00
                    Max.   :6.000      Max.   :2022          Max.   :26.00
     P_8020           P_8025          P_8019          P_8039          P_8016
 Min.   :0.000   Min.   :0.000   Min.   :0.00   Min.   :0.000   Min.   :0.000
 1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.00   1st Qu.:1.000   1st Qu.:1.000
 Median :2.000   Median :2.000   Median :2.00   Median :1.000   Median :1.000
 Mean   :1.889   Mean   :1.793   Mean   :1.76   Mean   :1.745   Mean   :1.654
 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.00   3rd Qu.:2.250   3rd Qu.:2.000
 Max.   :8.000   Max.   :7.000   Max.   :7.00   Max.   :7.000   Max.   :8.000
     P_8031          Cluster      Nivel_Productividad
 Min.   :0.000   Min.   :1.000   Length:208
 1st Qu.:1.000   1st Qu.:1.000   Class :character
 Median :2.000   Median :1.000   Mode  :character
 Mean   :1.904   Mean   :1.736
 3rd Qu.:3.000   3rd Qu.:3.000
 Max.   :6.000   Max.   :3.000
> str(data_productividad)
tibble [208 x 12] (S3: tbl_df/tbl/data.frame)
 $ Centro_Año          : chr [1:208] "9101-2010" "9127-2010" "9201-2010" "9202-2010" ...
 $ Cantidad_Proyectos  : num [1:208] 4 2 1 1 5 2 1 4 2 1 ...
 $ Fecha_inicio_Proyecto: num [1:208] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
 $ Cantidad_Productos  : num [1:208] 14 8 2 4 19 7 2 14 8 4 ...
 $ P_8020              : num [1:208] 1 1 0 0 6 0 0 2 0 1 ...
 $ P_8025              : num [1:208] 3 0 0 1 2 1 0 4 2 0 ...
 $ P_8019              : num [1:208] 1 0 0 0 2 2 1 2 2 2 ...
 $ P_8039              : num [1:208] 6 2 0 2 4 0 0 1 1 0 ...
 $ P_8016              : num [1:208] 0 3 0 0 2 1 0 4 3 0 ...
 $ P_8031              : num [1:208] 3 2 2 1 3 3 1 1 0 1 ...
 $ Cluster             : num [1:208] 2 1 1 1 2 1 1 3 1 1 ...
 $ Nivel_Productividad : chr [1:208] "Alta Productividad" "Baja Productividad" "Baja Productividad" "Baja Productividad" ...
```

Implementación del algoritmo K-NN en R statistics

Se convierten los datos en un data.frame

```
#=============================================================================
# Segundo paso: convertimos la variable de clasificación Gestión en factor
# Con el propósito de identificarla por niveles, pero primero convertimos los datos
# en data.frame
#=============================================================================
data_productividad <- as.data.frame(data_productividad)
str(data_productividad)
#=============================================================================
```

```
> data_productividad <- as.data.frame(data_productividad)
> str(data_productividad)
'data.frame':   208 obs. of  12 variables:
 $ Centro_Año           : chr  "9101-2010" "9127-2010" "9201-2010" "9202-2010" ...
 $ Cantidad_Proyectos   : num  4 2 1 1 5 2 1 4 2 1 ...
 $ Fecha_inicio_Proyecto: num  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
 $ Cantidad_Productos   : num  14 8 2 4 19 7 2 14 8 4 ...
 $ P_8020               : num  1 1 0 0 6 0 0 2 0 1 ...
 $ P_8025               : num  3 0 0 1 2 1 0 4 2 0 ...
 $ P_8019               : num  1 0 0 0 2 2 1 2 2 2 ...
 $ P_8039               : num  6 2 0 2 4 0 0 1 1 0 ...
 $ P_8016               : num  0 3 0 0 2 1 0 4 3 0 ...
 $ P_8031               : num  3 2 2 1 3 3 1 1 0 1 ...
 $ Cluster              : num  2 1 1 1 2 1 1 3 1 1 ...
 $ Nivel_Productividad  : chr  "Alta Productividad" "Baja Productividad" "Baja Productividad" "Baja Productividad" ...
```

Implementación del algoritmo K-NN en R statistics

Se realiza el escalamiento de los datos

```
#=============================================================================
#Escalamiento de datos
#=============================================================================
datos_TPProductividad <- scale(data_productividad[,c(5:10)])
View(datos_TPProductividad)
#Se adiciona nuevamente la columna de etiqueta Centro-Año, Cluster y Gestion
datos_scale_TPProductividad <- as.data.frame(cbind(data_productividad$Centro_Año, datos_TPProductividad, data_productividad$Cluster,
data_productividad$Nivel_Productividad))
View(datos_scale_TPProductividad)
str(datos_scale_TPProductividad)
dim(datos_scale_TPProductividad)
datos_scale_TPProductividad$P_8020 <- as.numeric(datos_scale_TPProductividad$P_8020)
datos_scale_TPProductividad$P_8025 <- as.numeric(datos_scale_TPProductividad$P_8025)
datos_scale_TPProductividad$P_8019 <- as.numeric(datos_scale_TPProductividad$P_8020)
datos_scale_TPProductividad$P_8039 <- as.numeric(datos_scale_TPProductividad$P_8025)
datos_scale_TPProductividad$P_8016 <- as.numeric(datos_scale_TPProductividad$P_8020)
datos_scale_TPProductividad$P_8031 <- as.numeric(datos_scale_TPProductividad$P_8025)
str(datos_scale_TPProductividad)
names(datos_scale_TPProductividad) <- names(data_productividad[,c(1, 5:12)])
View(datos_scale_TPProductividad)
row.names(datos_scale_TPProductividad) <- datos_scale_TPProductividad$Centro_Año
View(datos_scale_TPProductividad)
productividad <- datos_scale_TPProductividad[, c(-1, -8)]
View(productividad)
#=============================================================================
```

Implementación del algoritmo K-NN en R statistics

Tabla de datos escalada a clasificar con el algoritmo K-NN

Filter

	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031	Nivel_Productividad
9101-2010	-0.55877727	0.8552413	-0.55877727	0.8552413	-0.55877727	0.8552413	Alta Productividad
9127-2010	-0.55877727	-1.2709363	-0.55877727	-1.2709363	-0.55877727	-1.2709363	Baja Productividad
9201-2010	-1.18702415	-1.2709363	-1.18702415	-1.2709363	-1.18702415	-1.2709363	Baja Productividad
9202-2010	-1.18702415	-0.5622104	-1.18702415	-0.5622104	-1.18702415	-0.5622104	Baja Productividad
9203-2010	2.58245713	0.1465154	2.58245713	0.1465154	2.58245713	0.1465154	Alta Productividad
9204-2010	-1.18702415	-0.5622104	-1.18702415	-0.5622104	-1.18702415	-0.5622104	Baja Productividad
9205-2010	-1.18702415	-1.2709363	-1.18702415	-1.2709363	-1.18702415	-1.2709363	Baja Productividad
9206-2010	0.06946961	1.5639672	0.06946961	1.5639672	0.06946961	1.5639672	Mediana Productividad
9301-2010	-1.18702415	0.1465154	-1.18702415	0.1465154	-1.18702415	0.1465154	Baja Productividad
9401-2010	-0.55877727	-1.2709363	-0.55877727	-1.2709363	-0.55877727	-1.2709363	Baja Productividad
9402-2010	0.06946961	-0.5622104	0.06946961	-0.5622104	0.06946961	-0.5622104	Baja Productividad
9501-2010	-0.55877727	-1.2709363	-0.55877727	-1.2709363	-0.55877727	-1.2709363	Baja Productividad
9502-2010	-1.18702415	-0.5622104	-1.18702415	-0.5622104	-1.18702415	-0.5622104	Baja Productividad
9503-2010	-0.55877727	0.1465154	-0.55877727	0.1465154	-0.55877727	0.1465154	Mediana Productividad
9504-2010	0.06946961	1.5639672	0.06946961	1.5639672	0.06946961	1.5639672	Mediana Productividad
9549-2010	0.06946961	-0.5622104	0.06946961	-0.5622104	0.06946961	-0.5622104	Baja Productividad
9101-2011	-1.18702415	-1.2709363	-1.18702415	-1.2709363	-1.18702415	-1.2709363	Baja Productividad
9127-2011	0.06946961	-1.2709363	0.06946961	-1.2709363	0.06946961	-1.2709363	Baja Productividad

Showing 1 to 19 of 208 entries, 7 total columns

Implementación del algoritmo K-NN en R statistics

Se transforma la variable Nivel_Productividad en una variable factor de tres niveles

```
#=============================================================================
# La variable factor está indexada en la 7ª posición
#=============================================================================

productividad$Nivel_Productividad <- as.factor(productividad$Nivel_Productividad)
View(productividad)
str(productividad)

#=============================================================================
```

```
> str(productividad)
'data.frame':   208 obs. of  7 variables:
 $ P_8020             : num  -0.559 -0.559 -1.187 -1.187 2.582 ...
 $ P_8025             : num  0.855 -1.271 -1.271 -0.562 0.147 ...
 $ P_8019             : num  -0.559 -0.559 -1.187 -1.187 2.582 ...
 $ P_8039             : num  0.855 -1.271 -1.271 -0.562 0.147 ...
 $ P_8016             : num  -0.559 -0.559 -1.187 -1.187 2.582 ...
 $ P_8031             : num  0.855 -1.271 -1.271 -0.562 0.147 ...
 $ Nivel_Productividad: Factor w/ 3 levels "Alta Productividad",..: 1 2 2 2 1 2 2 3 2 2 ...
```

Implementación del algoritmo K-NN en R statistics

Se obtiene una tabla, donde se evidencia la cantidad de observaciones por nivel de productividad

```
#=============================================================================
table(productividad$Nivel_Productividad)
# Alta Productividad  Baja Productividad Media Productividad
#           39                   112                  57
#=============================================================================
```

```
> #=============================================================================
> table(productividad$Nivel_Productividad)

   Alta Productividad    Baja Productividad Mediana Productividad
                   39                   112                   57
> #=============================================================================
```

Porcentaje de centros por nivel de productividad

```
> round(prop.table(table(productividad$Nivel_Productividad)) * 100,digits = 1)

   Alta Productividad    Baja Productividad Mediana Productividad
                 18.8                  53.8                  27.4
```

Implementación del algoritmo K-NN en R statistics

Analítica de datos

```
#=========================================================================================
# Analitica de datos
#=========================================================================================

#=========================================================================================
# Paso 1: Preparación de datos: Escalado de datos y creación de conjuntos de
# datos de prueba y entrenamiento
#=========================================================================================

#=========================================================================================
# División aleatoria de los datos como 70 % de entrenamiento y
# 30 % de conjuntos de datos de prueba
#=========================================================================================
```

```
#=========================================================================================
# División aleatoria de los datos como 70 % de entrenamiento y
# 30 % de conjuntos de datos de prueba
#=========================================================================================

ind <- sample(2, nrow(productividad), replace=TRUE, prob=c(0.7, 0.3))
trainData <- productividad[ind==1,]
testData <-  productividad[ind==2,]
#=========================================================================================
```

Implementación del algoritmo K-NN en R statistics

Datos de entrenamiento

```
> testData <-  productividad[ind==2,]
> trainData
              P_8020     P_8025      P_8019     P_8039      P_8016     P_8031   Nivel_Productividad
9127-2010 -0.55877727 -1.2709363 -0.55877727 -1.2709363 -0.55877727 -1.2709363    Baja Productividad
9201-2010 -1.18702415 -1.2709363 -1.18702415 -1.2709363 -1.18702415 -1.2709363    Baja Productividad
9202-2010 -1.18702415 -0.5622104 -1.18702415 -0.5622104 -1.18702415 -0.5622104    Baja Productividad
9204-2010 -1.18702415 -0.5622104 -1.18702415 -0.5622104 -1.18702415 -0.5622104    Baja Productividad
9205-2010 -1.18702415 -1.2709363 -1.18702415 -1.2709363 -1.18702415 -1.2709363    Baja Productividad
9206-2010  0.06946961  1.5639672  0.06946961  1.5639672  0.06946961  1.5639672 Mediana Productividad
9401-2010 -0.55877727 -1.2709363 -0.55877727 -1.2709363 -0.55877727 -1.2709363    Baja Productividad
9502-2010 -1.18702415 -0.5622104 -1.18702415 -0.5622104 -1.18702415 -0.5622104    Baja Productividad
9503-2010 -0.55877727  0.1465154 -0.55877727  0.1465154 -0.55877727  0.1465154 Mediana Productividad
9504-2010  0.06946961  1.5639672  0.06946961  1.5639672  0.06946961  1.5639672 Mediana Productividad
9101-2011 -1.18702415 -1.2709363 -1.18702415 -1.2709363 -1.18702415 -1.2709363    Baja Productividad
9127-2011  0.06946961 -1.2709363  0.06946961 -1.2709363  0.06946961 -1.2709363    Baja Productividad
9202-2011  1.32596337  0.1465154  1.32596337  0.1465154  1.32596337  0.1465154 Mediana Productividad
9203-2011  0.06946961  0.1465154  0.06946961  0.1465154  0.06946961  0.1465154    Baja Productividad
9205-2011 -0.55877727 -0.5622104 -0.55877727 -0.5622104 -0.55877727 -0.5622104    Alta Productividad
9501-2011  1.32596337  0.1465154  1.32596337  0.1465154  1.32596337  0.1465154 Mediana Productividad
9503-2011 -0.55877727 -0.5622104 -0.55877727 -0.5622104 -0.55877727 -0.5622104    Baja Productividad
9504-2011  1.32596337  0.8552413  1.32596337  0.8552413  1.32596337  0.8552413 Mediana Productividad
9549-2011  0.06946961 -0.5622104  0.06946961 -0.5622104  0.06946961 -0.5622104    Baja Productividad
9101-2012 -1.18702415 -0.5622104 -1.18702415 -0.5622104 -1.18702415 -0.5622104    Alta Productividad
9127-2012  0.06946961  0.1465154  0.06946961  0.1465154  0.06946961  0.1465154    Baja Productividad
9201-2012 -0.55877727  0.8552413 -0.55877727  0.8552413 -0.55877727  0.8552413    Baja Productividad
9202-2012  0.69771649  0.1465154  0.69771649  0.1465154  0.69771649  0.1465154 Mediana Productividad
9203-2012  0.06946961 -1.2709363  0.06946961 -1.2709363  0.06946961 -1.2709363    Baja Productividad
9204-2012  1.32596337  1.5639672  1.32596337  1.5639672  1.32596337  1.5639672    Alta Productividad
9206-2012 -0.55877727  0.8552413 -0.55877727  0.8552413 -0.55877727  0.8552413    Baja Productividad
9301-2012 -1.18702415  0.1465154 -1.18702415  0.1465154 -1.18702415  0.1465154 Mediana Productividad
9401-2012 -0.55877727  0.1465154 -0.55877727  0.1465154 -0.55877727  0.1465154 Mediana Productividad
9501-2012  1.32596337  0.1465154  1.32596337  0.1465154  1.32596337  0.1465154    Baja Productividad
9502-2012  0.69771649  0.8552413  0.69771649  0.8552413  0.69771649  0.8552413 Mediana Productividad
9504-2012  0.06946961 -0.5622104  0.06946961 -0.5622104  0.06946961 -0.5622104    Baja Productividad
```

```
> dim(trainData)
[1] 134   7
```

Implementación del algoritmo K-NN en R statistics

Datos de prueba

```
> testData <-  productividad[ind==2,]
> testData
               P_8020     P_8025      P_8019     P_8039      P_8016     P_8031   Nivel_Productividad
9101-2010 -0.55877727  0.8552413 -0.55877727  0.8552413 -0.55877727  0.8552413    Alta Productividad
9203-2010  2.58245713  0.1465154  2.58245713  0.1465154  2.58245713  0.1465154    Alta Productividad
9301-2010 -1.18702415  0.1465154 -1.18702415  0.1465154 -1.18702415  0.1465154    Baja Productividad
9402-2010  0.06946961 -0.5622104  0.06946961 -0.5622104  0.06946961 -0.5622104    Baja Productividad
9501-2010 -0.55877727 -1.2709363 -0.55877727 -1.2709363 -0.55877727 -1.2709363    Baja Productividad
9549-2010  0.06946961 -0.5622104  0.06946961 -0.5622104  0.06946961 -0.5622104    Baja Productividad
9201-2011 -0.55877727 -1.2709363 -0.55877727 -1.2709363 -0.55877727 -1.2709363    Baja Productividad
9204-2011  0.06946961  0.8552413  0.06946961  0.8552413  0.06946961  0.8552413 Mediana Productividad
9206-2011 -1.18702415  2.9814190 -1.18702415  2.9814190 -1.18702415  2.9814190 Mediana Productividad
9301-2011  0.06946961 -0.5622104  0.06946961 -0.5622104  0.06946961 -0.5622104    Alta Productividad
9401-2011  1.95421025 -0.5622104  1.95421025 -0.5622104  1.95421025 -0.5622104    Baja Productividad
9402-2011  0.69771649 -1.2709363  0.69771649 -1.2709363  0.69771649 -1.2709363    Baja Productividad
9502-2011 -0.55877727 -0.5622104 -0.55877727 -0.5622104 -0.55877727 -0.5622104    Alta Productividad
9205-2012  0.69771649  2.2726931  0.69771649  2.2726931  0.69771649  2.2726931 Mediana Productividad
9402-2012  0.06946961  1.5639672  0.06946961  1.5639672  0.06946961  1.5639672 Mediana Productividad
9503-2012  0.06946961  0.8552413  0.06946961  0.8552413  0.06946961  0.8552413 Mediana Productividad
9201-2013  1.32596337  0.8552413  1.32596337  0.8552413  1.32596337  0.8552413    Alta Productividad
9202-2013 -1.18702415  0.1465154 -1.18702415  0.1465154 -1.18702415  0.1465154    Baja Productividad
9206-2013  0.06946961  2.2726931  0.06946961  2.2726931  0.06946961  2.2726931 Mediana Productividad
9101-2014 -1.18702415  0.1465154 -1.18702415  0.1465154 -1.18702415  0.1465154    Baja Productividad
9202-2014 -0.55877727  0.8552413 -0.55877727  0.8552413 -0.55877727  0.8552413    Baja Productividad
9205-2014  0.06946961  1.5639672  0.06946961  1.5639672  0.06946961  1.5639672 Mediana Productividad
9206-2014 -1.18702415 -0.5622104 -1.18702415 -0.5622104 -1.18702415 -0.5622104    Baja Productividad
9501-2014 -1.18702415  0.1465154 -1.18702415  0.1465154 -1.18702415  0.1465154    Baja Productividad
9503-2014 -1.18702415 -0.5622104 -1.18702415 -0.5622104 -1.18702415 -0.5622104    Baja Productividad
9127-2015  0.06946961  0.1465154  0.06946961  0.1465154  0.06946961  0.1465154 Mediana Productividad
9401-2015 -0.55877727 -0.5622104 -0.55877727 -0.5622104 -0.55877727 -0.5622104    Baja Productividad
```

```
> dim(testData)
[1] 74  7
```

Implementación del algoritmo K-NN en R statistics

Verificación de dimensiones de datos de entrenamiento y prueba

```
#=========================================================================================
# Removemos la variable tipo factor de los conjuntos de datos de entrenamiento y prueba
#=========================================================================================

trainData1 <- trainData[-7]
testData1 <- testData[-7]
#=========================================================================================

#=========================================================================================
# Verificamos las dimensiones de los conjuntos de datos de entrenamiento y prueba
#=========================================================================================

dim(trainData)
dim(trainData1)
dim(testData)
dim(testData1)

#=========================================================================================
```

```
> dim(trainData)
[1] 134   7
> dim(trainData1)
[1] 134   6
> dim(testData)
[1] 74  7
> dim(testData1)
[1] 74  6
```

Implementación del algoritmo K-NN en R statistics

Paso en el que se asegura que las etiquetas de entrenamiento y prueba sean tipo factor

```
#=============================================================================
# Paso 2: almacenar la variable objetivo para probar y entrenar datos:
# En este paso se debe asegurar que las etiquetas de entrenamiento y prueba
# sean datos de tipo de factor
#=============================================================================

data_productividad_train_labels <- trainData$Nivel_Productividad
dim(data_productividad_train_labels)
class(data_productividad_train_labels)

data_productividad_test_labels <- testData$Nivel_Productividad
dim(data_productividad_test_labels)
class(data_productividad_test_labels)

#=============================================================================
```

```
> data_productividad_train_labels <- trainData$Nivel_Productividad
> dim(data_productividad_train_labels)
NULL
> class(data_productividad_train_labels)
[1] "factor"
> data_productividad_test_labels <- testData$Nivel_Productividad
> dim(data_productividad_test_labels)
NULL
> class(data_productividad_test_labels)
[1] "factor"
```

Implementación del algoritmo K-NN en R statistics

Se implementa el algoritmo K-NN en R, mediante la función knn

```
#=====================================================================================
# Paso 3: entrenar un modelo con los datos
#=====================================================================================

install.packages("class") #Instalación de librerias
library(class)
data_procesos_test_pred1 <- knn(train = trainData1, test = testData1,
                                cl=data_procesos_train_labels, k = 3, prob=TRUE)
#KNN devuelve las etiquetas previstas para el conjunto de datos de prueba
#=====================================================================================
```

```
> data_productividad_test_pred1
 [1] Baja Productividad    Mediana Productividad Baja Productividad    Baja Productividad    Baja Productividad    Baja Productividad
 [7] Baja Productividad    Mediana Productividad Mediana Productividad Baja Productividad    Alta Productividad    Baja Productividad
[13] Baja Productividad    Mediana Productividad Mediana Productividad Mediana Productividad Mediana Productividad Baja Productividad
[19] Mediana Productividad Baja Productividad    Baja Productividad    Mediana Productividad Baja Productividad    Baja Productividad
[25] Baja Productividad    Baja Productividad    Baja Productividad    Alta Productividad    Baja Productividad    Baja Productividad
[31] Baja Productividad    Baja Productividad    Baja Productividad    Baja Productividad    Baja Productividad    Mediana Productividad
[37] Baja Productividad    Mediana Productividad Baja Productividad    Baja Productividad    Baja Productividad    Mediana Productividad
[43] Baja Productividad    Baja Productividad    Baja Productividad    Baja Productividad    Baja Productividad    Mediana Productividad
[49] Baja Productividad    Baja Productividad    Mediana Productividad Baja Productividad    Mediana Productividad Mediana Productividad
[55] Baja Productividad    Baja Productividad    Mediana Productividad Baja Productividad    Baja Productividad    Baja Productividad
[61] Alta Productividad    Mediana Productividad Baja Productividad    Mediana Productividad Baja Productividad    Mediana Productividad
[67] Baja Productividad    Baja Productividad    Mediana Productividad Baja Productividad    Mediana Productividad Mediana Productividad
[73] Baja Productividad    Baja Productividad
attr(,"prob")
 [1] 0.7500000 0.5000000 0.4444444 0.8888889 0.8750000 0.8888889 0.8750000 0.6000000 1.0000000 0.8888889 0.6666667 1.0000000 0.7857143
[14] 1.0000000 1.0000000 0.6000000 1.0000000 0.4444444 1.0000000 0.4444444 0.7500000 1.0000000 0.7500000 0.4444444 0.7500000 0.5000000
[27] 0.7857143 0.6666667 0.7857143 0.4375000 1.0000000 0.4444444 1.0000000 0.8750000 0.8888889 0.5714286 1.0000000 0.5000000 0.8333333
[40] 0.4375000 0.7857143 0.5000000 1.0000000 0.4285714 1.0000000 0.7500000 0.8000000 0.6666667 0.4285714 0.8750000 0.6000000 0.4375000
[53] 1.0000000 0.5000000 0.5000000 0.5000000 0.6666667 0.4375000 0.4000000 0.8750000 0.3333333 0.6000000 0.7500000 1.0000000 1.0000000
[66] 0.6666667 1.0000000 0.8888889 0.4285714 0.8888889 0.6000000 0.4285714 0.7500000 0.8750000
Levels: Alta Productividad Baja Productividad Mediana Productividad
```

Implementación del algoritmo K-NN en R statistics

Matriz de confusión

Matriz de confusión		Estimado por el modelo			
		Negativo (N)	*Positivo (P)*		
Real	*Negativo*	a: (TN)	b: (FP)		
	Positivo	c: (FN)	d: (TP)	**Precisión** ***("precision")*** Porcentaje predicciones positivas correctas:	d/(b+d)
		Sensibilidad, exhaustividad ***("Recall")*** Porcentaje casos positivos detectados	**Especifidad** ***(Specifity)*** Porcentaje casos negativos detectados	**Exactitud ("accuracy")** Porcentaje de predicciones correctas *(No sirve en datasets poco equilibrados)*	
		d/(d+c)	a/(a+b)	(a+d)/(a+b+c+d)	

Figura 4: Matriz de confusión y métricas asociadas

Implementación del algoritmo K-NN en R statistics

Métricas asociadas a la matriz de confusión

Exactitud

Si recordamos, la exactitud (o «accuracy«) representa el porcentaje de predicciones correctas frente al total. Por tanto, es el cociente entre los casos bien clasificados por el modelo (verdaderos positivos y verdaderos negativos, es decir, los valores en la diagonal de la matriz de confusión), y la suma de todos los casos.

Sin embargo, cuando un conjunto de datos es poco equilibrado, no es una métrica útil. Por ejemplo, si lo que intentamos predecir es una enfermedad rara, y nuestro algoritmo clasifica a todos los individuos como sanos, podría ser muy preciso (incluso un 99%), pero también, totalmente inútil.

Precisión

La **precisión,** (o"*precision*") se refiere a lo cerca que está el resultado de una predicción del valor verdadero. Por tanto, es el cociente entre los casos positivos bien clasificados por el modelo y el total de predicciones positivas.

Implementación del algoritmo K-NN en R statistics

Métricas asociadas a la matriz de confusión

Sensibilidad, exhaustividad

La sensibilidad (o recall) representa la tasa de verdaderos positivos (True Positive Rate) ó TP. Es la proporción entre los casos positivos bien clasificados por el modelo, respecto al total de positivos.

Especifidad

La especificidad, por su parte, es la tasa de verdaderos negativos, (“true negative rate”)o TN. Es la proporción entre los casos negativos bien clasificados por el modelo, respecto al total de negativos.

Matriz de confusión obtenida en R statistics

Matriz de confusión para el algoritmo K-NN en R, con k=3

```
> library(gmodels)
> validacion_modelo_productividad <- CrossTable(x = data_productividad_test_labels, y = data_productividad_test_pred1,prop.chisq=TRUE)


   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  74


                          | data_productividad_test_pred1
data_productividad_test_labels |    Alta Productividad |    Baja Productividad | Mediana Productividad |             Row Total |
-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
            Alta Productividad |                     2 |                     8 |                     8 |                    18 |
                               |                 2.211 |                 1.157 |                 1.034 |                       |
                               |                 0.111 |                 0.444 |                 0.444 |                 0.243 |
                               |                 0.667 |                 0.167 |                 0.348 |                       |
                               |                 0.027 |                 0.108 |                 0.108 |                       |
-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
            Baja Productividad |                     1 |                    35 |                     1 |                    37 |
                               |                 0.167 |                 5.042 |                 9.587 |                       |
                               |                 0.027 |                 0.946 |                 0.027 |                 0.500 |
                               |                 0.333 |                 0.729 |                 0.043 |                       |
                               |                 0.014 |                 0.473 |                 0.014 |                       |
-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
         Mediana Productividad |                     0 |                     5 |                    14 |                    19 |
                               |                 0.770 |                 4.353 |                11.095 |                       |
                               |                 0.000 |                 0.263 |                 0.737 |                 0.257 |
                               |                 0.000 |                 0.104 |                 0.609 |                       |
                               |                 0.000 |                 0.068 |                 0.189 |                       |
-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                  Column Total |                     3 |                    48 |                    23 |                    74 |
                               |                 0.041 |                 0.649 |                 0.311 |                       |
-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
```

Matriz de confusión obtenida en R statistics

Matriz de confusión para el algoritmo K-NN en R, con k=3, y sus respectivas métricas

```
> con=confusionMatrix(testData[,ncol(testData)], data_productividad_test_pred1, mode =  "everything")
> con
Confusion Matrix and Statistics

                     Reference
Prediction            Alta Productividad Baja Productividad Mediana Productividad
  Alta Productividad                   2                  8                    8
  Baja Productividad                   1                 35                    1
  Mediana Productividad                0                  5                   14

Overall Statistics

               Accuracy : 0.6892
                 95% CI : (0.571, 0.7917)
    No Information Rate : 0.6486
    P-Value [Acc > NIR] : 0.274045

                  Kappa : 0.4696

 Mcnemar's Test P-Value : 0.001076

Statistics by Class:

                     Class: Alta Productividad Class: Baja Productividad Class: Mediana Productividad
Sensitivity                            0.66667                   0.7292                        0.6087
Specificity                            0.77465                   0.9231                        0.9020
Pos Pred Value                         0.11111                   0.9459                        0.7368
Neg Pred Value                         0.98214                   0.6486                        0.8364
Precision                              0.11111                   0.9459                        0.7368
Recall                                 0.66667                   0.7292                        0.6087
F1                                     0.19048                   0.8235                        0.6667
Prevalence                             0.04054                   0.6486                        0.3108
Detection Rate                         0.02703                   0.4730                        0.1892
Detection Prevalence                   0.24324                   0.5000                        0.2568
Balanced Accuracy                      0.72066                   0.8261                        0.7553
```

Métricas del clasificador K-NN con k=3 en R statistics

```
#=========================================================================
# se almacena la matriz de confusion en una tabla
#=========================================================================

Tabla_matrix_productividad<-as.data.frame(matrix(data = con$table, nrow=3,ncol=3))
names(Tabla_matrix_productividad) <- c("Alta Productividad", "Baja Productividad", "Mediana Productividad")
row.names(Tabla_matrix_productividad) <- c("Alta Productividad", "Baja Productividad", "Mediana Productividad")
Tabla_matrix_productividad$Total <- 0
Tabla_matrix_productividad <- rbind(Tabla_matrix_productividad,c(0,0,0,0))
row.names(Tabla_matrix_productividad) <-c("Alta Productividad", "Baja Productividad", "Mediana Productividad", "Total")
Tabla_matrix_productividad$Total <- c(sum(Tabla_matrix_productividad[1,]), sum(Tabla_matrix_productividad[2,]), sum
(Tabla_matrix_productividad[3,]), 0)
Tabla_matrix_productividad[4,] <- c(sum(Tabla_matrix_productividad$`Alta Productividad`), sum(Tabla_matrix_productividad$`Baja
Productividad`), sum(Tabla_matrix_productividad$`Mediana Productividad`), sum(Tabla_matrix_productividad$Total))
#=========================================================================
```

```
#=========================================================================
# Tabla de metricas para las tres clases
#=========================================================================

Tabla_metricas_clase <- as.data.frame(matrix(data=t(con$byClass), nrow=11, ncol=3))
row.names(Tabla_metricas_clase) <- c("Sensitivity", "Specificity", "Pos Pred Value",
                                     "Neg Pred Value","Precision", "Recall", "F1",
                                     "Prevalence", "Detection Rate", "Detection Prevalence",
                                     "Balanced Accuracy" )
names(Tabla_metricas_clase) <- c("Alta Productividad", "Baja Productividad", "Mediana Productividad")

#=========================================================================
```

Métricas del clasificador K-NN con k=3 en R statistics

```
#=============================================================================
# Calculo de las métricas del clasificador
#=============================================================================

Accuracy <- con$overall[1]
clasificador_metricas <- data.frame("Metricas_del_Clasificador"=c(0,0,0,0,0,0))
k <- c(1,2,5,6,7)
for (i in 1:5)
{
  for (j in 1:3)
  {
    clasificador_metricas[i,1] <- clasificador_metricas[i,1] +
      (Tabla_matrix_productividad[j,4]/Tabla_matrix_productividad[4,4])*Tabla_metricas_clase[k[i],j]
  }

}
clasificador_metricas[6,1] <- Accuracy
row.names(clasificador_metricas) <- c("Sensitivity", "Specificity",
                                      "Precision", "Recall", "F1", "Accuracy" )
clasificador_metricas <- t(clasificador_metricas)

#=============================================================================
```

Implementación del algoritmo K-NN en R statistics

Validación del modelo a través de la matriz de confusión

```
#=================================================================================
# Matriz de Confusión

# Una vez obtenido el modelo podemos contrastar con el dataset de test su porcentaje
# de acierto, del 68,92%
#=================================================================================

con=confusionMatrix(testData[,ncol(testData)], data_productividad_test_pred1, mode =  "everything")

#=================================================================================
```

Implementación del algoritmo K-NN en R statistics – k = 3

Validación del modelo a través de la matriz de confusión

```
> con
Confusion Matrix and Statistics

                     Reference
Prediction            Alta Productividad Baja Productividad Mediana Productividad
  Alta Productividad                   2                  8                    8
  Baja Productividad                   1                 35                    1
  Mediana Productividad                0                  5                   14

Overall Statistics

               Accuracy : 0.6892
                 95% CI : (0.571, 0.7917)
    No Information Rate : 0.6486
    P-Value [Acc > NIR] : 0.274045

                  Kappa : 0.4696

 Mcnemar's Test P-Value : 0.001076

Statistics by Class:

                     Class: Alta Productividad Class: Baja Productividad Class: Mediana Productividad
Sensitivity                            0.66667                   0.7292                       0.6087
Specificity                            0.77465                   0.9231                       0.9020
Pos Pred Value                         0.11111                   0.9459                       0.7368
Neg Pred Value                         0.98214                   0.6486                       0.8364
Precision                              0.11111                   0.9459                       0.7368
Recall                                 0.66667                   0.7292                       0.6087
F1                                     0.19048                   0.8235                       0.6667
Prevalence                             0.04054                   0.6486                       0.3108
Detection Rate                         0.02703                   0.4730                       0.1892
Detection Prevalence                   0.24324                   0.5000                       0.2568
Balanced Accuracy                      0.72066                   0.8261                       0.7553
```

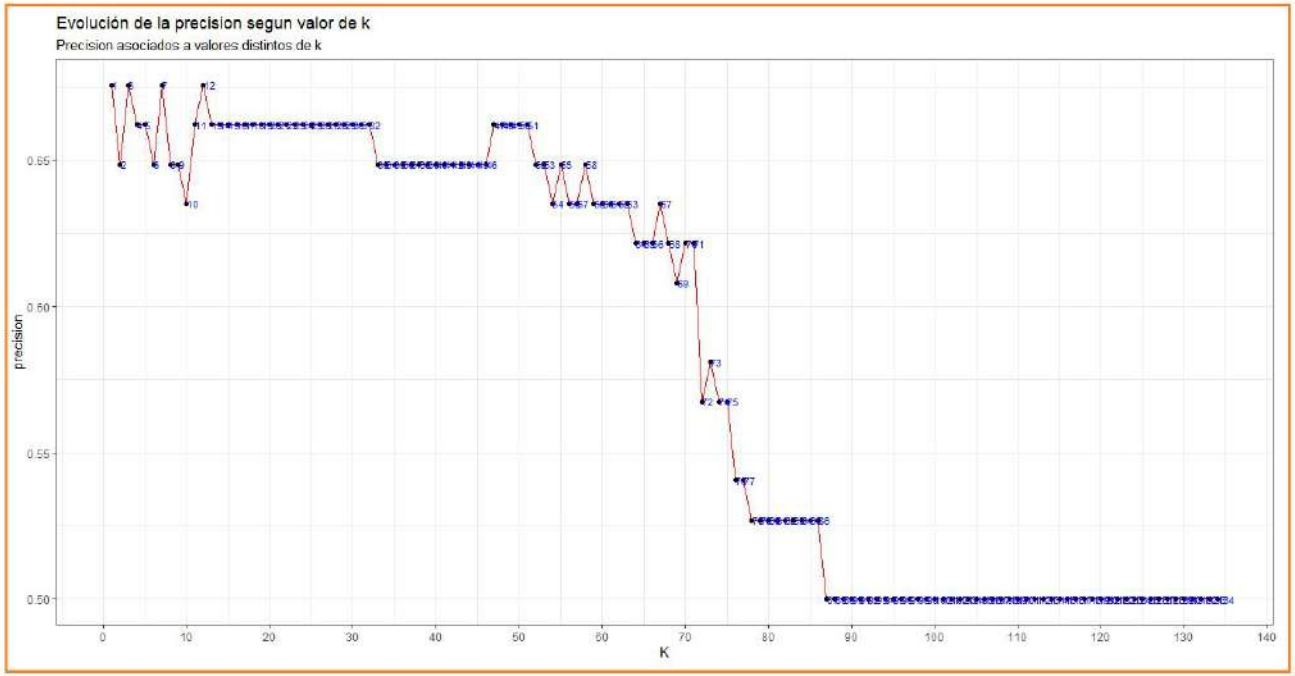
Calculo del valor de K óptimo para el K-NN en R statistics

Hasta el momento solo hemos utilizado un solo valor para k, 3 vecinos más cercanos. Sin embargo, podríamos ver la exactitud o *accuracy* para diferentes valores de k. Como tenemos 134 valores en nuestra data de entrenamiento veremos los aciertos tomando máximo 134 vecinos más cercanos.

```
#=============================================================================

k <- 1:134
resultado <- data.frame(k, precision = 0)

for(n in k){
  productividad_test_output_kNN <- knn(train =  trainData1,
                                       cl = data_productividad_train_labels,
                                       test = testData1,
                                       k = n)

  resultado$precision[n] <- mean(productividad_test_output_kNN == testData[,7])
}
resultado
library(extrafont)
#resultado %>%
ggplot(resultado,aes(x=k, y=precision))+
  labs(title = 'Evolución de la precision segun valor de k')+
  labs(subtitle = 'Precision asociados a valores distintos de k')+
  geom_line(color="red")+ geom_point()+
  scale_x_continuous(breaks=pretty(resultado$k, n = 10), name="K")+
  theme(legend.position="top", legend.text = element_text(colour="blue", size=6)) +
  geom_text(aes(label=k),hjust = 0, nudge_x = 0.05, color="blue",size=3)+
  theme(text=element_text(size=12, family="Calibri"))+
  theme_bw()
#=============================================================================
```

Calculo del valor de K óptimo para el K-NN en R statistics

Precisión del modelo para diferentes valores de k



Como vemos, para este caso, a partir de cierto número de vecinos más cercanos el nivel de aciertos de nuestro algoritmo empieza a reducirse. Dependerá de cada caso el escoger el mejor “k” para nuestro modelo, para nuestro caso k=3, ofrece la mejor precisión para nuestro modelo.

Matriz de confusión obtenida en R statistics

Matriz de confusión para el algoritmo K-NN en R, con k=7, y sus respectivas métricas

```
> con
Confusion Matrix and Statistics

                          Reference
Prediction                 Alta Productividad Baja Productividad Mediana Productividad
  Alta Productividad                        1                  8                     9
  Baja Productividad                        0                 36                     1
  Mediana Productividad                     0                  7                    12

Overall Statistics

               Accuracy : 0.6622
                 95% CI : (0.5428, 0.7681)
    No Information Rate : 0.6892
    P-Value [Acc > NIR] : 0.738

                  Kappa : 0.4133

 Mcnemar's Test P-Value : 8.288e-05

Statistics by Class:

                     Class: Alta Productividad Class: Baja Productividad Class: Mediana Productividad
Sensitivity                            1.00000                   0.7059                       0.5455
Specificity                            0.76712                   0.9565                       0.8654
Pos Pred Value                         0.05556                   0.9730                       0.6316
Neg Pred Value                         1.00000                   0.5946                       0.8182
Precision                              0.05556                   0.9730                       0.6316
Recall                                 1.00000                   0.7059                       0.5455
F1                                     0.10526                   0.8182                       0.5854
Prevalence                             0.01351                   0.6892                       0.2973
Detection Rate                         0.01351                   0.4865                       0.1622
Detection Prevalence                   0.24324                   0.5000                       0.2568
Balanced Accuracy                      0.88356                   0.8312                       0.7054
```

Implementación del algoritmo K-NN en R statistics

Evaluación del rendimiento del modelo: métricas asociadas al clasificador K-NN, con k=3

```
> clasificador_metricas
                          Sensitivity Specificity Precision    Recall        F1  Accuracy
Metricas_del_Clasificador   0.6830322   0.8815509 0.6891892 0.6830322 0.6292679 0.6891892
```

Implementación del algoritmo K-NN en R statistics

Análisis de resultados

Hemos clasificado 2 centros de Alta Productividad como Alta Productividad, lo que indica que está bien clasificado, 8 centros que siendo de Alta Productividad como de Baja Productividad y 8 centros que siendo de Alta productividad como de Mediana Productividad, lo que indica una incorrecta clasificación para estos. A su vez analizando los centros de Baja Productividad tenemos que la predicción en la clasificación arroja que 35 centros de Baja Productividad quedaron clasificados como de Baja Productividad lo cual indica que quedaron bien clasificados. Por otro lado, 1 centro siendo de Baja Productividad quedo clasificado como que fuera de Alta Productividad, asimismo otro centro siendo de Baja Productividad quedo clasificado en Mediana Productividad, lo que indica una incorrecta clasificación para estos. Por último 14 centros de Mediana Productividad fueron clasificados como de Mediana Productividad, lo cual indica una correcta clasificación y 5 centros pertenecientes a Mediana Productividad quedaron clasificados como si fueran de Baja Productividad, lo cual indica que esta incorrectamente clasificado.

Implementación del algoritmo K-NN en R statistics

Conclusiones

Sensibilidad de acuerdo al valor obtenido del 0.6830322, esto me indica **¿qué porcentaje de los centros de acuerdo a su nivel de productividad somos capaces de identificar?** Para lo cuál diríamos que el 68,30% de los centros se identifican acertadamente de acuerdo al nivel de productividad.

El valor **F1** se utiliza para combinar las medidas de **precision** y **recall (sensibilidad)** en un sólo valor. Asume que nos importa de igual forma la precisión y la exhaustividad. Esto no tiene que ser así en todos los problemas. Ya que necesitamos predecir si el nivel de productividad del centro se detecto adecuadamente, por lo tanto podemos valorar más la sensibilidad a riesgo de tener una peor precisión. Sin embargo para nuestro caso el F1 es igual 0,6292679 equivalente al 62,93% que no es muy aproximado al 68,30% de sensibilidad del modelo y a la precisión del modelo equivalente a 68,92%. Sin embargo el F1 permite comparar modelos donde si un modelo arroja un F1 alto representa el modelo que mejor puede clasificar las observaciones en clases.

Para nuestro caso un F1 alto indicaría que el modelo sería capaz de clasificar las observaciones de la clase nivel de productividad de manera más precisa.

Implementación del algoritmo K-NN en R statistics

Conclusiones

Exactitud (Accuracy): La exactitud (accuracy) mide el porcentaje de casos que el modelo ha acertado. Es decir, el modelo acierta el 68,92% de las veces. Como se observa, la exactitud es una métrica muy engañosa. De hecho, si tuviésemos un modelo que siempre predijera que el centro siempre va a estar en un nivel de productividad determinado, su accuracy podría ser del 80% o más y esto normalmente se observa en problemas con clases desbalanceadas por lo tanto es mucho mejor usar precision, recall y F1. Estas métricas dan una mejor idea de la calidad del modelo. IArtificial.net. (2020).

Referencias

IArtificial.net. (2020). https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/

Especificidad: el modelo tiene una alta capacidad para identificar los centros que no fueron clasificados correctamente según su nivel de productividad entre todos los centros cuyos niveles de productividad fueron clasificados correctamente. Para nuestro caso la especificidad del clasificador es de 88,16%.

Implementación del algoritmo K-NN en R statistics

Conclusiones

De lo anterior tenemos, el modelo es mucho más específico que sensible en la clasificación esto se traduce en que si un falso positivo significa que el modelo tiene una muy alta capacidad para identificar los centros que no fueron clasificados correctamente según su nivel de productividad evitando los falsos positivos.

GC-F-004v.05

7. Modelo de analítica de datos para medición de Gestión por centros – Algoritmo Random Forest

Funcionamiento del algoritmo Random Forest en R statistics

Random Forest

- Se considera como la “panacea” en todos los problemas de ciencia de datos.
- Útil para regresión y clasificación.
- Un grupo de modelos “débiles”, se combinan en un modelo robusto.
- Sirve como una técnica para reducción de la dimensionalidad.
- Se generan múltiples árboles (a diferencia de CART: Comparación entre árboles de regresión).
- Cada árbol da una clasificación (vota por una clase). Y el resultado es la clase con mayor número de votos en todo el **random forest.**
- Para regresión, se toma el promedio de las salidas (predicciones) de todos los árboles.

¿Cómo se construye un modelo random forest?

Cada árbol se construye así:

- Dado que el número de casos en el conjunto de entrenamiento es N. Una muestra de esos N casos se toma aleatoriamente pero **CON REEMPLAZO**. Esta muestra será el conjunto de entrenamiento para construir el árbol i.

Funcionamiento del algoritmo Random Forest en R statistics

¿Cómo se construye un modelo random forest?

Cada árbol se construye así:

- Si existen M variables de entrada, un número $m < M$ se especifica tal que para cada nodo, m variables se seleccionan aleatoriamente de M. La mejor división de estos m atributos es usado para ramificar el árbol. El valor m se mantiene constante durante la generación de todo el bosque.

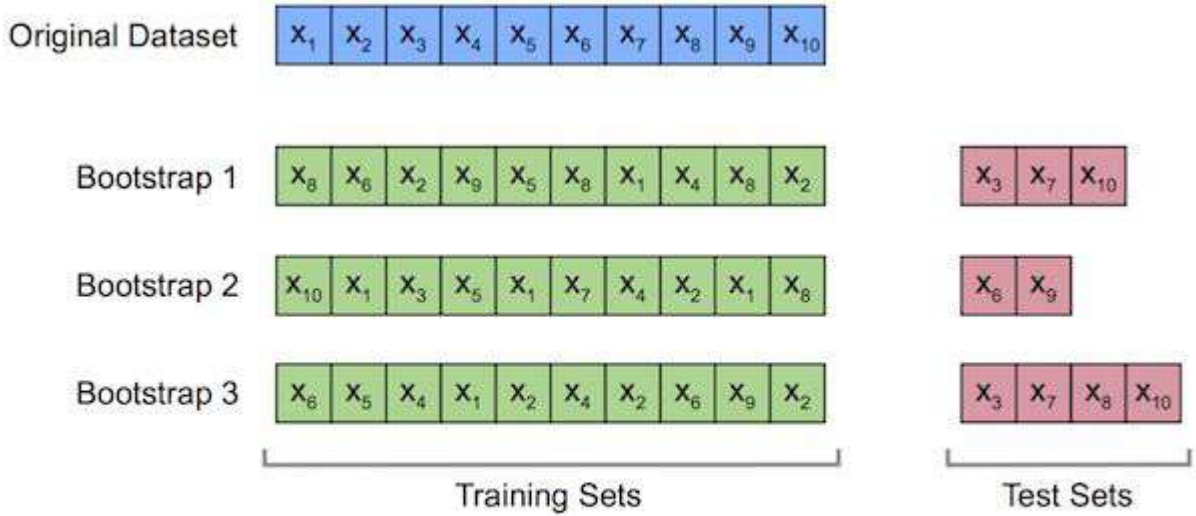
- Cada árbol crece hasta su máxima extensión posible y **NO hay proceso de poda**.

- Nuevas instancias se predicen a partir de la agregación de las predicciones de los x árboles (i.e., mayoría de votos para clasificación, promedio para regresión)

Funcionamiento del algoritmo Random Forest en R statistics

Out of bag samples y out of bag error (Muestras fuera de bolsa y error fuera de bolsa)

- El proceso de muestreo de los datos con reemplazo se denomina **bootstrap**.
- Un tercio de los datos no se usan para el entrenamiento y pueden ser usados para test.
- Este conjunto se denomina out of bag (OOB) samples.

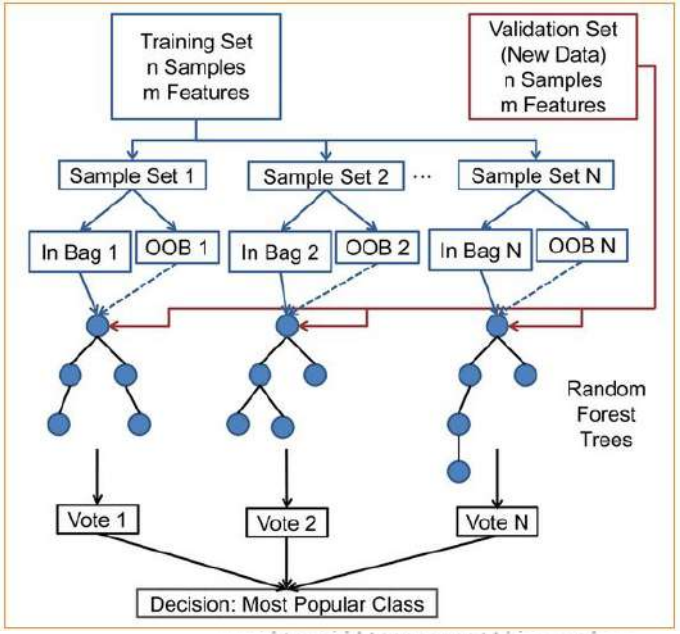


This work by Sebastian Raschka is licensed under a
Creative Commons Attribution 4.0 International License.

Funcionamiento del algoritmo Random Forest en R statistics

Out of bag samples y out of bag error (Muestras fuera de bolsa y error fuera de bolsa)

- El error estimado en estos out of bag samples se conoce como out of bag error (OOB error)
- Usar este conjunto de test (OOB) es tan preciso como si se usara un conjunto de test del mismo tamaño que el de entrenamiento.
- Sería posible no usar un conjunto de test adicional.



Funcionamiento del algoritmo Random Forest en R statistics

Hyper-parámetros

El hyper-parámetro más importante para ajustar es el número de variables candidatas a seleccionar para evaluar cada ramificación. Sin embargo, existen algunos adicionales que deben considerarse. Independientemente de los nombres en las distintas librerías, al menos los siguientes deberían estar presentes.

- **ntree:** número de árboles en el bosque. Se quiere estabilizar el error, pero usar demasiados árboles puede ser innecesariamente ineficiente.

- **mtry:** número de variables aleatorias como candidatas en cada ramificación.

- **sampsize:** el número de muestras sobre las cuales se va a entrenar. El valor por defecto es 63.25%. Valores más bajos podrían introducir sesgo y reducir el tiempo. Valores más altos podrían incrementar el rendimiento del modelo pero a riesgo de causar overfitting (sobreajuste). **Generalmente se mantiene en el rango 60-80%.**

- **nodesize:** mínimo número de muestras dentro de los nodos terminales. Equilibrio entre bias-varianza

Implementación del algoritmo Random Forest en R statistics

Ventajas de Random Forest

- Existen muy pocas suposiciones y por lo tanto la preparación de los datos es mínima.
- Puede manejar hasta miles de variables de entrada e identificar las más significativas. Método de reducción de dimensionalidad.
- Una de las salidas del modelo es la importancia de variables.
- Incorpora métodos efectivos para estimar valores faltantes.
- Es posible usarlo como método no supervisado (clustering) y detección de outliers.

Desventajas de Random Forest

- Pérdida de interpretación
- Bueno para clasificación, no tanto para regresión. Las predicciones no son de naturaleza continua.
- En regresión, no puede predecir más allá del rango de valores del conjunto de entrenamiento.
- Poco control en lo que hace el modelo (modelo caja negra para modeladores estadísticos)

Implementación del algoritmo Random Forest en R statistics

Lectura de base de datos Tabla_PProceso.xlsx

```
#=============================================================================
# Cargamos la base de datos procesos: Tabla_PProceso.xlsx
#=============================================================================

#=============================================================================
setwd("E:/Documents/Codigo en R/SENA paola")
data_procesos <- read_xlsx("Tabla_PProceso.xlsx")
View(data_procesos)
#=============================================================================
```

Filter

	Centro_Año	Q_Proyectos	QNA_5007	QNA_5009	QNA_5010	TV_Proyectos	Cluster	Gestion
1	9101-2010	1	1	2	0	465292449	1	Baja Gestion
2	9127-2010	2	2	4	1	392005181	1	Baja Gestion
3	9201-2010	1	4	1	1	455662749	1	Baja Gestion
4	9202-2010	3	5	5	2	407185228	3	Alta Gestion
5	9203-2010	2	4	3	1	290617466	1	Baja Gestion
6	9204-2010	3	5	7	1	564001653	2	Media Gestion
7	9205-2010	1	3	1	1	508395371	1	Baja Gestion
8	9206-2010	1	3	2	1	300860419	1	Baja Gestion
9	9301-2010	3	6	6	0	652146153	2	Media Gestion
10	9401-2010	3	2	4	1	584555799	1	Baja Gestion
11	9402-2010	2	2	6	2	469064097	1	Baja Gestion
12	9501-2010	2	1	4	1	550483170	1	Baja Gestion
13	9502-2010	3	3	7	1	518339355	2	Media Gestion
14	9503-2010	2	1	4	2	385085638	1	Baja Gestion
15	9504-2010	1	1	3	0	356850041	1	Baja Gestion
16	9549-2010	2	1	3	1	362870121	1	Baja Gestion
17	9101-2011	3	5	6	1	541412297	2	Media Gestion

Showing 1 to 18 of 208 entries, 8 total columns

Implementación del algoritmo Random Forest en R statistics

Exploración de los datos

```
#=============================================================================
# Primer paso: Exploración de los datos
#=============================================================================

#=============================================================================
summary(data_procesos)
str(data_procesos)
#=============================================================================
```

```
> summary(data_procesos)
  Centro_Año         Q_Proyectos       QNA_5007        QNA_5009        QNA_5010      TV_Proyectos          Cluster
 Length:208         Min.   :1.000   Min.   : 0.000   Min.   : 0.00   Min.   :0.000   Min.   :290617466   Min.   :1.000
 Class :character   1st Qu.:2.000   1st Qu.: 2.000   1st Qu.: 2.00   1st Qu.:0.000   1st Qu.:394601862   1st Qu.:1.000
 Mode  :character   Median :2.000   Median : 3.000   Median : 4.00   Median :1.000   Median :456991575   Median :1.000
                    Mean   :2.404   Mean   : 3.668   Mean   : 4.13   Mean   :1.255   Mean   :480505237   Mean   :1.639
                    3rd Qu.:3.000   3rd Qu.: 5.000   3rd Qu.: 6.00   3rd Qu.:2.000   3rd Qu.:541300336   3rd Qu.:2.000
                    Max.   :6.000   Max.   :14.000   Max.   :13.00   Max.   :5.000   Max.   :932664228   Max.   :3.000
   Gestion
 Length:208
 Class :character
 Mode  :character


> str(data_procesos)
tibble [208 × 8] (S3: tbl_df/tbl/data.frame)
 $ Centro_Año  : chr [1:208] "9101-2010" "9127-2010" "9201-2010" "9202-2010" ...
 $ Q_Proyectos : num [1:208] 1 2 1 3 2 3 1 1 3 3 ...
 $ QNA_5007    : num [1:208] 1 2 4 5 4 5 3 3 6 2 ...
 $ QNA_5009    : num [1:208] 2 4 1 5 3 7 1 2 6 4 ...
 $ QNA_5010    : num [1:208] 0 1 1 2 1 1 1 1 0 1 ...
 $ TV_Proyectos: num [1:208] 4.65e+08 3.92e+08 4.56e+08 4.07e+08 2.91e+08 ...
 $ Cluster     : num [1:208] 1 1 1 3 1 2 1 1 2 1 ...
 $ Gestion     : chr [1:208] "Baja Gestion" "Baja Gestion" "Baja Gestion" "Alta Gestion" ...
```

Implementación del algoritmo Random Forest en R statistics

Se realiza escalamiento de datos

```
#=============================================================================
#Escalamiento de datos
#=============================================================================
datos_TPProceso <- scale(data_procesos[,c(2:6)])
View(datos_TPProceso)
#Se adiciona nuevamente la columna de etiqueta Centro-Año, Cluster y Gestion
datos_scale_TPProceso <- as.data.frame(cbind(data_procesos$Centro_Año, datos_TPProceso, data_procesos$Cluster, data_procesos$Gestion))
View(datos_scale_TPProceso)
str(datos_scale_TPProceso)
dim(datos_scale_TPProceso)
datos_scale_TPProceso$Q_Proyectos <- as.numeric(datos_scale_TPProceso$Q_Proyectos)
datos_scale_TPProceso$QNA_5007 <- as.numeric(datos_scale_TPProceso$QNA_5007)
datos_scale_TPProceso$QNA_5009 <- as.numeric(datos_scale_TPProceso$QNA_5009)
datos_scale_TPProceso$QNA_5010 <- as.numeric(datos_scale_TPProceso$QNA_5010)
datos_scale_TPProceso$TV_Proyectos <- as.numeric(datos_scale_TPProceso$TV_Proyectos)
str(datos_scale_TPProceso)
names(datos_scale_TPProceso) <- names(data_procesos)
View(datos_scale_TPProceso)
row.names(datos_scale_TPProceso) <- datos_scale_TPProceso$Centro_Año
View(datos_scale_TPProceso)
proceso <- datos_scale_TPProceso[, c(-1,-7)]
View(proceso)
str(proceso)
#=============================================================================
```

Implementación del algoritmo Random Forest en R statistics

Convertimos la variable etiquetas en una variable tipo categórico

```
#=============================================================================
# Segundo paso: convertimos la variable de clasificación Gestión en factor
# Con el propósito de identificarla por niveles, pero primero convertimos los datos
# en data.frame
#=============================================================================

#=============================================================================
proceso <- as.data.frame(proceso)
str(proceso)
proceso$Gestion <- as.factor(proceso$Gestion)
str(proceso)
View(proceso)
#=============================================================================
```

```
> str(proceso)
'data.frame':   208 obs. of  6 variables:
 $ Q_Proyectos : num  -1.257 -0.362 -1.257 0.534 -0.362 ...
 $ QNA_5007    : num  -1.103 -0.689 0.137 0.55 0.137 ...
 $ QNA_5009    : num  -0.8238 -0.0502 -1.2106 0.3366 -0.437 ...
 $ QNA_5010    : num  -1.108 -0.225 -0.225 0.658 -0.225 ...
 $ TV_Proyectos: num  -0.126 -0.733 -0.206 -0.607 -1.573 ...
 $ Gestion     : Factor w/ 3 levels "Alta Gestion",..: 2 2 2 1 2 3 2 2 3 2 ...
```

Implementación del algoritmo Random Forest en R statistics

Se realiza escalamiento de datos

```
> str(proceso)
'data.frame':   208 obs. of  6 variables:
 $ Q_Proyectos : num  -1.257 -0.362 -1.257 0.534 -0.362 ...
 $ QNA_5007    : num  -1.103 -0.689 0.137 0.55 0.137 ...
 $ QNA_5009    : num  -0.8238 -0.0502 -1.2106 0.3366 -0.437 ...
 $ QNA_5010    : num  -1.108 -0.225 -0.225 0.658 -0.225 ...
 $ TV_Proyectos: num  -0.126 -0.733 -0.206 -0.607 -1.573 ...
 $ Gestion     : Factor w/ 3 levels "Alta Gestion",..: 2 2 2 1 2 3 2 2 3 2 ...
```

	Q_Proyectos	QNA_5007	QNA_5009	QNA_5010	TV_Proyectos	Gestion
9101-2010	-1.2572745	-1.1026653	-0.82382836	-1.1079853	-0.12603990	Baja Gestion
9127-2010	-0.3616817	-0.6894141	-0.05021076	-0.2249932	-0.73323429	Baja Gestion
9201-2010	-1.2572745	0.1370881	-1.21063716	-0.2249932	-0.20582319	Baja Gestion
9202-2010	0.5339111	0.5503392	0.33659804	0.6579990	-0.60746565	Alta Gestion
9203-2010	-0.3616817	0.1370881	-0.43701956	-0.2249932	-1.57324448	Baja Gestion
9204-2010	0.5339111	0.5503392	1.11021564	-0.2249932	0.69177849	Media Gestion
9205-2010	-1.2572745	-0.2761630	-1.21063716	-0.2249932	0.23107333	Baja Gestion
9206-2010	-1.2572745	-0.2761630	-0.82382836	-0.2249932	-1.48838030	Baja Gestion
9301-2010	0.5339111	0.9635904	0.72340684	-1.1079853	1.42206695	Media Gestion
9401-2010	0.5339111	-0.6894141	-0.05021076	-0.2249932	0.86207222	Baja Gestion
9402-2010	-0.3616817	-0.6894141	0.72340684	0.6579990	-0.09479131	Baja Gestion
9501-2010	-0.3616817	-1.1026653	-0.05021076	-0.2249932	0.57977613	Baja Gestion
9502-2010	0.5339111	-0.2761630	1.11021564	-0.2249932	0.31346051	Media Gestion
9503-2010	-0.3616817	-1.1026653	-0.05021076	0.6579990	-0.79056358	Baja Gestion
9504-2010	-1.2572745	-1.1026653	-0.43701956	-1.1079853	-1.02449911	Baja Gestion
9549-2010	-0.3616817	-1.1026653	-0.43701956	-0.2249932	-0.97462198	Baja Gestion
9101-2011	0.5339111	0.5503392	0.72340684	-0.2249932	0.50462279	Media Gestion

Showing 1 to 18 of 208 entries, 6 total columns

Implementación del algoritmo Random Forest en R statistics

Obtenemos la distribución de clases, ya que este es un problema multiclase, podemos usar el generador de tablas de datos de clasificación múltiple, para evidenciar la distribución de datos para cada clase

```
#=============================================================================
table(proceso$Gestion)
#=============================================================================
```

```
> table(proceso$Gestion)

 Alta Gestion  Baja Gestion Media Gestion
           47           122            39
```

Porcentaje de centros por distribución para cada clase

```
#=============================================================================
round(prop.table(table(proceso$Gestion)) * 100,digits = 1)
#=============================================================================
```

```
> round(prop.table(table(proceso$Gestion)) * 100,digits = 1)

 Alta Gestion  Baja Gestion Media Gestion
         22.6          58.7          18.8
```

Implementación del algoritmo Random Forest en R statistics

Analítica de datos

Primer paso: Dividimos los datos en datos de prueba y datos de entrenamiento
Definimos un índice para dividir los dos conjuntos de datos, se creó un conjunto de entrenamiento del 70% para entrenar el modelo y un 30% restante como datos de validación para probar el modelo.

```
#=================================================================================
ind <- sample(2, nrow(proceso), replace=TRUE, prob=c(0.7, 0.3))
# Valores asignados para las filas (1: Entrenamiento, 2: Prueba)
trainData <- proceso[ind==1,]
testData <-  proceso[ind==2,]
dim(trainData)
dim(testData)
#=================================================================================
```

```
> dim(trainData)
[1] 146   6
> # [1]  146  6
> dim(testData)
[1] 62  6
```

Implementación del algoritmo Random Forest en R statistics

Se carga la libreria (randomForest)

```
#=============================================================================
library(randomForest)
#=============================================================================

#=============================================================================
proceso_classifier <- randomForest(Gestion ~.,
                                   data = trainData, # Conjunto de datos de entrenamiento
                                   importance = T)
proceso_classifier                 # Matrix de confusion: Evaluacion de la prediccion
#=============================================================================
```

```
> proceso_classifier

Call:
 randomForest(formula = Gestion ~ ., data = trainData, importance = T)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 3.42%
Confusion matrix:
              Alta Gestion Baja Gestion Media Gestion class.error
Alta Gestion            33            0             1  0.02941176
Baja Gestion             0           82             1  0.01204819
Media Gestion            3            0            26  0.10344828
```

Implementación del algoritmo Random Forest en R statistics

Se carga la libreria (randomForest)

```
#=================================================================================
library(randomForest)
#=================================================================================

#=================================================================================
proceso_classifier <- randomForest(Gestion ~.,
                                   data = trainData, # Conjunto de datos de entrenamiento
                                   importance = T)
proceso_classifier                 # Matrix de confusion: Evaluacion de la prediccion
#=================================================================================
```

```
#=================================================================================
# Encontramos las características más importantes del nivel de Gestion para los centros
importance = importance(proceso_classifier)
varImportance = data.frame(Variables = row.names(importance),
                           Importance = round(importance[, "MeanDecreaseAccuracy"],2))

rankImportance = varImportance%>%mutate(Rank=paste('#',dense_rank(desc(Importance))))

ggplot(rankImportance,aes(x=reorder(Variables,Importance),
  y=Importance,fill=Importance)) +
  geom_bar(stat='identity') +
  geom_text(aes(x = Variables, y = 0.5, label = Rank),
            hjust=0, vjust=0.55, size = 4, colour = 'white') +
  labs(x = 'Variables') +
  coord_flip() +
  theme_classic()
#=================================================================================
```

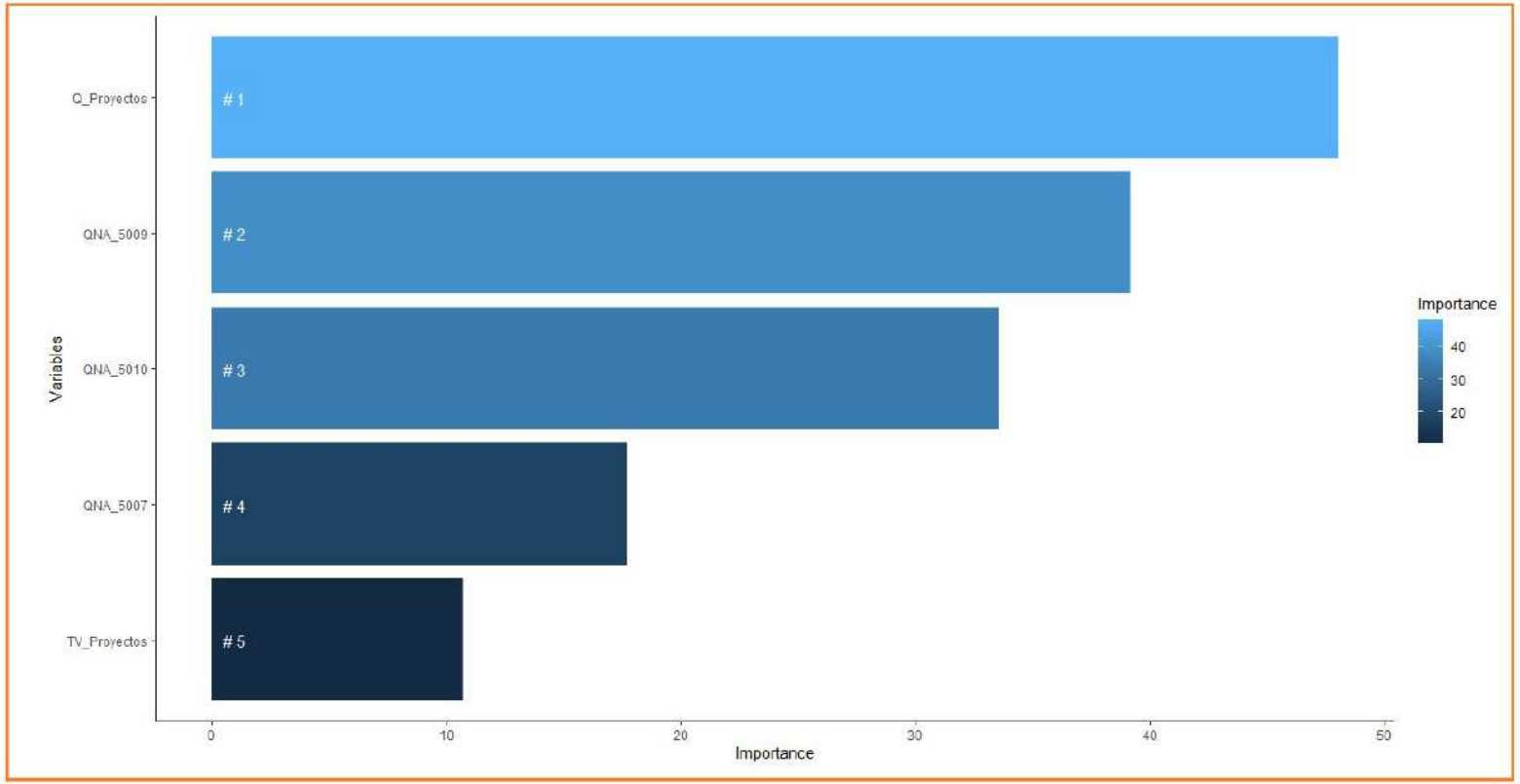
Implementación del algoritmo Random Forest en R statistics

Nivel de importancia de los predictores

```
> importance
              Alta Gestion Baja Gestion Media Gestion MeanDecreaseAccuracy MeanDecreaseGini
Q_Proyectos      39.451556    35.038121     30.040761             48.05972         30.82437
QNA_5007          8.525961    10.799348     13.637488             17.71016         11.61786
QNA_5009         12.985370    10.648546     44.295556             39.20556         18.73913
QNA_5010         36.863203    11.371012      7.626616             33.59858         11.86977
TV_Proyectos      7.074760     8.366456      3.638730             10.73385         11.01981
```

Implementación del algoritmo Random Forest en R statistics

Se entrega el gráfico donde se evidencia las variables con mayor importancia en el modelo



Implementación del algoritmo Random Forest en R statistics

Matriz de confusión para validar el modelo

Tercer paso: El modelo es relativamente preciso. Sin embargo, ahora sabemos qué tan bien se desempeña el modelo en el conjunto de entrenamiento, necesitamos validar esto con una matriz de confusión. Random Forest muestra que ha sido entrenado en más de > 2 clases, por lo que pasa de un modelo binario a un modelo de clasificación múltiple. Las funciones contenidas en el paquete funcionan con métodos binarios y de clasificación múltiple.

Uso de la matriz de confusión del paquete CARET: La matriz de confusión almacena una serie de elementos de la lista que deben utilizarse juntos para comparar el rendimiento del ajuste del modelo a lo largo del tiempo para asegurarse de que no haya un desfase de características subyacentes ni una regresión en el rendimiento. Para evitar lo anterior, hacemos una predicción sobre el modelo ajustado con los datos de prueba

```
#==========================================================================
rf_class <- predict(proceso_classifier, newdata = testData)
predictions <- cbind(data.frame(train_preds=rf_class,
                                testData$Gestion))
#==========================================================================
```

Implementación del algoritmo Random Forest en R statistics

Matriz de confusión para validar el modelo

Cuarto paso: Crear un objeto de matriz de confusión

```
#==============================================================================
cm <- caret::confusionMatrix(predictions$train_preds, predictions$testData.Gestion)
print(cm)
#==============================================================================
```

```
> cm <- caret::confusionMatrix(predictions$train_preds, predictions$testData.Gestion)
> print(cm)
Confusion Matrix and Statistics

                Reference
Prediction       Alta Gestion Baja Gestion Media Gestion
  Alta Gestion             12            2             0
  Baja Gestion              0           37             0
  Media Gestion             1            0            10

Overall Statistics

               Accuracy : 0.9516
                 95% CI : (0.865, 0.9899)
    No Information Rate : 0.629
    P-Value [Acc > NIR] : 2.784e-09

                  Kappa : 0.9118

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Alta Gestion Class: Baja Gestion Class: Media Gestion
Sensitivity                       0.9231              0.9487               1.0000
Specificity                       0.9592              1.0000               0.9808
Pos Pred Value                    0.8571              1.0000               0.9091
Neg Pred Value                    0.9792              0.9200               1.0000
Prevalence                        0.2097              0.6290               0.1613
Detection Rate                    0.1935              0.5968               0.1613
Detection Prevalence              0.2258              0.5968               0.1774
Balanced Accuracy                 0.9411              0.9744               0.9904
```

Implementación del algoritmo Random Forest en R statistics

Se almacena la matriz de confusión en una tabla

```
Tabla_matrix<-as.data.frame(matrix(data = cm$table, nrow=3,ncol=3))
names(Tabla_matrix) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")
row.names(Tabla_matrix) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")
Tabla_matrix$Total <- 0
Tabla_matrix <- rbind(Tabla_matrix,c(0,0,0,0))
row.names(Tabla_matrix) <-c("Alta Gestion", "Baja Gestion", "Media Gestion", "Total")
Tabla_matrix$Total <- c(sum(Tabla_matrix[1,]), sum(Tabla_matrix[2,]), sum(Tabla_matrix[3,]), 0)
Tabla_matrix[4,] <- c(sum(Tabla_matrix$`Alta Gestion`), sum(Tabla_matrix$`Baja Gestion`), sum(Tabla_matrix$`Media
Gestion`), sum(Tabla_matrix$Total))
```

Tabla de métricas para las tres clases

```
Tabla_metricas_clase <- as.data.frame(matrix(data=t(cm$byClass), nrow=11, ncol=3))
row.names(Tabla_metricas_clase) <- c("Sensitivity", "Specificity", "Pos Pred Value",
                                     "Neg Pred Value","Precision", "Recall", "F1",
                                     "Prevalence", "Detection Rate", "Detection Prevalence",
                                     "Balanced Accuracy" )
names(Tabla_metricas_clase) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")
```

Implementación del algoritmo Random Forest en R statistics

Cálculo de las métricas del clasificador

```
Accuracy <- cm$overall[1]
clasificador_metricas <- data.frame("Metricas_del_Clasificador"=c(0,0,0,0,0,0))
k <- c(1,2,5,6,7)
for (i in 1:5)
{
  for (j in 1:3)
  {
    clasificador_metricas[i,1] <- clasificador_metricas[i,1] +
      (Tabla_matrix[j,4]/Tabla_matrix[4,4])*Tabla_metricas_clase[k[i],j]
  }

}
clasificador_metricas[6,1] <- Accuracy
row.names(clasificador_metricas) <- c("Sensitivity", "Specificity",
                                      "Precision", "Recall", "F1", "Accuracy" )
clasificador_metricas <- t(clasificador_metricas)
```

```
> clasificador_metricas
                        Sensitivity Specificity Precision    Recall        F1  Accuracy
Metricas_del_Clasificador   0.9520265   0.9873715 0.9516129 0.9520265 0.9507573 0.9516129
```

GC-F-004v.05

8. Modelo de analítica de datos para medir el nivel de productividad por centros – Algoritmo Random Forest

Implementación del algoritmo Random Forest en R statistics

Lectura de base de datos Tabla_Nivel_Productividad.xlsx

```
#=============================================================================
# Base de datos procesos: Tabla_Nivel_Productividad.xlsx
#=============================================================================
setwd("E:/Documents/Codigo en R/SENA paola")
data_productividad <- read_xlsx("Tabla_Nivel_Productividad.xlsx")
View(data_productividad)
#=============================================================================
```

Filter

	Centro_Año	Cantidad_Proyectos	Fecha_inicio_Proyecto	Cantidad_Productos	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031	Cluster	Nivel_Productividad
1	9101-2010	4	2010	14	1	3	1	6	0	3	2	Alta Productividad
2	9127-2010	2	2010	8	1	0	0	2	3	2	1	Baja Productividad
3	9201-2010	1	2010	2	0	0	0	0	0	2	1	Baja Productividad
4	9202-2010	1	2010	4	0	1	0	2	0	1	1	Baja Productividad
5	9203-2010	5	2010	19	6	2	2	4	2	3	2	Alta Productividad
6	9204-2010	2	2010	7	0	1	2	0	1	3	1	Baja Productividad
7	9205-2010	1	2010	2	0	0	1	0	0	1	1	Baja Productividad
8	9206-2010	4	2010	14	2	4	2	1	4	1	3	Mediana Productividad
9	9301-2010	2	2010	8	0	2	2	1	3	0	1	Baja Productividad
10	9401-2010	1	2010	4	1	0	2	0	0	1	1	Baja Productividad
11	9402-2010	2	2010	7	2	1	1	1	0	2	1	Baja Productividad
12	9501-2010	3	2010	10	1	0	1	5	0	3	1	Baja Productividad
13	9502-2010	1	2010	3	0	1	0	0	2	0	1	Baja Productividad
14	9503-2010	3	2010	10	1	2	5	0	1	1	3	Mediana Productividad
15	9504-2010	4	2010	14	2	4	2	2	2	2	3	Mediana Productividad
16	9549-2010	3	2010	11	2	1	1	2	0	5	1	Baja Productividad
17	9101-2011	2	2011	6	0	0	1	3	1	1	1	Baja Productividad
18	9127-2011	2	2011	11	2	0	3	0	0	6	1	Baja Productividad
19	9201-2011	1	2011	6	1	0	2	2	0	1	1	Baja Productividad

Showing 1 to 20 of 206 entries, 12 total columns

Implementación del algoritmo Random Forest en R statistics

Exploración de los datos

```
#=================================================================================
# Primer paso: Exploración de los datos
#=================================================================================
summary(data_productividad)
str(data_productividad)
#=================================================================================
```

```
> summary(data_productividad)
  Centro_Año        Cantidad_Proyectos Fecha_inicio_Proyecto Cantidad_Productos      P_8020          P_8025          P_8019
 Length:208         Min.   :1.000      Min.   :2010          Min.   : 2.00      Min.   :0.000   Min.   :0.000   Min.   :0.00
 Class :character   1st Qu.:1.000      1st Qu.:2013          1st Qu.: 7.00      1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.00
 Mode  :character   Median :2.000      Median :2016          Median :10.00      Median :2.000   Median :2.000   Median :2.00
                    Mean   :2.404      Mean   :2016          Mean   :10.75      Mean   :1.889   Mean   :1.793   Mean   :1.76
                    3rd Qu.:3.000      3rd Qu.:2019          3rd Qu.:14.00      3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.00
                    Max.   :6.000      Max.   :2022          Max.   :26.00      Max.   :8.000   Max.   :7.000   Max.   :7.00
     P_8039          P_8016          P_8031          Cluster      Nivel_Productividad
 Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :1.000   Length:208
 1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   Class :character
 Median :1.000   Median :1.000   Median :2.000   Median :1.000   Mode  :character
 Mean   :1.745   Mean   :1.654   Mean   :1.904   Mean   :1.736
 3rd Qu.:2.250   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:3.000
 Max.   :7.000   Max.   :8.000   Max.   :6.000   Max.   :3.000
> str(data_productividad)
tibble [208 × 12] (S3: tbl_df/tbl/data.frame)
 $ Centro_Año           : chr [1:208] "9101-2010" "9127-2010" "9201-2010" "9202-2010" ...
 $ Cantidad_Proyectos   : num [1:208] 4 2 1 1 5 2 1 4 2 1 ...
 $ Fecha_inicio_Proyecto: num [1:208] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
 $ Cantidad_Productos   : num [1:208] 14 8 2 4 19 7 2 14 8 4 ...
 $ P_8020               : num [1:208] 1 1 0 0 6 0 0 2 0 1 ...
 $ P_8025               : num [1:208] 3 0 0 1 2 1 0 4 2 0 ...
 $ P_8019               : num [1:208] 1 0 0 0 2 2 1 2 2 2 ...
 $ P_8039               : num [1:208] 6 2 0 2 4 0 0 1 1 0 ...
 $ P_8016               : num [1:208] 0 3 0 0 2 1 0 4 3 0 ...
 $ P_8031               : num [1:208] 3 2 2 1 3 3 1 1 0 1 ...
 $ Cluster              : num [1:208] 2 1 1 1 2 1 1 3 1 1 ...
 $ Nivel_Productividad  : chr [1:208] "Alta Productividad" "Baja Productividad" "Baja Productividad" "Baja Productividad" ...
```

Implementación del algoritmo Random Forest en R statistics

Segundo paso: convertimos la variable de clasificación nivel de productividad en factor, con el propósito de identificarla por niveles, pero primero convertimos los datos en un data.frame

```
#=============================================================================
data_productividad <- as.data.frame(data_productividad)
str(data_productividad)
#=============================================================================
```

```
> str(data_productividad)
'data.frame':   208 obs. of  12 variables:
 $ Centro_Año           : chr  "9101-2010" "9127-2010" "9201-2010" "9202-2010" ...
 $ Cantidad_Proyectos   : num  4 2 1 1 5 2 1 4 2 1 ...
 $ Fecha_inicio_Proyecto: num  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
 $ Cantidad_Productos   : num  14 8 2 4 19 7 2 14 8 4 ...
 $ P_8020               : num  1 1 0 0 6 0 0 2 0 1 ...
 $ P_8025               : num  3 0 0 1 2 1 0 4 2 0 ...
 $ P_8019               : num  1 0 0 0 2 2 1 2 2 2 ...
 $ P_8039               : num  6 2 0 2 4 0 0 1 1 0 ...
 $ P_8016               : num  0 3 0 0 2 1 0 4 3 0 ...
 $ P_8031               : num  3 2 2 1 3 3 1 1 0 1 ...
 $ Cluster              : num  2 1 1 1 2 1 1 3 1 1 ...
 $ Nivel_Productividad  : chr  "Alta Productividad" "Baja Productividad" "Baja Productividad" "Baja Productividad" ...
```

Implementación del algoritmo Random Forest en R statistics

Se realiza escalamiento de datos

```
#=============================================================================
#Escalamiento de datos
#=============================================================================
datos_TPProductividad <- scale(data_productividad[,c(5:10)])
View(datos_TPProductividad)
#Se adiciona nuevamente la columna de etiqueta Centro-Año, Cluster y Gestion
datos_scale_TPProductividad <- as.data.frame(cbind(data_productividad$Centro_Año, datos_TPProductividad, data_productividad$Cluster,
data_productividad$Nivel_Productividad))
View(datos_scale_TPProductividad)
str(datos_scale_TPProductividad)
dim(datos_scale_TPProductividad)
datos_scale_TPProductividad$P_8020 <- as.numeric(datos_scale_TPProductividad$P_8020)
datos_scale_TPProductividad$P_8025 <- as.numeric(datos_scale_TPProductividad$P_8025)
datos_scale_TPProductividad$P_8019 <- as.numeric(datos_scale_TPProductividad$P_8020)
datos_scale_TPProductividad$P_8039 <- as.numeric(datos_scale_TPProductividad$P_8025)
datos_scale_TPProductividad$P_8016 <- as.numeric(datos_scale_TPProductividad$P_8020)
datos_scale_TPProductividad$P_8031 <- as.numeric(datos_scale_TPProductividad$P_8025)
str(datos_scale_TPProductividad)
names(datos_scale_TPProductividad) <- names(data_productividad[,c(1, 5:12)])
View(datos_scale_TPProductividad)
row.names(datos_scale_TPProductividad) <- datos_scale_TPProductividad$Centro_Año
View(datos_scale_TPProductividad)
productividad <- datos_scale_TPProductividad[, c(-1, -8)]
View(productividad)
#=============================================================================
```

Implementación del algoritmo Random Forest en R statistics

Datos escalados

Filter

	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031	Nivel_Productividad
9101-2010	-0.55877727	0.8552413	-0.55877727	0.8552413	-0.55877727	0.8552413	Alta Productividad
9127-2010	-0.55877727	-1.2709363	-0.55877727	-1.2709363	-0.55877727	-1.2709363	Baja Productividad
9201-2010	-1.18702415	-1.2709363	-1.18702415	-1.2709363	-1.18702415	-1.2709363	Baja Productividad
9202-2010	-1.18702415	-0.5622104	-1.18702415	-0.5622104	-1.18702415	-0.5622104	Baja Productividad
9203-2010	2.58245713	0.1465154	2.58245713	0.1465154	2.58245713	0.1465154	Alta Productividad
9204-2010	-1.18702415	-0.5622104	-1.18702415	-0.5622104	-1.18702415	-0.5622104	Baja Productividad
9205-2010	-1.18702415	-1.2709363	-1.18702415	-1.2709363	-1.18702415	-1.2709363	Baja Productividad
9206-2010	0.06946961	1.5639672	0.06946961	1.5639672	0.06946961	1.5639672	Mediana Productividad
9301-2010	-1.18702415	0.1465154	-1.18702415	0.1465154	-1.18702415	0.1465154	Baja Productividad
9401-2010	-0.55877727	-1.2709363	-0.55877727	-1.2709363	-0.55877727	-1.2709363	Baja Productividad
9402-2010	0.06946961	-0.5622104	0.06946961	-0.5622104	0.06946961	-0.5622104	Baja Productividad
9501-2010	-0.55877727	-1.2709363	-0.55877727	-1.2709363	-0.55877727	-1.2709363	Baja Productividad
9502-2010	-1.18702415	-0.5622104	-1.18702415	-0.5622104	-1.18702415	-0.5622104	Baja Productividad
9503-2010	-0.55877727	0.1465154	-0.55877727	0.1465154	-0.55877727	0.1465154	Mediana Productividad
9504-2010	0.06946961	1.5639672	0.06946961	1.5639672	0.06946961	1.5639672	Mediana Productividad
9549-2010	0.06946961	-0.5622104	0.06946961	-0.5622104	0.06946961	-0.5622104	Baja Productividad

Showing 1 to 16 of 208 entries, 7 total columns

Implementación del algoritmo Random Forest en R statistics

La variable factor está indexada en la 7ª posición

```
productividad$Nivel_Productividad <- as.factor(productividad$Nivel_Productividad)
View(productividad)
str(productividad)
```

```
> str(productividad)
'data.frame':   208 obs. of  7 variables:
 $ P_8020             : num  -0.559 -0.559 -1.187 -1.187 2.582 ...
 $ P_8025             : num  0.855 -1.271 -1.271 -0.562 0.147 ...
 $ P_8019             : num  -0.559 -0.559 -1.187 -1.187 2.582 ...
 $ P_8039             : num  0.855 -1.271 -1.271 -0.562 0.147 ...
 $ P_8016             : num  -0.559 -0.559 -1.187 -1.187 2.582 ...
 $ P_8031             : num  0.855 -1.271 -1.271 -0.562 0.147 ...
 $ Nivel_Productividad: Factor w/ 3 levels "Alta Productividad",..: 1 2 2 2 1 2 2 3 2 2 ...
```

Implementación del algoritmo Random Forest en R statistics

Obtenemos la distribución de clases, ya que este es un problema multiclase, podemos usar el generador de tablas de datos de clasificación múltiple, para evidenciar la distribución de datos para cada clase

```
#=============================================================================
table(productividad$Nivel_Productividad)
#=============================================================================
```

```
> table(productividad$Nivel_Productividad)

   Alta Productividad    Baja Productividad Mediana Productividad
                   39                   112                    57
```

Porcentaje de centros por distribución para cada clase

```
> round(prop.table(table(productividad$Nivel_Productividad)) * 100,digits = 1)

   Alta Productividad    Baja Productividad Mediana Productividad
                 18.8                  53.8                  27.4
```

Implementación del algoritmo Random Forest en R statistics

Analítica de datos

Primer paso: Dividimos los datos en datos de prueba y datos de entrenamiento
Definimos un índice para dividir los dos conjuntos de datos, se creó un conjunto de entrenamiento del 70% para entrenar el modelo y un 30% restante como datos de validación para probar el modelo.

```
#================================================================================
ind1 <- sample(2, nrow(productividad), replace=TRUE, prob=c(0.7, 0.3))
# Valores asignados para las filas (1: Entrenamiento, 2: Prueba)
trainData1 <- productividad[ind==1,]
testData1 <-  productividad[ind==2,]
dim(trainData1)
dim(testData1)
#================================================================================
```

```
> dim(trainData1)
[1] 146   7
> dim(testData1)
[1] 62  7
```

Implementación del algoritmo Random Forest en R statistics

Se carga la librería (randomForest)

```
#=============================================================================
library(randomForest)
#=============================================================================

#=============================================================================
productividad_classifier <- randomForest(Nivel_Productividad ~.,
                                         data = trainData1, # Conjunto de datos de entrenamiento
                                         importance = T)
productividad_classifier                 # Matrix de confusion: Evaluacion de la prediccion
#=============================================================================
```

```
> productividad_classifier                # Matrix de confusion: Evaluacion de la prediccion

Call:
 randomForest(formula = Nivel_Productividad ~ ., data = trainData1,      importance = T)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 30.14%
Confusion matrix:
                     Alta Productividad Baja Productividad Mediana Productividad class.error
Alta Productividad                    2                 17                     7  0.92307692
Baja Productividad                    3                 75                     3  0.07407407
Mediana Productividad                 5                  9                    25  0.35897436
```

Implementación del algoritmo Random Forest en R statistics

Se carga la libreria (randomForest)

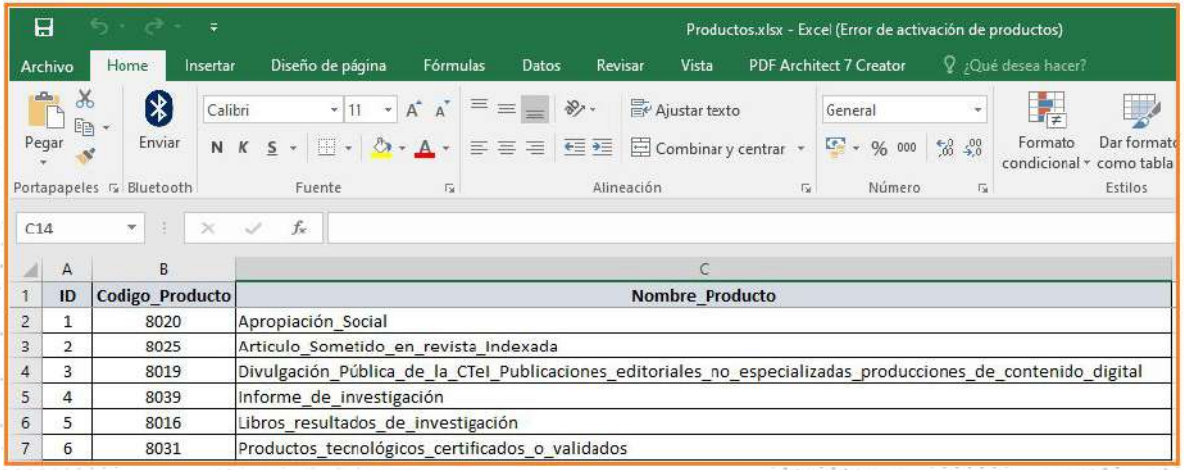
```
#=============================================================================
library(randomForest)
#=============================================================================

#=============================================================================
proceso_classifier <- randomForest(Gestion ~.,
                                   data = trainData, # Conjunto de datos de entrenamiento
                                   importance = T)
proceso_classifier                 # Matrix de confusion: Evaluacion de la prediccion
#=============================================================================
```

```
#=============================================================================
# Las características más importantes del nivel de productividad para los centros
importance_prod = importance(productividad_classifier)
importance_prod
#=============================================================================
varImportance_prod = data.frame(Variables = row.names(importance_prod),
                                Importance = round(importance_prod[, "MeanDecreaseAccuracy"],2))

rankImportance_prod = varImportance_prod%>%mutate(Rank=paste('#',dense_rank(desc(Importance))))

ggplot(rankImportance_prod,aes(x=reorder(Variables,Importance),
                               y=Importance,fill=Importance)) +
  geom_bar(stat='identity') +
  geom_text(aes(x = Variables, y = 0.5, label = Rank),
            hjust=0, vjust=0.55, size = 4, colour = 'white') +
  labs(x = 'Variables') +
  coord_flip() +
  theme_classic()
#=============================================================================
```

Implementación del algoritmo Random Forest en R statistics

Nivel de importancia de los predictores

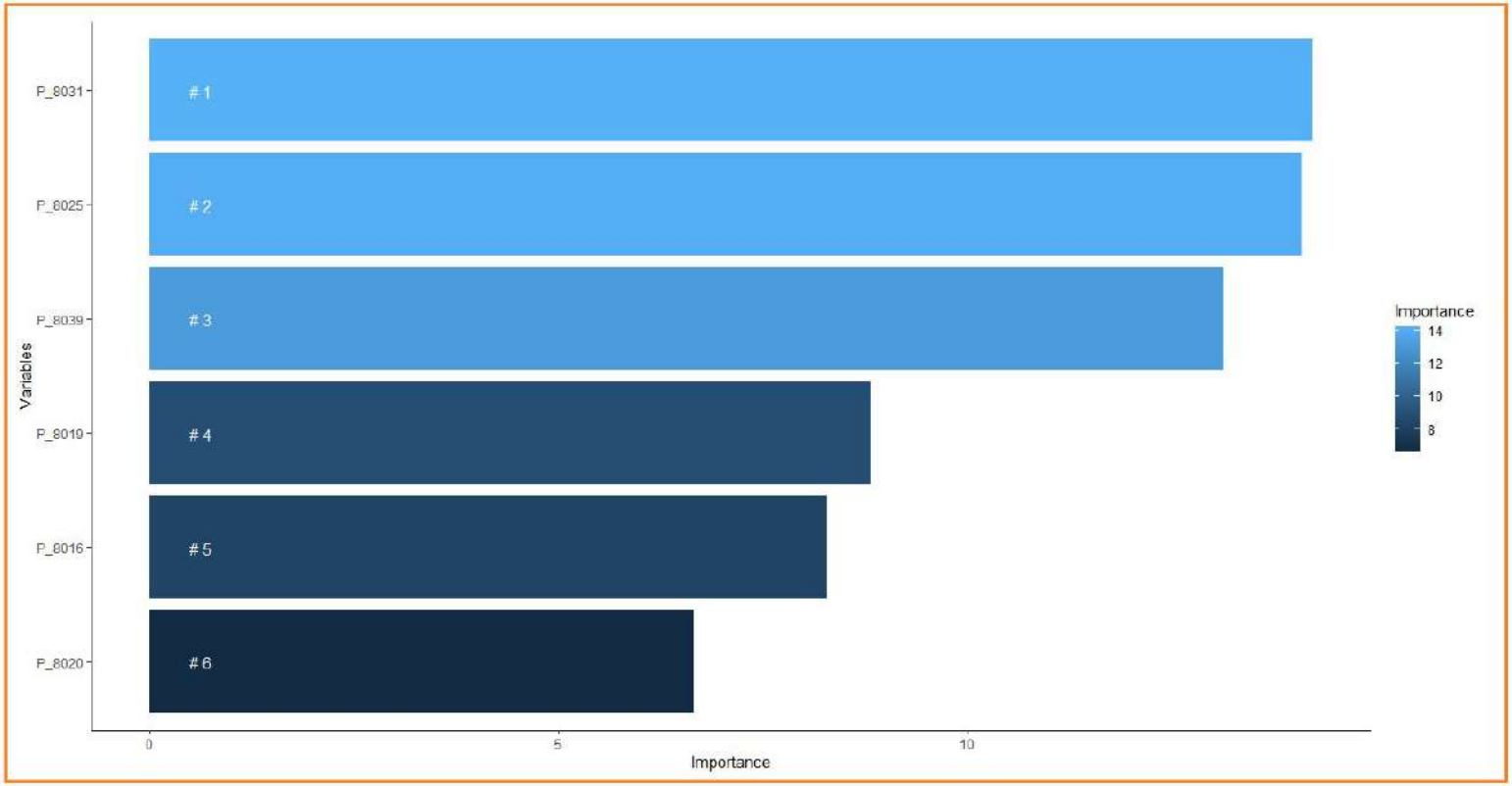
```
> importance_prod
        Alta Productividad Baja Productividad Mediana Productividad MeanDecreaseAccuracy
P_8020          -1.741486           5.252055             5.686394             6.658020
P_8025          -5.367219           8.423517            16.097664            14.109163
P_8019           2.727407           4.832161             8.075860             8.834543
P_8039          -5.502712           7.694560            14.976803            13.153378
P_8016           1.536152           5.386000             7.126662             8.297906
P_8031          -4.610135           8.465576            15.764616            14.243584
        MeanDecreaseGini
P_8020         5.881731
P_8025         9.367019
P_8019         5.696320
P_8039         9.209668
P_8016         6.028610
P_8031         9.486834
```



	A	B	C
1	ID	Codigo_Producto	Nombre_Producto
2	1	8020	Apropiación_Social
3	2	8025	Articulo_Sometido_en_revista_Indexada
4	3	8019	Divulgación_Pública_de_la_CTeI_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital
5	4	8039	Informe_de_investigación
6	5	8016	Libros_resultados_de_investigación
7	6	8031	Productos_tecnológicos_certificados_o_validados

Implementación del algoritmo Random Forest en R statistics

Se entrega el gráfico donde se evidencia las variables con mayor importancia en el modelo



Implementación del algoritmo Random Forest en R statistics

Matriz de confusión para validar el modelo

Tercer paso: El modelo es relativamente preciso. Sin embargo, ahora sabemos qué tan bien se desempeña el modelo en el conjunto de entrenamiento, necesitamos validar esto con una matriz de confusión. Random Forest muestra que ha sido entrenado en más de > 2 clases, por lo que pasa de un modelo binario a un modelo de clasificación múltiple. Las funciones contenidas en el paquete funcionan con métodos binarios y de clasificación múltiple.

Uso de la matriz de confusión del paquete CARET: La matriz de confusión almacena una serie de elementos de la lista que deben utilizarse juntos para comparar el rendimiento del ajuste del modelo a lo largo del tiempo para asegurarse de que no haya un desfase de características subyacentes ni una regresión en el rendimiento. Para evitar lo anterior, hacemos una predicción sobre el modelo ajustado con los datos de prueba

```
#==========================================================================
rf_class <- predict(productividad_classifier, newdata = testData1)
predictions_productividad <- cbind(data.frame(train_preds=rf_class,
                                   testData1$Nivel_Productividad))
#==========================================================================
```

Implementación del algoritmo Random Forest en R statistics

Matriz de confusión para validar el modelo

Cuarto paso: Crear un objeto de matriz de confusión

```
#==========================================================================
cm_productividad <- caret::confusionMatrix(predictions_productividad$train_preds,
                                           predictions_productividad$testData1.Nivel_Productividad)
print(cm_productividad)
#==========================================================================
```

```
> print(cm_productividad)
Confusion Matrix and Statistics

                          Reference
Prediction                 Alta Productividad Baja Productividad Mediana Productividad
  Alta Productividad                        5                  1                     4
  Baja Productividad                        6                 29                     7
  Mediana Productividad                     2                  1                     7

Overall Statistics

               Accuracy : 0.6613
                 95% CI : (0.5299, 0.7767)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : 0.007567

                  Kappa : 0.4167

 Mcnemar's Test P-Value : 0.032984

Statistics by Class:

                     Class: Alta Productividad Class: Baja Productividad Class: Mediana Productividad
Sensitivity                            0.38462                   0.9355                       0.3889
Specificity                            0.89796                   0.5806                       0.9318
Pos Pred Value                         0.50000                   0.6905                       0.7000
Neg Pred Value                         0.84615                   0.9000                       0.7885
Prevalence                             0.20968                   0.5000                       0.2903
Detection Rate                         0.08065                   0.4677                       0.1129
Detection Prevalence                   0.16129                   0.6774                       0.1613
Balanced Accuracy                      0.64129                   0.7581                       0.6604
```

Implementación del algoritmo Random Forest en R statistics

Se almacena la matriz de confusión en una tabla

```
#=================================================================================
# se almacena la matriz de confusion en una tabla
#=================================================================================

matrix_productividad<-as.data.frame(matrix(data = cm_productividad$table, nrow=3,ncol=3))
names(matrix_productividad) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")
row.names(matrix_productividad) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")
matrix_productividad$Total <- 0
matrix_productividad <- rbind(matrix_productividad,c(0,0,0,0))
row.names(matrix_productividad) <-c("Alta Gestion", "Baja Gestion", "Media Gestion", "Total")
matrix_productividad$Total <- c(sum(matrix_productividad[1,]), sum(matrix_productividad[2,]), sum
(matrix_productividad[3,]), 0)
matrix_productividad[4,] <- c(sum(matrix_productividad$`Alta Gestion`), sum
(matrix_productividad$`Baja Gestion`), sum(matrix_productividad$`Media Gestion`), sum
(matrix_productividad$Total))
#=================================================================================
```

Tabla de métricas para las tres clases

```
#=================================================================================
# Tabla de metricas para las tres clases
#=================================================================================

metricas_clase_productividad <- as.data.frame(matrix(data=t(cm_productividad$byClass), nrow=11, ncol
=3))
row.names(metricas_clase_productividad) <- c("Sensitivity", "Specificity", "Pos Pred Value",
                                             "Neg Pred Value","Precision", "Recall", "F1",
                                             "Prevalence", "Detection Rate", "Detection Prevalence",
                                             "Balanced Accuracy" )
names(metricas_clase_productividad) <- c("Alta Gestion", "Baja Gestion", "Media Gestion")

#=================================================================================
```

Implementación del algoritmo Random Forest en R statistics

Cálculo de las métricas del clasificador

```
Accuracy <- cm_productividad$overall[1]
clasificador_metricas_productividad <- data.frame("Metricas_del_Clasificador"=c(0,0,0,0,0,0))
k <- c(1,2,5,6,7)
for (i in 1:5)
{
  for (j in 1:3)
  {
    clasificador_metricas_productividad[i,1] <- clasificador_metricas_productividad[i,1] +
      (matrix_productividad[j,4]/matrix_productividad[4,4])*metricas_clase_productividad[k[i],j]
  }

}
clasificador_metricas_productividad[6,1] <- Accuracy
row.names(clasificador_metricas_productividad) <- c("Sensitivity", "Specificity",
                                                    "Precision", "Recall", "F1", "Accuracy" )
clasificador_metricas_productividad <- t(clasificador_metricas_productividad)
clasificador_metricas_productividad
```

```
> clasificador_metricas_productividad
                      Sensitivity Specificity Precision    Recall        F1  Accuracy
Metricas_del_Clasificador   0.7584736   0.6884657 0.6612903 0.7584736 0.688995 0.6612903
```

GC-F-004v.05

Divulgación

Actividades

- Proyecto Promoción de Ecosistemas de Innovación en América Latina - Componentes C3 OTC (Oficina de Transferencia de Conocimiento) y C4 (Gestión de Capital Documental).

- Participación en sesión formativa REDIT Proyectos I+D

- Cursos de transferencia de conocimientos dictados por la Universidad Politécnica de Valencia.
  - La función de transferencia de conocimiento en ecosistemas de innovación
  - Gestión del capital intelectual en SENNOVA.

- Elaboración de actas de los prototipos.

Pendiente

- Aportes para el artículo de visualización de la información

Actividades

2. Se revisó el artículo y se realizaron aportes al mismo

Visualización geoespacial mediante analítica de datos usando Python para la medición de indicadores de Ciencia, Tecnología e Innovación en el SENA Regional Antioquia

Geospatial visualization through data analytics using Python for the measurement of indicators of Science, Technology and Innovation in the SENA Regional Antioquia

Clustering and geospatial visualization of KPIs using Python in the SENA regional Antioquia

Resumen

Las actividades de Ciencia, Tecnología e Innovación - ACTI se relacionan con la realización de acciones que involucran la investigación, el desarrollo experimental, apoyo a la formación y capacitación científica y tecnológica, actividades de innovación, servicios científicos y tecnológicos, administración y otras actividades de apoyo en Ciencia Tecnología e Innovación - CTeI, las cuales son realizadas por entidades públicas y privadas para aportar a la generación y apropiación del conocimiento. En ese sentido, debido al aumento exponencial que ha tenido la generación de productos de nuevo conocimiento a nivel educativo y en pro de dar respuesta de esta forma a una necesidad evidente de aportar nuevas tecnologías y conocimientos al sector productivo de la región, el SENA a través de su Sistema de Investigación SENNOVA potencializa tanto los recursos humanos como tecnológicos con los que cuenta la entidad para aportar al crecimiento económico y social de la región. Teniendo en cuenta lo anterior, en la regional Antioquia del SENA se está trabajando en el desarrollo de la visualización de un modelo de analítica de datos para la medición de indicadores KPIs, donde se identifique de manera gráfica las diferentes categorías de ACTI que se vienen trabajando en los últimos años en la entidad, dentro de las que se encuentran los procesos de gestión y productividad de proyectos por centro en un periodo de tiempo, con el propósito de facilitar la medición y análisis de los resultados obtenidos desde las perspectiva tanto de la estadística descriptiva como de la ciencia de datos, con técnicas tales como clusterización por K-means y Análisis de Componentes Principales (ACP), que permitieron agrupar los indicadores KPIs encontrados. Se utilizó Folium de Python como herramienta de visualización que permite generar comparaciones de manera gráfica entre los diferentes centros SENA de la Regional Antioquia en un periodo de tiempo determinado. Se observó que mediante estas técnicas se presenta mayor facilidad de interpretación, comparación y análisis de resultados estadísticos y de analítica de datos para medidores de CTeI frente a herramientas de visualización de la información usadas convencionalmente.

Palabras clave: Visualización de datos, Ciencia de datos, Folium, CTeI,Técnicas de visualización, Herramientas de visualización, Geoespacial, Python, indicadores.

Abstract

Keywords: Data visualization, Data science, Folium, Visualization Techniques, Visualization Tools, Geospatial,

Keywords: Clustering, Folium, CTeI, KPIs, K-means, PCA, ACTI, SENA

usuario
Excede 15 palabras, referencia:
https://scioti.ufd.cl/article.php?secid=revsci_articulos&id=82664.2125159990013900118
Actualmente tiene 23, se sugiere reducir 8 palabras

usuario
Se propone este título que cumple con cantidad de palabras, es más breve y llamativo. Además, porque en Ciencia, Tecnología e innovación los indicadores se miden como KPIs

usuario
Por qué usar la expresión KPIs, significado. Referencio artículo Q1:
https://www.sciencedirect.com/science/article/pii/S2214716021001974
Q1 según Scimago Journal:
https://www.scimagojr.com/journalsearch.php?q=21100974811&tip=sid&clean=0

usuario
Sugiero revisión de este párrafo al Ing John Jairo y Julián que conocen el decreto 1666.
Referencias:
https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=124851
Articulando de una mejor forma este párrafo a las actividades de Ciencia, Tecnología e innovación en el SENA como institución pública y al sector privado.

usuario
El resumen no pasa de 350 palabras, según la universidad de Harvard no debe ser de más de 350. Referencia:
https://www.hms.harvard.edu/DMP/documents/SubmissionTitlePageAbstract.pdf
Resumen corregido a 339 palabras y evitando redundancias semánticas.

usuario
Deben ser mínimo 4 a 6, referencia: Universidad de Yale: https://www.aje.com/arc/editing-tips-choosing-effective-keywords/
Para que lo tengan en cuenta por favor en la edición del artículo.
Esta es mi sugerencia según la temática que se va a tratar

usuario
Se recomienda esto

Introducción

Imagen de visualización sugerida:

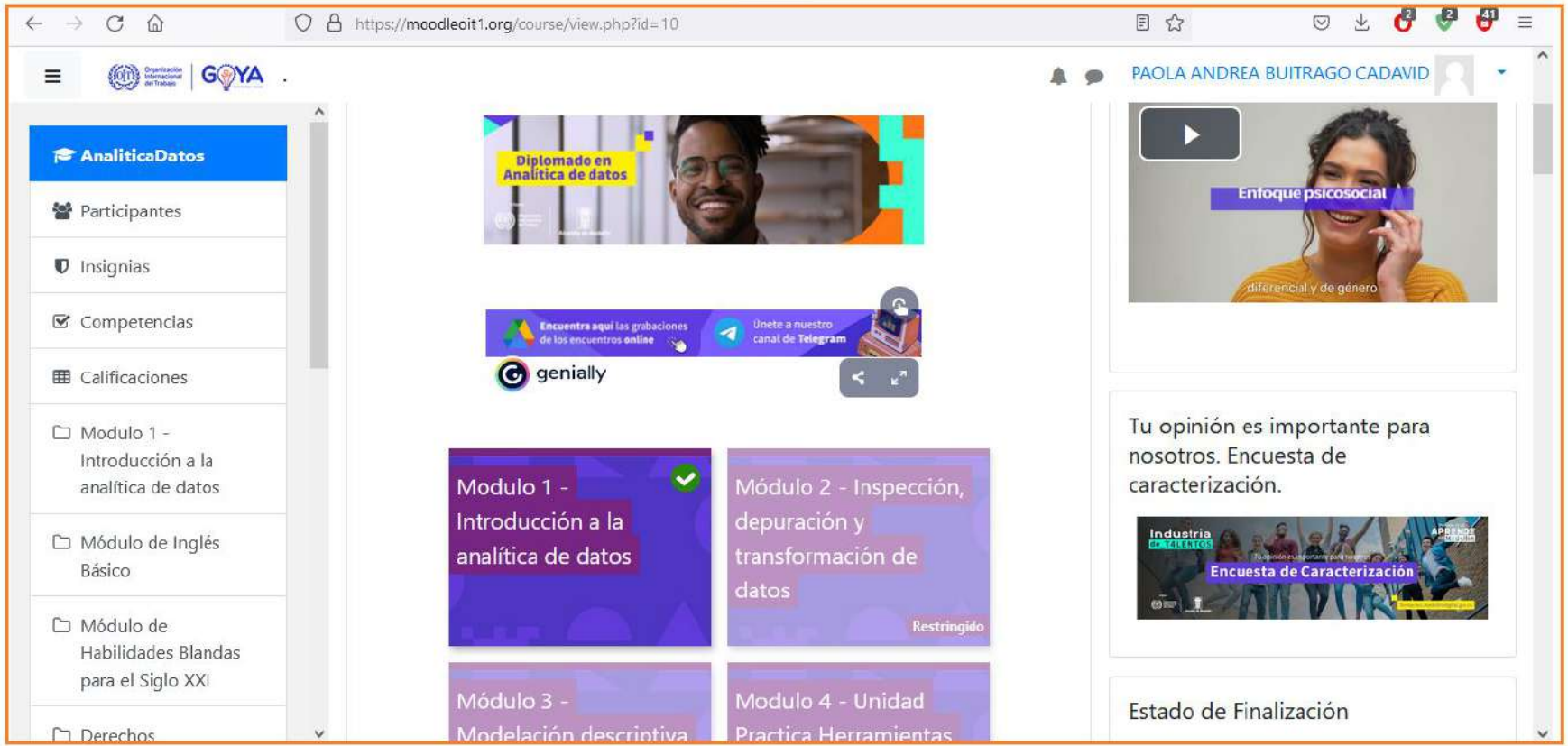
En la plataforma actual utilizada, con el ingeniero John Feria se puede incorporar el mapa, la estadística descriptiva y el respectivo indicador.

Referencia: https://s3-us-west-1.amazonaws.com/powr/defaults-images-optimized/dashboard-template.jpg?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIDGAJDZWIIYJQWIIIQ%2F20220826%2Fus-west-1%2Fs3%2Faws4_request&X-Amz-Date=20220826T230256Z&X-Amz-Expires=86400&X-Amz-Signature=7338fa7f7ccdc8f4d51aea8f2a1f8f2c1d0d2f6d8f2f0008d3e7a1d2e1

Bibliografía

Actividades

1. Diplomado en Analitica de Datos Industria OIT



Actividades plan de transferencia del proyecto 9748

Acción	Programas de formación	Semillero	Fecha
Divulgación de tecnologías para ofrecer soluciones acorde a las necesidades del sector productivo	Aprendices de programas de formación relacionados con áreas administrativas	SIPI	24 de agosto
Divulgación de buenas prácticas tecnológicas en la aplicación de analítica de datos	Aprendices de programas de formación relacionados con áreas de desarrollo de software	MERLIN	21 de septiembre

GRACIAS

Línea de atención al ciudadano: 018000 910270
Línea de atención al empresario: 018000 910682

@SENAcomunica

www.sena.edu.co

ANEXOS 5. Actas de reuniones.

ACTA No. 115		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Seguimiento Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 12 de julio de 2022	**HORA INICIO:** 10:30 a.m.	**HORA FIN:** 12:15 p.m.
LUGAR Y/O ENLACE: Laboratorio de Inteligencia Artificial Piso 5 https://cutt.ly/nLElq8J	**DIRECCIÓN GENERAL / REGIONAL / CENTRO** Centro de Servicios y Gestión Empresarial	

AGENDA O PUNTOS PARA DESARROLLAR:

1. Avance de la integración del prototipo con el modelo de analítica de datos.
2. Avances de la redacción del artículo de visualización de información. actividades de divulgación

OBJETIVO(S) DE LA REUNIÓN:

Realizar seguimiento a la ejecución de las actividades del Proyecto SGPS 9748 Prototipo Innovación.

DESARROLLO DE LA REUNIÓN

1. Avance de la integración del prototipo con el modelo de analítica de datos.

Para iniciar el desarrollador John Ferla socializó que fueron realizados los ajustes a la información de productividad y procesos, los cuales fueron revisados y ajustados con los investigadores Paola Buitrago y Jesús Arbey Benavides. También comentó que guardó el código en un espacio propio y lo compartirá al resto del equipo, además mencionó que el hosting está en proceso de adquisición, sin embargo, en caso de que se requiera se instalará en un equipo de manera local. Se sugirió que en vez de nombrar los clustering como mapas agregarle el nombre de conglomerados.

Se acordó realizar ajustes en la distribución de la información en el aplicativo y le van a compartir el estilo de gráficas que se requieren para el aplicativo y que sean fáciles de comprender para el público objetivo, el objetivo es que el desarrollador tenga el espacio para trabajar esta semana y terminarlo la próxima semana.

2. Avances de la redacción del artículo de visualización de información. actividades de divulgación

Se presentó el artículo de visualización de información y revisó el avance del artículo, se sugirieron ajustes en la redacción para que se especifique que los datos son sintéticos y que se cuide no sobrepasar el máximo de hojas que se aceptan en el evento. Igualmente, se sugiere no especificar demasiado en la explicación de componentes principales y clusterización.

El aporte de la investigadora Erika Restrepo al artículo es verificar que se cumpla con las especificaciones de forma solicitadas en el evento para el artículo, se tienen las siguientes fechas para envíos, el 15 de julio enviar resumen del capítulo de libro para expotecnológica, 31 de julio para la revista retos y se debe solicitar el pago de la participación en el evento de Foristom hasta el 17 de julio.

La investigadora Erika Restrepo realizó las siguientes actividades:

- Participación en el Proyecto Promoción de Ecosistemas de Innovación en América Latina - Componentes C3 OTC (Oficina de Transferencia de Conocimiento) y C4 (Gestión de Capital Documental).
- Realización de los cursos de transferencia de conocimientos dictados por la Elaboración de actas de los prototipos.

Tiene pendiente realizar las siguientes actividades:

- Aportes para el artículo de visualización de la información
- Compartir el resumen para expotecnológica

CONCLUSIONES

Se continúa avanzando con la integración del aplicativo de recolección de información con el modelo de analítica que se eligió.
La escritura del artículo está en su recta final, falta agregar algunas referencias, revisar la redacción y la forma general del artículo.

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales se debe garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento. El Servicio Nacional de Aprendizaje SENA solicita la siguiente clasificación de la información:

La información de este documento se debe clasificar como:

PÚBLICA ☐ **PRIVADO** ☐ **SEMIPRIVADO** ☐ **SENSIBLE** ☐

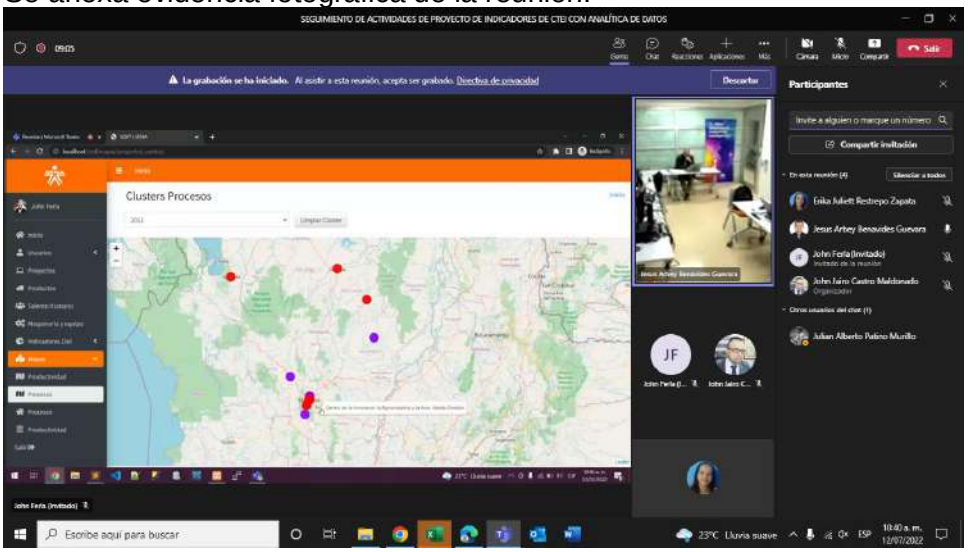
Nota: Antes de contestar esta informacional por favor remitirse al instructivo

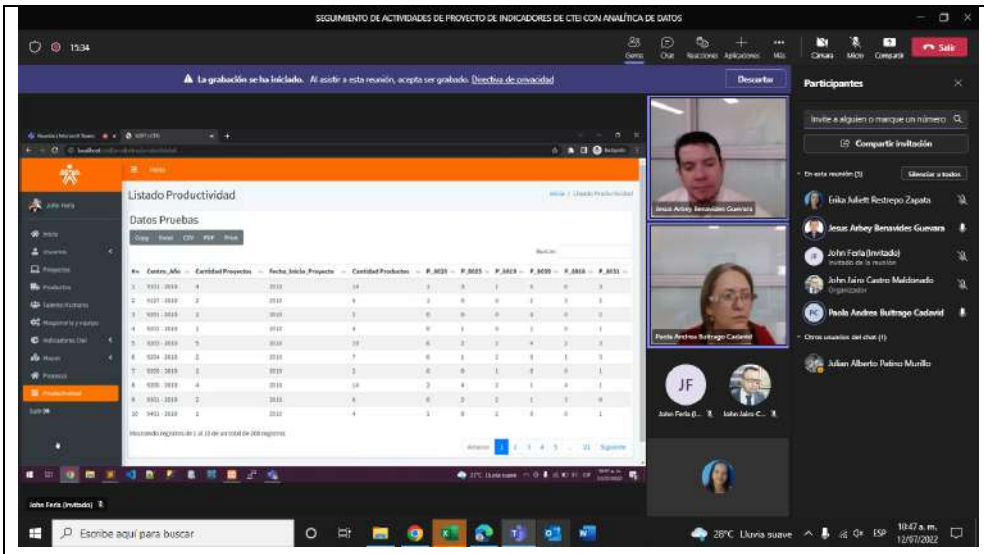
COMPROMISOS

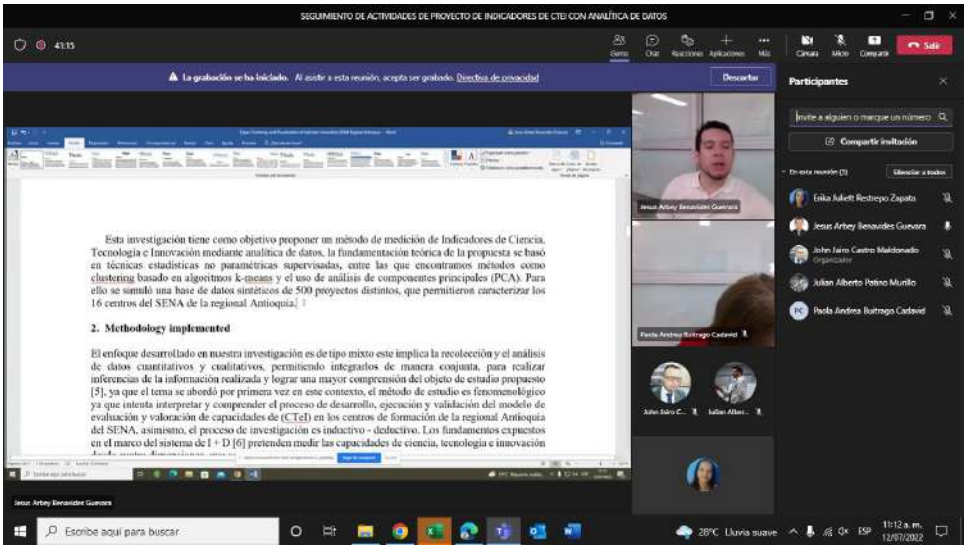
ACTIVIDAD	RESPONSABLE	FECHA
Continuar apoyando el desarrollo y la elaboración de la documentación del prototipo.	Jesús Arbey Benavides Paola Andrea Buitrago	19 de agosto de 2022
Continuar integrando el modelo de analítica en el aplicativo	John Ferla	19 de agosto de 2022
Continuar con las actividades de divulgación del prototipo.	Erika Restrepo	19 de agosto de 2022

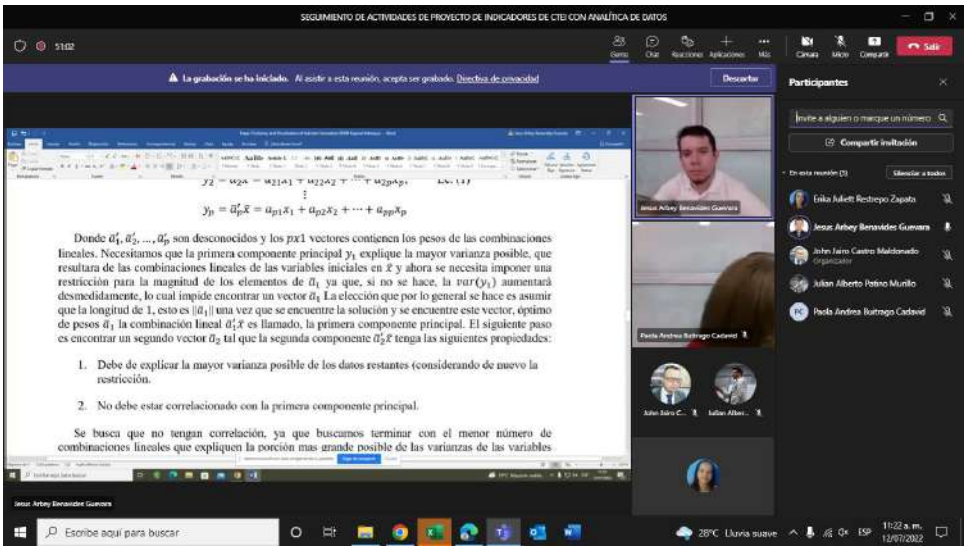
ASISTENTES: (Incorporar registro de asistencia)

Se anexa evidencia fotográfica de la reunión:




SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANALÍTICA DE DATOS
15:34
La grabación se ha iniciado. Al asistir a esta reunión, acepta ser grabado. Directiva de privacidad
Descartar
Participantes
Invite a alguien o marque un número
Compartir invitación
Listado Productividad
Datos Pruebas
Escribe aquí para buscar
28°C Lluvia suave
10:47 a. m.
12/07/2022


SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANALÍTICA DE DATOS
41:15
La grabación se ha iniciado. Al asistir a esta reunión, acepta ser grabado. Directiva de privacidad
Descartar
Participantes
Esta investigación tiene como objetivo proponer un método de medición de Indicadores de Ciencia, Tecnología e Innovación mediante analítica de datos, la fundamentación teórica de la propuesta se basó en técnicas estadísticas no paramétricas supervisadas, entre las que encontramos métodos como clustering basado en algoritmos k-means y el uso de análisis de componentes principales (PCA). Para ello se simuló una base de datos sintéticos de 500 proyectos distintos, que permitieron caracterizar los 16 centros del SENA de la regional Antioquia.
2. Methodology implemented
El enfoque desarrollado en nuestra investigación es de tipo mixto este implica la recolección y el análisis de datos cuantitativos y cualitativos, permitiendo integrarlos de manera conjunta, para realizar inferencias de la información realizada y lograr una mayor comprensión del objeto de estudio propuesto [5], ya que el tema se abordó por primera vez en este contexto, el método de estudio es fenomenológico ya que intenta interpretar y comprender el proceso de desarrollo, ejecución y validación del modelo de evaluación y valoración de capacidades de (CTeI) en los centros de formación de la regional Antioquia del SENA, asimismo, el proceso de investigación es inductivo - deductivo. Los fundamentos expuestos en el marco del sistema de I + D [6] pretenden medir las capacidades de ciencia, tecnología e innovación
Escribe aquí para buscar
28°C Lluvia suave
11:12 a. m.
12/07/2022


SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANALÍTICA DE DATOS
51:02
La grabación se ha iniciado. Al asistir a esta reunión, acepta ser grabado. Directiva de privacidad
Descartar
Participantes
$y_p = \vec{a}_p'\vec{x} = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p$
Donde $\vec{a}_1', \vec{a}_2', \ldots, \vec{a}_p'$ son desconocidos y los $px1$ vectores contienen los pesos de las combinaciones lineales. Necesitamos que la primera componente principal y_1 explique la mayor varianza posible, que resultara de las combinaciones lineales de las variables iniciales en $\vec{x}$ y ahora se necesita imponer una restricción para la magnitud de los elementos de $\vec{a}_1$ ya que, si no se hace, la $var(y_1)$ aumentará desmedidamente, lo cual impide encontrar un vector $\vec{a}_1$ La elección que por lo general se hace es asumir que la longitud de 1, esto es $||\vec{a}_1||$ una vez que se encuentre la solución y se encuentre este vector, óptimo de pesos $\vec{a}_1$ la combinación lineal $\vec{a}_1'\vec{x}$ es llamado, la primera componente principal. El siguiente paso es encontrar un segundo vector $\vec{a}_2$ tal que la segunda componente $\vec{a}_2'\vec{x}$ tenga las siguientes propiedades:
1. Debe de explicar la mayor varianza posible de los datos restantes (considerando de nuevo la restricción.
2. No debe estar correlacionado con la primera componente principal.
Se busca que no tengan correlación, ya que buscamos terminar con el menor número de combinaciones lineales que expliquen la porción mas grande posible de las varianzas de las variables
Escribe aquí para buscar
28°C Lluvia suave
11:22 a. m.
12/07/2022

Nota: Julián Patiño, líder del proyecto, se conectó de manera remota a la reunión.

Enlaces de grabación:

SENA

REGISTRO DE ASISTENCIA Y APROBACIÓN DEL ACTA No. 123 DEL DÍA 12 DEL MES DE Julio DEL AÑO 2022.

OBJETIVO (S): Realizar seguimiento a la ejecución de las actividades del proyecto SGPS9748 -Prototipo innovación

No.	NOMBRES Y APELLIDOS	No. DOCUMENTO	PLANTA	CONTRATISTA	OTRO ¿CUAL?	DEPENDENCIA/ EMPRESA	CORREO ELECTRÓNICO	TELÉFONO/EXT.SENA	AUTORIZA GRABACIÓN	FIRMA O PARTICIPACIÓN VIRTUAL
1	Jesus Arbey Benavides	1085260713		X		SENA	Jabenavides@sena.edu.co	3105595765	Si	Jesus Arbey
2	John Ferla	1056621874		X		SENA	Joferla@sena.edu.co	3146864277	SI	
3	John Jairo Castro M.	88.265.552	X			SENA CESGE	jcastrom@sena.edu.co	42850	si	
4	Erika Juliett Restrepo	43973982		X		SENA SENNOVA	erestrepoz@sena.edu.co	3005077253	Si	

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales, el Servicio Nacional de Aprendizaje SENA, se compromete a garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento, y les dará el tratamiento correspondiente en cumplimiento de lo establecido legalmente.

ACTA No. 115		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Seguimiento Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 05 de julio de 2022	**HORA INICIO:** 10:30 a.m.	**HORA FIN:** 10:59 p.m.
LUGAR Y/O ENLACE: Laboratorio de Inteligencia Artificial Piso 5 https://bit.ly/3yBZYEV	**DIRECCIÓN GENERAL / REGIONAL / CENTRO** Centro de Servicios y Gestión Empresarial	

AGENDA O PUNTOS PARA DESARROLLAR:

1. Avance de la integración del prototipo con el modelo de analítica de datos.
2. Avances de la redacción del artículo de visualización de información. actividades de divulgación

OBJETIVO(S) DE LA REUNIÓN:

Realizar seguimiento a la ejecución de las actividades del Proyecto SGPS 9748 Prototipo Innovación.

DESARROLLO DE LA REUNIÓN

1. Avance de la integración del prototipo con el modelo de analítica de datos.

Para iniciar se solucionaron inquietudes que presentaba el desarrollador John Ferla, quien además presentó los avances de la visualización de la información de las bases de datos por año ya integradas en el prototipo de recolección de información. En el clúster de procesos también se presentará la información por años y está pendiente validar que lo presentado esté acorde a lo que se vienen trabajando. También creó dos tablas de procesos y productividad por centros que permite validar los datos. Los investigadores Paola Buitrago y Jesús Benavides se reunirán con John Ferla hoy para solucionar inquietudes adicionales.

John Ferla mencionó que según su cronograma de trabajo se está dentro de los tiempos para entregar el prototipo el 12 de julio y que quede tiempo suficiente para realizar el manual técnico, el manual de usuario y demás documentos.

Por otro lado, se mencionó que aún no se tiene un espacio para alojar el aplicativo y se consultará a Julián Patiño, igual se tiene la opción de instalarlo de manera local o un servidor que adquiera el centro. Por último, mencionó que en el prototipo ambiental lo están solicitando para realizar ajustes y solicitó espacio para realizarlos.

La investigadora Paola Buitrago propuso ir adelantando los documentos para el registro del aplicativo ante la Dirección Nacional de Derechos de Autor, se debe gestionar una nueva carta donde se autorice a la subdirectora Rosalía Suescún para realizar el trámite del registro.

2. Avances de la redacción del artículo de visualización de información. actividades de divulgación

Se presentó y revisó el avance del artículo de visualización de información se presentaron los apartados del artículo, están pendientes las conclusiones y algunos ajustes de forma, fue necesario cambiar un poco el enfoque para que quede más acorde a lo que se viene trabajando.

La investigadora Erika Restrepo realizó las siguientes actividades:

- Participación en el Proyecto Promoción de Ecosistemas de Innovación en América Latina - Componentes C3 OTC (Oficina de Transferencia de Conocimiento) y C4 (Gestión de Capital Documental).

- Realización de los cursos de transferencia de conocimientos dictados por la Universidad Politécnica de Valencia:
  - La función de transferencia de conocimiento en ecosistemas de innovación
  - Gestión del capital intelectual en SENNOVA.
- Elaboración de actas de los prototipos.

Tiene pendiente realizar las siguientes actividades:

- Aportes para el artículo de visualización de la información

CONCLUSIONES

Se está avanzando con la integración del aplicativo de recolección de información con el modelo de analítica que se eligió.
Se está avanzando con la escritura del artículo tener en cuenta la tarea transversal de escribir los artículos.

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales se debe garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento. El Servicio Nacional de Aprendizaje SENA solicita la siguiente clasificación de la información:

La información de este documento se debe clasificar como:

PÚBLICA ☐ **PRIVADO** ☐ **SEMIPRIVADO** ☐ **SENSIBLE** ☐

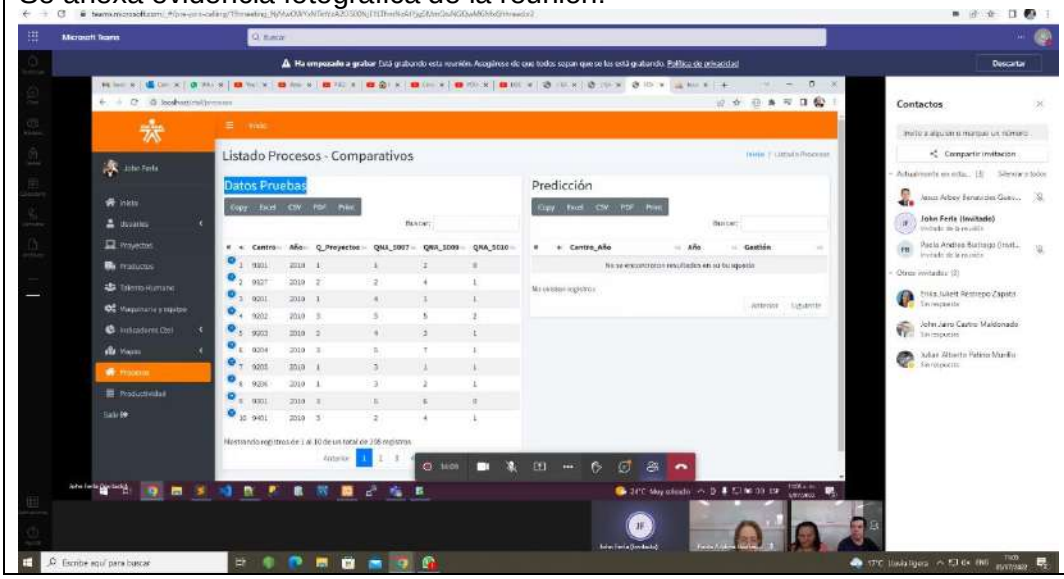
Nota: Antes de contestar esta informacional por favor remitirse al instructivo

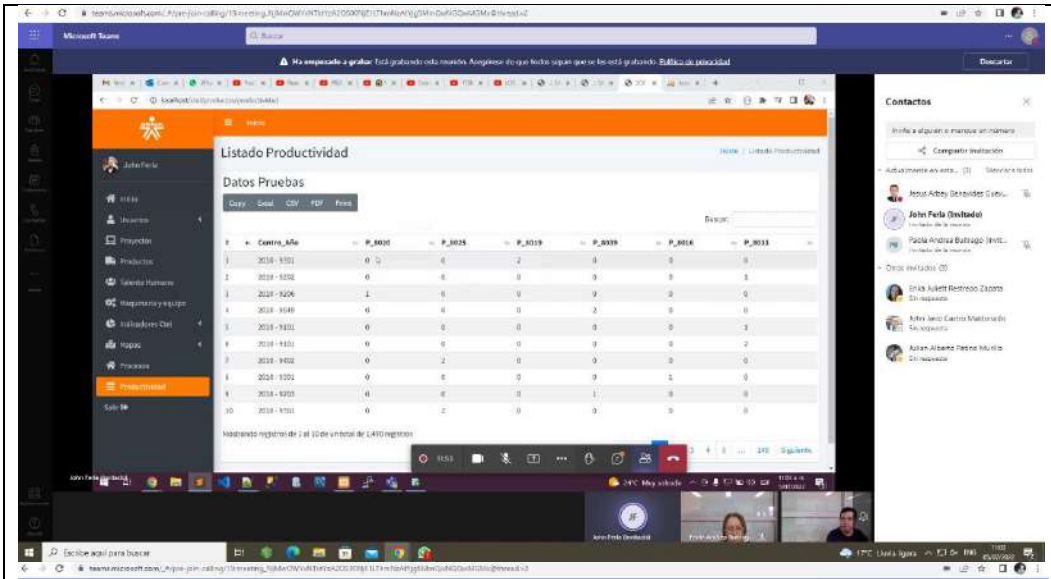
COMPROMISOS

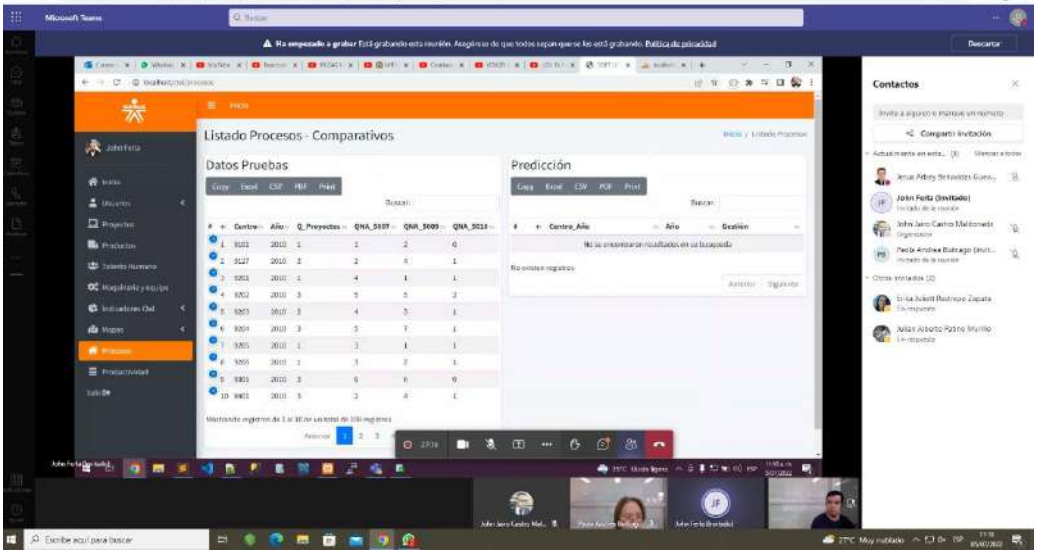
ACTIVIDAD	RESPONSABLE	FECHA
Continuar trabajando en lo que se requiera de la integración del aplicativo con el modelo de analítica de datos.	Jesús Arbey Benavides Paola Andrea Buitrago John Ferla	12 de julio de 2022
Continuar con las actividades de divulgación del prototipo.	Erika Restrepo	12 de julio de 2022

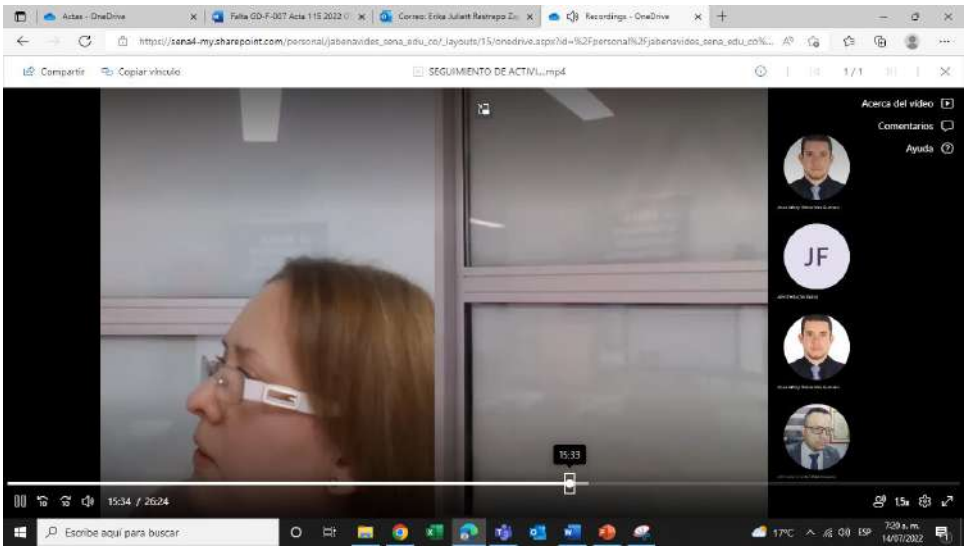
ASISTENTES: (Incorporar registro de asistencia)

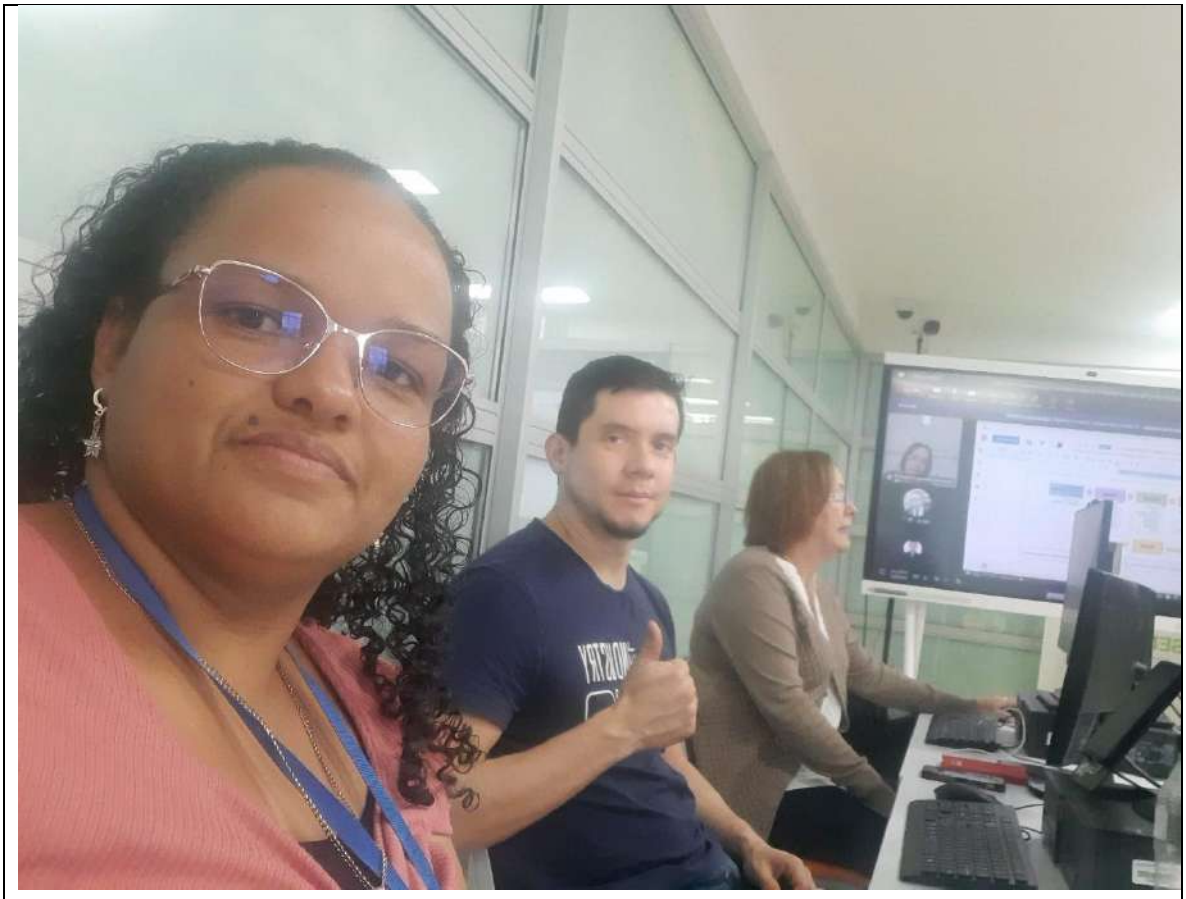
Se anexa evidencia fotográfica de la reunión:




Microsoft Teams
Ha empezado a grabar Está grabando esta reunión. Asegúrese de que todos sepan que se les está grabando. Política de privacidad
Descartar
Listado Productividad
Datos Pruebas
Contactos


Microsoft Teams
Ha empezado a grabar Está grabando esta reunión. Asegúrese de que todos sepan que se les está grabando. Política de privacidad
Descartar
Listado Procesos - Comparativos
Datos Pruebas
Predicción
Contactos


SEGUIMIENTO DE ACTIVI...mp4
Acerca del vídeo
Comentarios
Ayuda
JF
15:34 / 26:24
Escribe aquí para buscar



Nota: John Ferla, Desarrollador del proyecto, se conectó de manera remota a la reunión.

Enlaces de grabación:
https://sena4-my.sharepoint.com/:v:/r/personal/jabenavides_sena_edu_co/Documents/Recordings/SEGUIMIENTO%20DE%20ACTIVIDADES%20DE%20PROYECTO%20DE%20INDICADORES%20DE%20CTEI%20CON%20ANAL%C3%8DTICA%20DE%20DATOS-20220705_105934-Grabaci%C3%B3n%20de%20la%20reuni%C3%B3n.mp4?csf=1&web=1&e=Nfb7BH
https://sena4-

SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANALÍTICA DE DATOS-20220705_105934-Grabación de la reunión 1.mp4

SENA

REGISTRO DE ASISTENCIA Y APROBACIÓN DEL ACTA No- 718 DEL DÍA 05 DEL MES DE 07 DEL AÑO 2022

OBJETIVO (S): Realizar seguimiento a la ejecución de actividades del proyecto 9748 - Prototipo innovación

No.	NOMBRES Y APELLIDOS	No. DOCUMENTO	PLANTA	CONTRATISTA	OTRO ¿CUAL?	DEPENDENCIA/ EMPRESA	CORREO ELECTRÓNICO	TELÉFONO/EXT.SENA	AUTORIZA GRABACIÓN	FIRMA O PARTICIPACIÓN VIRTUAL
1	Jesus Arbey Benavides G.	1085260713		X		SENA	jabenavidesg@sena.edu.co	3105595765	SI	Jesus Arbey B.
2	Pabla Andrea Buitrago	43641579		X		SENA	pabuitragoc@sena.edu.co	3113671040	SI	Pabla Buitrago
3	John Jairo Castro Maldonado	88265-552	X			SENA	jcastrom@sena.edu.co.	3204185552	SI	
4	Erika Juliett Restrepo Zapata	43973982		X		SENA	erestrepoz@sena.edu.co	3005477253	SI	Erestrepo.

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales, el Servicio Nacional de Aprendizaje SENA, se compromete a garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento, y les dará el tratamiento correspondiente en cumplimiento de lo establecido legalmente.