

| | | |
|--|---|-----------------------------|
| UDECA <small>UNIVERSIDAD DE CUNDINAMARCA</small> | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 1 de 12 |

16.

| | |
|--------------|---------------------------------|
| FECHA | martes, 12 de diciembre de 2023 |
|--------------|---------------------------------|

Señores
UNIVERSIDAD DE CUNDINAMARCA
 BIBLIOTECA
 Facatativá

| | |
|---|------------------------|
| UNIDAD REGIONAL | Extensión Facatativá |
| TIPO DE DOCUMENTO | Trabajo De Grado |
| FACULTAD | Ingeniería |
| NIVEL ACADÉMICO DE FORMACIÓN O PROCESO | Pregrado |
| PROGRAMA ACADÉMICO | Ingeniería de Sistemas |

El Autor(Es):

| APELLIDOS COMPLETOS | NOMBRES COMPLETOS | No. DOCUMENTO DE IDENTIFICACIÓN |
|----------------------------|--------------------------|--|
| Vergara Serrato | Luis Felipe | 1193128701 |
| Mahecha Hernández | David Santiago | 1004006687 |
| | | |

Director(Es) y/o Asesor(Es) del documento:

| APELLIDOS COMPLETOS | NOMBRES COMPLETOS |
|----------------------------|--------------------------|
| Barahona Rodríguez | Cesar Yesid |
| | |
| | |

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca
 Teléfono: (091) 8281483 Línea Gratuita: 018000180414
www.ucundinamarca.edu.co E-mail: info@ucundinamarca.edu.co
 NIT: 890.680.062-2

*Documento controlado por el Sistema de Gestión de la Calidad
 Asegúrese que corresponde a la última versión consultando el Portal Institucional*

| | | |
|--|---|-----------------------------|
| | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 2 de 12 |

| TÍTULO DEL DOCUMENTO |
|--|
| <p>APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE DESEMPEÑO EN LA PRUEBAS SABER</p> |


| SUBTÍTULO (Aplica solo para Tesis, Artículos Científicos, Disertaciones, Objetos Virtuales de Aprendizaje) |
|---|
| |

| EXCLUSIVO PARA PUBLICACIÓN DESDE LA DIRECCIÓN INVESTIGACIÓN | |
|---|--------|
| INDICADORES | NÚMERO |
| ISBN | |
| ISSN | |
| ISMN | |

| AÑO DE EDICION DEL DOCUMENTO | NÚMERO DE PÀGINAS |
|------------------------------|-------------------|
| 20/11/2023 | 205 |

| DESCRIPTORES O PALABRAS CLAVES EN ESPAÑOL E INGLÉS (Usar 6 descriptores o palabras claves) | |
|---|----------------------|
| ESPAÑOL | INGLÉS |
| 1.Mineria de datos | Data mining |
| 2.Desempeño académico | Academic performance |
| 3.Pruebas saber | Knowledge tests |
| 4.Educacion básica | Basic education |
| 5.Prediccion | Prediction |
| 6. | |

| FUENTES (Todas las fuentes de su trabajo, en orden alfabético) |
|---|
| <p>¿Qué es PostgreSQL y por qué llevarlo a Cloud? Arsys. (s. f.). Blog de arsys.es. https://www.arsys.es/blog/postgresql-servidores#:~:text=PostgreSQL%20es%20un%20sistema%20de,consistente%20y%20tolerante%20a%20fallos.</p> <p>Acero Caro B. L., & Casas M., A. F. (2013). Análisis de las diferencias de género en el desempeño de estudiantes colombianos en matemáticas y lenguaje.”</p> <p>Adebayo, K. A., Ntokozo, N., & Grace, N. Z. (2020). Availability of Educational Resources and Student Academic Performances in South Africa. Universal Journal of Educational Research, 8(8), 3768–3781. doi: 10.13189/ujer.2020.080858</p> <p>ADMINISTRACIÓN DE BASES DE DATOS CON POSTGRESQL - Mariuxi Paola Zea Ordóñez, Jimmy Rolando Molina Ríos, Fausto Fabían Redrován Castillo - Google Libros.</p> |

| | | |
|---|---|-----------------------------|
|  | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 3 de 12 |

(n.d.). Retrieved August 17, 2022, from https://books.google.com.co/books?id=5-mkDgAAQBAJ&pg=PA79&dq=postgresql&hl=es&sa=X&ved=2ahUKEwjm_Kf59dH5AhU8fjABHaTgBToQ6AF6BAgBEAI#v=onepage&q=postgresql&f=false

Ajibade, S. S. M., & Bahiah Binti Ahmad, N. (2019). Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods. IOP Conference Series: Materials Science and Engineering, 551(1), 012061.

Alam, T. M., Mushtaq, M., & Shaukat, K. (2021). A Novel Method for Performance Measurement of Public Educational Institutions Using Machine Learning Models. Applied Sciences, 11(19), 9296.

Amrieh E. A. 2017 Int'l. Journal of Database Theory and Application 9 119-136.

Así ha afectado el Covid-19 la educación en Colombia - Forbes Colombia. (n.d.). Retrieved from <https://forbes.co/2020/04/30/actualidad/asi-ha-afectado-el-covid-19-la-educacion-en-colombia/>

Bai, X., Zhang, F., Li, J., Guo, T., Aziz, A., Jin, A., & Xia, F. (2021). Educational Big Data: Predictions, Applications and Challenges. Big Data Research, 26, 100270. <https://doi.org/10.1016/j.bdr.2021.100270>

Barrientos Rocío. (2009). Árboles de decisión como herramienta en el diagnóstico médico. <https://1library.co/document/zkw51d4m-%C3%A1rboles-decisi%C3%B3n-herramienta-diagn%C3%B3stico-m%C3%A9dico.html>

Bester, G., & Kuyper, N. (2020). The Influence of Additional Educational Support on Poverty-Stricken Adolescents' Resilience and Academic Performance. <https://doi.org/10.1080/18146627.2019.1689149>

Cabeza María Alejandra. (2004). INDICADORES DE GESTIÓN EN LA EDUCACIÓN SUPERIOR COMO HERRAMIENTA DE LA PLANIFICACIÓN ESTRATÉGICA. <http://www.redalyc.org/articulo.oa?id=36410206>

Cambi alvarado Jessica, & Zuñiga garcia Jessica. (2016). Comparacion de diferentes bases de datos.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., CRISP-DM 1.0: Step-by-step data mining guide: CRISP-DM Consortium., 2009.

Cobos, C., Zuñiga, J., Guarín, J., León, E., Mendoza, M., & English, I. (2010). CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos.

Content, R. R. (2019, May 6). ¿Qué es la usabilidad web? Conoce este concepto y por qué es importante para la estrategia digital de las empresas. Rock Content - ES; Rock Content. <https://rockcontent.com/es/blog/usabilidad/>

Contreras Bravo, L. E., Fuentes López, H. J., & Rivas Trujillo, E. (2022). Análisis del rendimiento académico mediante técnicas de aprendizaje automático con métodos de ensamble. Revista Boletín Redipe, 10(13), 171–190.

Devi, K., Ratnoo, S., & Bajaj, A. (2022). Impact of Socio-Economic Factors on Students' Academic Performance: A Case Study of Jawahar Navodaya Vidyalaya. Lecture Notes in Networks and Systems, 419 LNNS, 774–785. doi: 10.1007/978-3-030-96299-9_73

Efecto de la pandemia sobre el sistema educativo: El caso de Colombia | Banco de la República. (n.d.). Retrieved from <https://www.banrep.gov.co/es/borrador-1179>

El ODS 4: El objetivo de desarrollo sostenible de la educación | Compartir Palabra maestra. (2020). Retrieved May 11, 2022, from <https://www.compartirpalabramaestra.org/actualidad/articulos-informativos/el-ods-4-el-objetivo-de-desarrollo-sostenible-de-la-educacion>


Fulgencio Cánovas García, Francisco Alonso Sarría, & Francisco Gomariz Castillo. (2016). MODIFICACIÓN DEL ALGORITMO RANDOM FOREST PARA SU EMPLEO EN CLASIFICACIÓN DE IMÁGENES DE TELEDETECCIÓN.

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca

Teléfono: (091) 8281483 Línea Gratuita: 018000180414

www.ucundinamarca.edu.co E-mail: info@ucundinamarca.edu.co

NIT: 890.680.062-2

| | | |
|---|---|-----------------------------|
|  | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 4 de 12 |

Fundación iS+D La finalidad de la metodología cuantitativa. (2018). Retrieved from <https://isdfundacion.org/2018/11/01/finalidad-metodologia-cuantitativa/>

Gómez Moreno Antonio Steeven, Vínces Sánchez Luis Fernando, & Honores Tapia Joofer Antonio. (2018). Comparacion De Tendencias Tecnologicas En Aplicaciones.

Han, J. and Kamber, M. "Data Mining: Concepts and Techniques. Second Edition", Morgan Kaufmann Publishers, San Francisco, 2010.

Hussain, S. (2018, 1 febrero). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA | Hussain | Indonesian Journal of Electrical Engineering and Computer Science. Indonesian Journal. Ingeniería e Investigación, 30(3), 45–56. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092010000300004&lng=en&nrm=iso&tlng=es

Java. (s. f.). Home de DesarrolloWeb.com. <https://desarrolloweb.com/home/java>

Jiang, P. (2021). Gender differences in mathematics academic performance of high school students in western China. Journal of Physics: Conference Series, 1978(1). doi: 10.1088/1742-6596/1978/1/012038

Jin, Y., & Yang, X. (2021). Educational Data Mining: Discovering Principal Factors for Better Academic Performance. 2021 the 3rd International Conference on Big Data Engineering and Technology (BDET).

Khoirom, S., Sonia, M., Laikhuram, B., Laishram, J., & Singh, D. (2020). Comparative Analysis of Python and Java for Beginners Cite this paper Comparative Analysis of Python and Java for Beginners. International Research Journal of Engineering and Technology. www.irjet.net

kmeans [en línea]. (sin fecha). Universidad de Oviedo - Universidad de Oviedo, la universidad de Asturias. [Consultado el 26 de agosto de 2022]. Disponible en: https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html#:~:text=K-means%20es%20un%20algoritmo,suele%20usar%20la%20distancia%20cuadrática.

Kotomina, O. v., & Sazhina, A. I. (2021). The Influence of Family Factors on the Academic Performance of Schoolchildren and University Students: Review of Foreign Studies. Education and Self Development, 16(4), 74–92. doi: 10.26907/esd.16.4.07

La participación de los padres en la enseñanza — Observatorio | Instituto para el Futuro de la Educación. (n.d.). Retrieved from <https://observatorio.tec.mx/edu-news/la-importancia-de-la-participacion-de-los-padres-en-la-educacion>

Ley 1266 de 2008, (2008). http://www.secretariassenado.gov.co/senado/basedoc/ley_1266_2008.html

LEY NÚMERO 23 DE 1982, (1982).


Más de 600.000 estudiantes presentan las pruebas Icfes en todo el país. (n.d.). Retrieved May 8, 2022, from <https://www.larepublica.co/economia/mas-de-600000-estudiantes-presentan-las-pruebas-icfes-en-todo-el-pais-2759151>

Matzavela, V., & Alepis, E. (2021, 5 octubre). E-Biblioteca Ucindinamarca. ScienceDirect. Retrieved from <https://login.ucundinamarca.basesdedatoszproxy.com/login?qurl=https://www.sciencedirect.com%2fscience%2farticle%2fpii%2fS2666920X21000291%3fvia%253Dihub>

Metodología Scrum: cómo aplicar el método Scrum | APD. (2022). Retrieved from <https://www.apd.es/metodologia-scrum-que-es>

Mitra S, Acharya T. Data Mining. Multimedia, Soft Computing, and Bioinformatics. John Wiley & Sons, Inc., Hoboken, New Jersey; 2013.

Mitra, S., & Acharya, Tinku. (2003). Data mining: multimedia, soft computing, and bioinformatics. 401. <https://www.wiley.com/en->

| | | |
|---|---|-----------------------------|
|  | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 5 de 12 |

us/Data+Mining%3A+Multimedia%2C+Soft+Computing%2C+and+Bioinformatics-p-9780471460541

Modelos de minería de datos – Acervo Lima. (2020). Retrieved from <https://es.acervolima.com/modelos-de-mineria-de-datos/>

Moisa V. 2018 Journal of Mobile Embedded and Distributed Systems 5 70-77

Muelle, L. (2020). Socioeconomic and contextual factors associated with low academic performance of Peruvian students in PISA 2015. *Apuntes*, 47(86), 111–146. doi: 10.21678/APUNTES.86.943

Nuankaew, P., & Sararat, W. (2022). Student Performance Prediction Model for Predicting Academic Achievement of High School Students. *European Journal of Educational Research*, 11(2), 949–963.

Objetivo 4: Educación de calidad | Sustainable Development Goals Fund. (n.d.). Retrieved May 11, 2022, from <https://www.sdgfund.org/es/objetivo-4-educaci%C3%B3n-de-calidad>

Pandey, U.K. and Pal, S., 2011. Data Mining: A prediction of performer or underperformer using classification. (IJCSIT) *International Journal of Computer Science and Information Technologies*, Vol. 2 (2), 2011, 686- 690.

Patiño Pérez, D., Silva Bustillos, R., Munive Mora, C., & Botto-Tobar, M. (2020). Predicción de Covid19 con el uso del Algoritmo Random Forest y Redes Neuronales Artificiales. *Ecuadorian Science Journal*, 4(2), 101–110. <https://doi.org/10.46480/esj.4.2.41>

PostgreSQL vs Oracle: Critical Differences - Learn | Hevo [en línea]. (sin fecha). Learn | Hevo. [Consultado el 26 de agosto de 2022]. Disponible en: <https://hevo.com/learn/postgresql-vs-oracle/#support>

Regresan las Pruebas Saber 3°, 5°, 7° y 9° para cerca de 200 mil estudiantes, de 1.300 sedes educativas de todo el país. (2022, 8 abril). Icfes. Retrieved from <https://bit.ly/38y4JWw>

Regresan las Pruebas Saber 3°, 5°, 7° y 9° para cerca de 200 mil estudiantes, de 1.300 sedes educativas de todo el país. (n.d.). Retrieved May 8, 2022, from <https://www.mineducacion.gov.co/portal/salaprensa/Noticias/410085:Regresan-las-Pruebas-Saber-3-5-7-y-9-para-cerca-de-200-mil-estudiantes-de-1-300-sedes-educativas-de-todo-el-pais>

Rocha, J. L. M., Zela, M. A. C., Torres, N. I. V., Rojas, J. T. R., Valderrama, S. O. M., & Medina, G. S. (2021). Virtual Education: Impact of Socio-emotional and Pedagogical Factors on Academic Performance based on Neural Networks and Stepwise Regression. *Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology*, 2021-July. doi: 10.18687/LACCEI2021.1.1.459


Rocha, Á., Adeli, H., Dzemyda, G., Moreira, F. y Ramalho Correia, A. M., eds., (2021). *Trends and Applications in Information Systems and Technologies* [en línea]. Cham: Springer International Publishing. [Consultado el 26 de agosto de 2022]. Disponible en: doi: 10.1007/978-3-030-72651-5

Rodríguez Diego, Ordoñez Ruber, Hidalgo Mario. (2021). Academic performance determinants of high school students in the Department of Nariño, Colombia. *Lecturas de Economía*, 94, 87–126. doi: 10.17533/UDEA.LE.N94A341834

Rodríguez Suárez, Y., & Amador, A. D. (2009). *Herramientas de Minería de Datos Data Mining Tools*. 3(3), 73–80.

Rodríguez, J. V., Rodado, D. N., Crissien Borrero, T., & Parody, A. (2022). Multidimensional indicator to measure quality in education. *International Journal of Educational Development*, 89. doi: 10.1016/j.ijedudev.2021.102541

Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and

| | | |
|---|---|-----------------------------|
|  | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 6 de 12 |

predictor evaluation. Computers and Education: Artificial Intelligence, 2. doi: 10.1016/j.caeai.2021.100018

Rodríguez-Rodríguez, D., & Guzmán, R. (2019). Socio-familial risk factors and personal protective variables of academic performance in Secondary Education students. Psicothema, 31(2), 142–148. doi: 10.7334/psicothema2018.213

Rodríguez-Rodríguez, D., & Guzmán, R. (2021). Academic performance of secondary education students in socio-familial risk contexts. Suma Psicologica, 28(2), 104–111. doi: 10.14349/sumapsi.2021.v28.n2.5

Rud, O. P. (2012). Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management (1.a ed.). John Wiley & Sons Inc.

Sabzevari M. 2018 Cornell Uni. arXiv preprint arXiv:1802.07877

Samsudin, N. A. M. (2021). Modeling Student's Academic Performance During Covid-19 Based on Classification in Support Vector Machine. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(5), 1798-1804.

Sarmiento Espinel, J. A., Silva Arias, A. C., & van Gameren, E. (2019). Evolution of the inequality of educational opportunities from secondary education to university. International Journal of Educational Development, 66, 193–202. doi: 10.1016/j.ijedudev.2018.09.006

Selenium y la automatización de las pruebas. (2020). Juntadeandalucia.es. from <https://www.juntadeandalucia.es/servicios/madeja/contenido/recurso/381>

Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research, 12(1), 217-222.

Sofía J. Vallejos y Mgter. David Luis la Red Martínez. Minería de Datos. Universidad Nacional del Nordeste Facultad de Ciencias Exactas, Naturales y Agrimensura, (2006).

Subaira A. 2016 IEEE 8th Int. Conf. on Intelligent System and Control (ISCO) 978 274-280

Sunita, B., & Lobo, Mr. (2011). Data Mining in Educational System using WEKA. Data Mining in Educational System using WEKA, 20-25.

Testing con Cucumber: Cómo afianzar la fiabilidad en los desarrollos. (2020, August 10). Chakray. <https://www.chakray.com/es/testing-cucumber-afianzar-fiabilidad-desarrollos/>

Thierry Groussard. (n.d.). JAVA 7: Los fundamentos del lenguaje Java - Thierry Groussard - Google Libros. Retrieved August 17, 2022, from https://books.google.es/books?hl=es&lr=lang_es&id=JaPTzKZxbN4C&oi=fnd&pg=PA165&dq=JAVA&ots=pW5Bwhvk2b&sig=r0C1scygJio3-HS2oHIMSxFBhJQ#v=onepage&q=JAVA&f=false

Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. (2020). Academic performance patterns of middle school students in the knowledge natural science test 11 with decision trees. RISTI - Revista De Sistemas e Tecnologias De Informacao, 2020(E32), 190-201.

Vanessa Berlanga Silvente, & María José Rubio Hurtado. (2013). Cómo aplicar árboles de decisión en SPSS. <http://diposit.ub.edu/dspace/bitstream/2445/43762/1/618361.pdf>

Verma, M., Srivastava, M., Chack, N., Kumar Diswar, A., & Gupta, N. (2012). A Comparative Study of Various Clustering Algorithms in Data Mining. In International Journal of Engineering Research and Applications (IJERA) (Vol. 2). www.ijera.com

Wang, Y., Xia, M., Guo, W., Xu, F., & Zhao, Y. (2022). Academic performance under COVID-19: The role of online learning readiness and emotional competence. Current Psychology. doi: 10.1007/s12144-022-02699-7


Witten, I. H., Frank, E., & Hall, M. (2005). Data mining: Practical machine learning tools and techniques. Miami: Elsevier

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca

Teléfono: (091) 8281483 Línea Gratuita: 018000180414

www.ucundinamarca.edu.co E-mail: info@ucundinamarca.edu.co

NIT: 890.680.062-2

| | | |
|---|---|-----------------------------|
|  | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 7 de 12 |

Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.). Morgan Kaufmann Publishers.

Zeng, L., & Luo, H. (2021). Online Academic Performance during the COVID-19: Evidence from a Rural High School in Western China.

RESUMEN DEL CONTENIDO EN ESPAÑOL E INGLÉS

(Máximo 250 palabras – 1530 caracteres, aplica para resumen en español):

Resumen

En el presente proyecto de investigación se quiere determinar factores e indicadores de desempeño académico, por ende, se plantea el uso de minería de datos dentro del aprendizaje escolar ya que este campo permite:

- Predecir y analizar los resultados académicos de los estudiantes.
- Analizar el desempeño estudiantil.
- Identificar patrones que permitan determinar el rendimiento académico.
- Asociar el proceso de formación con los datos socioeconómicos y familiar del estudiante.

Analizando que el uso de modelos de minería de datos son de gran fiabilidad ya que permiten detectar información oculta, ya que no es fácil encontrar en bases de datos por la gran cantidad de información que se encuentran en estas.

Para mayor seguridad del modelo se trabajará con información pública que viene del ICFES, estos datos contarán no solamente con los resultados generales y por núcleos temáticos, sino que por otro lado se dispondrán también de datos familiares, socioeconómicos y personales de los estudiantes, los cuales serán tratados y más adelante relacionados entre sí; con el fin de determinar la trayectoria de estos y determinar en donde hay una mayor agrupación de la información.

Abstract

In this research project we want to determine factors and indicators of academic performance, therefore, the use of data mining within school learning is proposed since this field allows:

- predict and analyze the academic results of students.
- Analyze student performance.
- Identify patterns to determine academic performance.
- Associate the training process with the student's socio-economic and family data.


Analyzing that the use of data mining models are very reliable as they allow to detect hidden information, since it is not easy to find in databases because of the large amount of information that are in them.

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca

Teléfono: (091) 8281483 Línea Gratuita: 018000180414

www.ucundinamarca.edu.co E-mail: info@ucundinamarca.edu.co

NIT: 890.680.062-2

| | | |
|---|---|-----------------------------|
|  | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 8 de 12 |

For greater security of the model will work with public information coming from the ICFES, these data will not only have the general results and thematic cores, but on the other hand will also have family, socioeconomic and personal data of the students, which will be treated and later related to each other; in order to determine the trajectory of these and determine where there is a greater grouping of information.


AUTORIZACIÓN DE PUBLICACIÓN

Por medio del presente escrito autorizo (Autorizamos) a la Universidad de Cundinamarca para que, en desarrollo de la presente licencia de uso parcial, pueda ejercer sobre mí (nuestra) obra las atribuciones que se indican a continuación, teniendo en cuenta que, en cualquier caso, la finalidad perseguida será facilitar, difundir y promover el aprendizaje, la enseñanza y la investigación.

En consecuencia, las atribuciones de usos temporales y parciales que por virtud de la presente licencia se autoriza a la Universidad de Cundinamarca, a los usuarios de la Biblioteca de la Universidad; así como a los usuarios de las redes, bases de datos y demás sitios web con los que la Universidad tenga perfeccionado una alianza, son: Marque con una "X":

| AUTORIZO (AUTORIZAMOS) | SI | NO |
|--|----|----|
| 1. La reproducción por cualquier formato conocido o por conocer. | X | |
| 2. La comunicación pública, masiva por cualquier procedimiento o medio físico, electrónico y digital. | X | |
| 3. La inclusión en bases de datos y en sitios web sean éstos onerosos o gratuitos, existiendo con ellos previa alianza perfeccionada con la Universidad de Cundinamarca para efectos de satisfacer los fines previstos. En este evento, tales sitios y sus usuarios tendrán las mismas facultades que las aquí concedidas con las mismas limitaciones y condiciones. | X | |
| 4. La inclusión en el Repositorio Institucional. | X | |

De acuerdo con la naturaleza del uso concedido, la presente licencia parcial se otorga a título gratuito por el máximo tiempo legal colombiano, con el propósito de que en dicho lapso mi (nuestra) obra sea explotada en las condiciones aquí estipuladas y para los fines indicados, respetando siempre la titularidad de los derechos patrimoniales y morales correspondientes, de acuerdo con los usos

| | | |
|---|---|-----------------------------|
|  | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 9 de 12 |

honrados, de manera proporcional y justificada a la finalidad perseguida, sin ánimo de lucro ni de comercialización.

Para el caso de las Tesis, Trabajo de Grado o Pasantía, de manera complementaria, garantizo(garantizamos) en mi(nuestra) calidad de estudiante(s) y por ende autor(es) exclusivo(s), que la Tesis, Trabajo de Grado o Pasantía en cuestión, es producto de mi(nuestra) plena autoría, de mi(nuestro) esfuerzo personal intelectual, como consecuencia de mi(nuestra) creación original particular y, por tanto, soy(somos) el(los) único(s) titular(es) de la misma. Además, aseguro (aseguramos) que no contiene citas, ni transcripciones de otras obras protegidas, por fuera de los límites autorizados por la ley, según los usos honrados, y en proporción a los fines previstos; ni tampoco contempla declaraciones difamatorias contra terceros; respetando el derecho a la imagen, intimidad, buen nombre y demás derechos constitucionales. Adicionalmente, manifiesto (manifestamos) que no se incluyeron expresiones contrarias al orden público ni a las buenas costumbres. En consecuencia, la responsabilidad directa en la elaboración, presentación, investigación y, en general, contenidos de la Tesis o Trabajo de Grado es de mí (nuestra) competencia exclusiva, eximiendo de toda responsabilidad a la Universidad de Cundinamarca por tales aspectos.

Sin perjuicio de los usos y atribuciones otorgadas en virtud de este documento, continuaré (continuaremos) conservando los correspondientes derechos patrimoniales sin modificación o restricción alguna, puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación de los derechos patrimoniales derivados del régimen del Derecho de Autor.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “*Los derechos morales sobre el trabajo son propiedad de los autores*”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables. En consecuencia, la Universidad de Cundinamarca está en la obligación de RESPETARLOS Y HACERLOS RESPETAR, para lo cual tomará las medidas correspondientes para garantizar su observancia.

NOTA: (Para Tesis, Trabajo de Grado o Pasantía):

Información Confidencial:

Esta Tesis, Trabajo de Grado o Pasantía, contiene información privilegiada, estratégica, secreta, confidencial y demás similar, o hace parte de la investigación que se adelanta y cuyos resultados finales no se han publicado.

SI NO

| | | |
|--|---|-----------------------------|
| | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 10 de 12 |

En caso afirmativo expresamente indicaré (indicaremos) en carta adjunta, expedida por la entidad respectiva, la cual informa sobre tal situación, lo anterior con el fin de que se mantenga la restricción de acceso.

LICENCIA DE PUBLICACIÓN

Como titular(es) del derecho de autor, confiero(erimos) a la Universidad de Cundinamarca una licencia no exclusiva, limitada y gratuita sobre la obra que se integrará en el Repositorio Institucional, que se ajusta a las siguientes características:

- a) Estará vigente a partir de la fecha de inclusión en el repositorio, por un plazo de 5 años, que serán prorrogables indefinidamente por el tiempo que dure el derecho patrimonial del autor. El autor podrá dar por terminada la licencia solicitándolo a la Universidad por escrito. (Para el caso de los Recursos Educativos Digitales, la Licencia de Publicación será permanente).
- b) Autoriza a la Universidad de Cundinamarca a publicar la obra en formato y/o soporte digital, conociendo que, dado que se publica en Internet, por este hecho circula con un alcance mundial.
- c) Los titulares aceptan que la autorización se hace a título gratuito, por lo tanto, renuncian a recibir beneficio alguno por la publicación, distribución, comunicación pública y cualquier otro uso que se haga en los términos de la presente licencia y de la licencia de uso con que se publica.
- d) El(Los) Autor(es), garantizo(amos) que el documento en cuestión es producto de mi(nuestra) plena autoría, de mi(nuestro) esfuerzo personal intelectual, como consecuencia de mi (nuestra) creación original particular y, por tanto, soy(somos) el(los) único(s) titular(es) de la misma. Además, aseguro(aseguramos) que no contiene citas, ni transcripciones de otras obras protegidas, por fuera de los límites autorizados por la ley, según los usos honrados, y en proporción a los fines previstos; ni tampoco contempla declaraciones difamatorias contra terceros; respetando el derecho a la imagen, intimidad, buen nombre y demás derechos constitucionales. Adicionalmente, manifiesto (manifestamos) que no se incluyeron expresiones contrarias al orden público ni a las buenas costumbres. En consecuencia, la responsabilidad directa en la elaboración, presentación, investigación y, en general, contenidos es de mí (nuestro) competencia exclusiva, eximiendo de toda responsabilidad a la Universidad de Cundinamarca por tales aspectos.
- e) En todo caso la Universidad de Cundinamarca se compromete a indicar siempre la autoría incluyendo el nombre del autor y la fecha de publicación.
- f) Los titulares autorizan a la Universidad para incluir la obra en los índices y buscadores que estimen necesarios para promover su difusión.

| | | |
|--|---|-----------------------------|
| | MACROPROCESO DE APOYO | CÓDIGO: AAAR113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 11 de 12 |

g) Los titulares aceptan que la Universidad de Cundinamarca pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

h) Los titulares autorizan que la obra sea puesta a disposición del público en los términos autorizados en los literales anteriores bajo los límites definidos por la universidad en el “Manual del Repositorio Institucional AAAM003”

i) Para el caso de los Recursos Educativos Digitales producidos por la Oficina de Educación Virtual, sus contenidos de publicación se rigen bajo la Licencia Creative Commons: Atribución- No comercial- Compartir Igual.



j) Para el caso de los Artículos Científicos y Revistas, sus contenidos se rigen bajo la Licencia Creative Commons Atribución- No comercial- Sin derivar.



Nota:

Si el documento se basa en un trabajo que ha sido patrocinado o apoyado por una entidad, con excepción de Universidad de Cundinamarca, los autores garantizan que se ha cumplido con los derechos y obligaciones requeridos por el respectivo contrato o acuerdo.

La obra que se integrará en el Repositorio Institucional está en el(los) siguiente(s) archivo(s).

| Nombre completo del Archivo Incluida su Extensión (Ej. Nombre completo del proyecto.pdf) | Tipo de documento (ej. Texto, imagen, video, etc.) |
|---|---|
| 1. APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE DESEMPEÑO EN LA PRUEBAS SABER.pdf | Texto |
| 2. | |
| 3. | |
| 4. | |

En constancia de lo anterior, Firmo (amos) el presente documento:

| | | |
|--|---|-----------------------------|
| | MACROPROCESO DE APOYO | CÓDIGO: AAAr113 |
| | PROCESO GESTIÓN APOYO ACADÉMICO | VERSIÓN: 6 |
| | DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL | VIGENCIA: 2021-09-14 |
| | | PAGINA: 12 de 12 |

| APELLIDOS Y NOMBRES COMPLETOS | FIRMA (autógrafo) |
|--------------------------------------|--------------------------|
| Mahecha Hernández David Santiago | |
| Vergara Serrato Luis Felipe | |
| | |
| | |
| | |

21.1-51-20.

**APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE DESEMPEÑO
EN LAS PRUEBAS SABER**

**LUIS FELIPE VERGARA SERRATO
DAVID SANTIAGO MAHECHA HERNÁNDEZ**

**UNIVERSIDAD DE CUNDINAMARCA FACULTAD DE INGENIERÍA PROGRAMA
INGENIERÍA DE SISTEMAS FACULTATIVO
2023**

**APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE DESEMPEÑO
EN LAS PRUEBAS SABER**

**LUIS FELIPE VERGARA SERRATO
DAVID SANTIAGO MAHECHA HERNÁNDEZ**

DIRECTOR:

CESAR YESID BARAHONA RODRIGUEZ
Ingeniero en telecomunicaciones, Especialista de Gestión de Proyectos

**UNIVERSIDAD DE CUNDINAMARCA FACULTAD DE INGENIERÍA PROGRAMA
INGENIERÍA DE SISTEMAS FACULTATIVA
2023**

AGRADECIMIENTOS

Inicialmente agradezco a mis padres y a mis hermanos por haberme apoyado durante el logro de mis metas y objetivos, también quiero agradecer al profesor Cesar Yesid Barahona quien con todo su conocimiento me guio durante la realización del proyecto de grado, adicionalmente por haber confiado y creído en mis habilidades y capacidades.

Agradezco igualmente a todos los docentes que me acompañaron durante toda esta de aprendizaje, llenaron de conocimientos y experiencias mi formación y por supuesto me impulsaron a llegar a este punto.

COMPROMISO DEL AUTOR

Yo, **Vergara Serrato Luis Felipe** identificado con cédula de ciudadanía No. **1193128701** y con cód. **461219185** y **Mahecha Hernández David Santiago** identificado con cedula de ciudadanía No. **1004006687** y con cód. **461219239** estudiantes del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, declaro que: El contenido del presente documento es un reflejo de mi trabajo personal y manifiesto que, ante cualquier notificación de plagio, copia o falta a la fuente original, soy responsable directo legal, económico y administrativo sin afectar al Director del trabajo, a la Universidad y a cuantas instituciones hayan colaborado en dicho trabajo, asumiendo las consecuencias derivadas de tales prácticas.

Firma



RESUMEN

El presente proyecto está enfocado a determinar qué factores influyen en el desempeño en la educación básica primaria y secundaria, para ello se tuvieron en cuenta los resultados académicos y socioeconómicos de las pruebas saber 3°, 5°, 9° y 11°; con esta información se realizó por medio de técnicas de minería de datos en la educación, un análisis en el cual se identifica la incidencia de cada variable sobre el comportamiento de los estudiantes. Para el análisis descriptivo se implementó la herramienta power bi la cual ayuda con la visualización de los resultados de esta manera se identificaron los factores que tienen incidencia en el desempeño académico de los estudiantes.

Asimismo, se realizó un análisis predictivo con la información socioeconómica de los estudiantes, se realizó un análisis por medio del algoritmo J48, el cual da como resultado un árbol de decisión; de este se extrajeron las reglas predictivas de los puntajes a obtener de acuerdo con cada una de las variables analizadas.

Con la información obtenida de los análisis descriptivos y predictivos se procedió a realizar un aplicativo, en el cual se muestra gráficamente los resultados y la relación de estos frente a cada año y tipo de prueba.

Palabras clave

Minería de datos, desempeño académico, pruebas saber, educación básica, predicción.

ABSTRACT

This project is focused on determining the factors that influence performance in primary and secondary education. To achieve this, academic and socioeconomic results from the Saber tests in 3rd, 5th, 9th, and 11th grades were taken into account. Using data mining techniques in education, an analysis was conducted to identify the impact of each variable on students' performance. For the descriptive analysis, the Power BI tool was implemented to aid in visualizing the results. This helped identify the factors that affect students' academic performance.

Furthermore, predictive analysis was carried out using socioeconomic information from the students. An analysis was conducted using the J48 algorithm, which resulted in a decision tree. Predictive rules for scores were extracted based on each of the analyzed variables. Using the information obtained from the descriptive and predictive analyses, an application was developed to graphically display the results and their relationship with each year and type of test.

Keywords

Data mining, academic performance, knowledge tests, basic education, clustering, prediction

Contenido

| | |
|---|----|
| Introducción..... | 18 |
| I. Informe De Investigación..... | 18 |
| I.I Estado del arte | 18 |
| I.II. Línea de investigación..... | 25 |
| I.III. Planteamiento del problema y pregunta de investigación | 26 |
| I.IV. Objetivo general y objetivos específicos..... | 27 |
| I.V. Alcance e impacto del proyecto..... | 27 |
| I.VI. Metodología..... | 27 |
| I.VII. Marcos de referencia | 28 |
| I.VII.I Marco teórico | 28 |
| I.VII.I.I Indicadores de desempeño..... | 29 |
| I.VII.I.II. Minería de datos | 29 |
| I.VII.I.III. Técnicas de minería de datos | 30 |
| I.VII.I.IV. Árboles de decisión | 30 |
| I.VII.I.V. Clustering..... | 31 |
| I.VII.I.VI. Random forest..... | 32 |
| I.VII.I.VII. Metodologías | 32 |
| I.VII.I.VII.I. CRISP-DM | 33 |
| I.VII.I.VIII. WEKA..... | 33 |
| I.VII.I.IX. JAVA | 34 |
| I.VII.I.X. PostgreSQL | 36 |
| I.VII.II. Marco legal | 39 |
| I.VII.II.I. Ley 23 de 1982 | 39 |
| I.VII.II.III. Decisión 351 de 1993 Acuerdo de Cartagena | 41 |
| I.VII.II.V. Ley 1712 de 2014..... | 41 |
| II. Limpieza De Datos | 41 |
| II.I. Script de limpieza | 42 |
| II.II. Selección del archivo | 42 |
| II.III. Procesamiento de datos | 42 |
| II.IV. Descarga de los datos limpios | 43 |

| | | |
|--------------|--|----|
| III. | Selección De Indicadores | 43 |
| III.I. | Configuración Weka | 43 |
| III.II. | Selección De Indicadores | 46 |
| III.II.I | Indicador Acceso A Internet | 46 |
| III.II.II | Indicador Acceso A Computador..... | 48 |
| IV. | Algoritmo De Analisis Predictivo | 49 |
| IV.I. | Validación algoritmo predictivo | 50 |
| IV.I.I | Prueba indicadorde estrato socioeconómico..... | 50 |
| IV.I.I.I | Algoritmo J48 | 50 |
| IV.I.I.II. | Algoritmo RandomForest | 51 |
| IV.I.I.III. | Algoritmo RandomTree..... | 52 |
| IV.I.I.IV. | Algoritmo REPTree | 53 |
| IV.I.II. | Prueba de indicador de acceso a internet..... | 53 |
| IV.I.II.I. | Algoritmo J48..... | 53 |
| IV.I.II.II. | Algoritmo RandomForest..... | 54 |
| IV.I.II.III. | Algoritmo RandomTree..... | 55 |
| IV.I.II.IV. | Algoritmo REPTree..... | 55 |
| V. | Documentación Del Software | 56 |
| V.I. | Determinación de requerimientos..... | 56 |
| V.II. | Especificación del diseño | 59 |
| V.II.I. | Diagrama de clases de la base de datos | 59 |
| V.II.I.I. | Modelo entidad relación de la base de datos..... | 60 |
| V.II.I.II | Modelo relacional de la base de datos..... | 61 |
| V.II.II. | Roles..... | 61 |
| V.II.II.I. | Administrador..... | 61 |
| V.II.II.II. | Usuario | 62 |
| V.II.II.III. | Aplicativo | 62 |
| V.II.III. | Diagramas de casos de uso..... | 62 |
| V.I.III.I. | Casos de uso: Usuario..... | 62 |
| V.I.III.II. | Casos de uso: Administrador..... | 64 |
| V.I.III.III. | Casos de uso: Aplicativo..... | 67 |

| | | |
|--------------|---|-----|
| V.I.IV. | Diagramas de secuencia..... | 67 |
| V.I.IV.I | Diagrama de secuencia de inicio de sesión | 68 |
| V.I.IV.II | Diagrama de secuencia de cerrar sesión | 69 |
| V.I.IV.III | Diagrama de secuencia de creación de usuarios | 70 |
| V.I.IV.IV | Diagrama de secuencia de consultar usuarios | 71 |
| V.I.IV.V | Diagrama de secuencia de modificar usuarios | 72 |
| V.I.IV.VI | Diagrama de secuencia de eliminar usuarios | 74 |
| V.I.IV.VII | Diagrama de secuencia de cargarlos resultados..... | 75 |
| V.I.IV.VIII | Diagrama de secuencia de consultar listado de pruebas | 75 |
| V.I.IV.IX | Diagrama de secuencia del análisis predictivo | 76 |
| V.I.IV.X | Diagrama de secuencia de los reportes..... | 77 |
| V.II.V. | Diagramas de actividad..... | 77 |
| V.II.V.I. | Diagrama de actividad de inicio de sesión | 78 |
| V.II.V.II. | Diagrama de actividad de cerrar sesión..... | 79 |
| V.II.V.III. | Diagrama de actividad de la creación de usuarios | 80 |
| V.II.V.IV. | Diagrama de actividad de modificación de usuarios..... | 81 |
| V.II.V.V. | Diagrama de actividad de consultar los usuarios..... | 83 |
| V.II.V.VI. | Diagrama de actividad de la eliminación de usuarios | 84 |
| V.II.V.VII. | Diagrama de actividad del cargue de los resultados | 85 |
| V.II.V.VIII. | Diagrama de actividad de consultar el listado de pruebas..... | 86 |
| V.II.V.IX. | Diagrama de actividad del análisis predictivo | 87 |
| V.II.V.X. | Diagrama de actividad de los reportes..... | 88 |
| V.III. | Diseño de los casos de prueba | 88 |
| V.III.I. | Usabilidad del aplicativo web | 88 |
| V.III.I.I. | MAXQDA | 92 |
| V.III.II. | Pruebas de verificación del software | 97 |
| V.III.II.I. | Sonarcloud..... | 102 |
| VI. | Estimación de recursos con punto de función COSMIC | 105 |
| VI.I. | Identificación de Requerimientos funcionales: | 105 |
| VI.II. | Identificación de datos de entrada y salida | 105 |
| VI.III. | Estimación de puntos de función | 106 |

| | | |
|----------------|--|-----|
| VI.IV. | Roles de equipo | 106 |
| VI.V. | Horas de Trabajo por Semana (por Investigador): | 107 |
| VI.VI. | Semanas de Trabajo para 10 meses | 107 |
| VI.VII. | Tasas de Pago por Hora: | 107 |
| VI.VIII. | Determinación de Esfuerzo:..... | 107 |
| VI.IX. | Distribución del Esfuerzo: | 107 |
| VI.X. | Estimación de Costos Por Rol (para 10 meses):..... | 107 |
| VI.XI. | Costo Total para 10 meses: | 107 |
| VI.XII. | Valor Monetario Total para 10 meses (Incluyendo Gastos Asociados): | 107 |
| VII. | Resultados..... | 108 |
| VII.I. | Resultados descriptivos | 109 |
| VII.I.I. | Saber 3° | 109 |
| VII.I.I.I. | Puntaje en lenguaje | 109 |
| VII.I.I.II. | Puntaje matemáticas | 111 |
| VII.I.II. | Saber 5 | 113 |
| VII.I.II.I. | Puntaje lenguaje | 113 |
| VII.I.II.II. | Puntaje matemáticas | 115 |
| VII.I.II.III. | Puntaje ciencias naturales..... | 117 |
| VII.I.III. | Saber 9 | 118 |
| VII.I.III.I. | Puntaje lenguaje | 118 |
| VII.I.III.II. | Puntaje matemáticas | 120 |
| VII.I.III.III. | Puntaje ciencias naturales..... | 122 |
| VII.I.IV. | Saber 11 | 123 |
| VII.I.IV.I. | Indicador Acceso a computador | 123 |
| VII.I.IV.II. | Indicador Acceso a internet | 124 |
| VII.I.IV.III. | Indicador educación padres..... | 125 |
| VII.I.IV.IV. | Indicador trabajo padres | 126 |
| VII.I.IV.V. | Indicador estrato vivienda..... | 127 |
| VII.II. | Resultados predictivos saber once | 128 |
| VII.II.I. | Indicador acceso a computador | 128 |
| VII.II.II. | Indicador acceso a internet..... | 129 |

| | |
|---|-----|
| VII.II.III. Indicador educación padres..... | 129 |
| VII.II.IV. Indicador estrato vivienda | 130 |
| Referencias | 132 |

LISTA DE TABLAS

| | |
|---|----|
| Tabla 1. Fuentes de indicadores de desempeño estudiantil | 21 |
| Tabla 2. Fuentes obtenidas de minería de datos | 24 |
| Tabla 3. Java vs Python vs go vs php vs c#..... | 34 |
| Tabla 4. PosgreSQL vs Oracle vs MySQL | 36 |
| Tabla 5. Requerimientos de software | 56 |
| Tabla 6. Diagrama de clases de la Base de datos resultados básico | 59 |
| Tabla 7. Descripción relaciones | 60 |
| Tabla 8. Relaciones base de datos..... | 61 |
| Tabla 9. Descripción de caso de uso inicio de sesión..... | 62 |
| Tabla 10. Descripción de caso de uso reportes | 63 |
| Tabla 11. Descripción de caso de uso cerrar sesión | 63 |
| Tabla 12. Descripción de caso de uso inicio de sesión..... | 64 |
| Tabla 13. Descripción de caso de uso cerrar sesión | 65 |
| Tabla 14. Descripción de caso de uso creación de usuarios | 65 |
| Tabla 15. Descripción de caso de uso consulta de usuarios | 65 |
| Tabla 16. Descripción de caso de uso modificación de usuarios..... | 65 |
| Tabla 17. Descripción de caso de uso eliminación usuarios..... | 66 |
| Tabla 18. Descripción de caso de uso cargar pruebas..... | 66 |
| Tabla 19. Descripción de caso de uso consultar listado de pruebas | 66 |
| Tabla 20. Caso de uso análisis predictivo..... | 67 |
| Tabla 21. Caso de uso Limpieza de datos..... | 67 |
| Tabla 22. Descripción diagrama de secuencia de inicio de sesión..... | 68 |
| Tabla 23. Descripción diagrama de secuencia cerrar sesión..... | 69 |
| Tabla 24. Descripción diagrama de secuencias de creación de usuarios..... | 70 |
| Tabla 25. Descripción diagrama de secuencia: consulta usuarios | 71 |
| Tabla 26. Descripción diagrama de secuencia: modificación usuarios..... | 73 |

| | | |
|-----------|--|-----|
| Tabla 27. | Descripción diagrama de secuencia: eliminación usuarios..... | 74 |
| Tabla 28. | Descripción diagrama de secuencia: consultar listado de pruebas..... | 76 |
| Tabla 29. | Descripción diagrama de actividad: inicio de sesión..... | 78 |
| Tabla 30. | Descripción diagrama de actividad: cerrar sesión..... | 79 |
| Tabla 31. | Descripción diagrama de actividad: creación usuarios | 80 |
| Tabla 32. | Descripción diagrama de actividad: modificación usuarios..... | 82 |
| Tabla 33. | Descripción diagrama de actividad: consulta usuarios | 83 |
| Tabla 34. | Descripción diagrama de actividad: eliminación usuarios..... | 84 |
| Tabla 35. | Descripción diagrama de actividad: consultar listado de pruebas..... | 86 |
| Tabla 36. | Descripción diagrama de actividad: análisis predictivo | 87 |
| Tabla 37. | Análisis de Cobertura de Segmentos en Entrevistas de Usuarios | 93 |
| Tabla 38. | Análisis de Cobertura de Segmentos en Entrevistas de Usuarios | 94 |
| Tabla 39. | Análisis de Cobertura de Segmentos en Entrevistas de Usuarios | 95 |
| Tabla 40. | Análisis de Cobertura de Segmentos en Entrevistas de Usuarios | 96 |
| Tabla 41. | Análisis de Cobertura de Segmentos en Entrevistas de Usuarios | 97 |
| Tabla 42. | Valor en puntos de función | 106 |

LISTA DE FIGURAS

| | | |
|------------|--|----|
| Figura 1. | Analisis por escuela rural o urbana..... | 19 |
| Figura 2. | Modelo random forest..... | 32 |
| Figura 3. | Función de limpieza de datos | 42 |
| Figura 4. | Configuración del archivo “DatabaseUtils.props” | 44 |
| Figura 5. | Interfaz inicial de weka..... | 44 |
| Figura 6. | Interfaz explorer..... | 45 |
| Figura 7. | Menu clasificadores weka..... | 45 |
| Figura 8. | Analisis algoritmo simpleKmeans | 46 |
| Figura 9. | Ejemplo de clasificación Acceso a internet | 47 |
| Figura 10. | Clusters del indicador de acceso a internet | 47 |
| Figura 11. | Ejemplo de clasificación acceso a computador | 48 |
| Figura 12. | Clusters Indicador acceso a computador..... | 48 |
| Figura 13. | Selección del Algoritmo J48..... | 49 |
| Figura 14. | Arbol generado con el algoritmo J48 | 50 |
| Figura 15. | Porcentaje de clasificación Algoritmo J48..... | 50 |
| Figura 16. | Porcentaje de clasificación algoritmo randomForest | 51 |
| Figura 17. | Porcentaje de clasificación algoritmo randomTree | 52 |
| Figura 18. | Algoritmo de clasificación REPTree..... | 53 |
| Figura 19. | Porcentaje de clasificación algoritmo J48 | 53 |
| Figura 20. | Porcentaje de clasificación algoritmo RandomForest | 54 |
| Figura 21. | Porcentaje de clasificación del algoritmo RandomTree | 55 |
| Figura 22. | Porcentaje de clasificación del algoritmo REPTree | 55 |
| Figura 23. | Modelo del diagrama de clases de la base de datos AWID | 59 |
| Figura 24. | MER base de datos | 60 |
| Figura 25. | Modelo relacional base de datos awid | 61 |
| Figura 26. | Diagrama casos de uso usuario..... | 62 |
| Figura 27. | Diagrama caso de uso administrador | 64 |
| Figura 28. | Diagrama de secuencia de inicio de sesión..... | 68 |
| Figura 29. | Diagrama de secuencia: cerrar Sesión | 69 |
| Figura 30. | Diagrama de secuencia: creación usuarios | 70 |

| | | |
|------------|--|----|
| Figura 31. | Diagrama de secuencia: consulta usuarios | 71 |
| Figura 32. | Diagrama de secuencia de la modificación de usuarios | 72 |
| Figura 33. | Diagrama de secuencia de eliminar usuarios | 74 |
| Figura 34. | Diagrama de secuencia: cargar pruebas | 75 |
| Figura 35. | Diagrama de secuencia: consultar listado de pruebas..... | 75 |
| Figura 36. | Diagrama de secuencia: análisis predictivo | 76 |
| Figura 37. | Diagrama de secuencia: reportes..... | 77 |
| Figura 38. | Diagrama de actividad: inicio de sesión..... | 78 |
| Figura 39. | Diagrama de actividad: cerrar sesión | 79 |
| Figura 40. | Diagrama de actividad: creación usuarios | 80 |
| Figura 41. | Diagrama de actividad: modificación usuarios..... | 81 |
| Figura 42. | Diagrama de actividad: consulta usuarios | 83 |
| Figura 43. | Diagrama de actividad: eliminación usuarios..... | 84 |
| Figura 44. | Diagrama de actividad: cargar resultados..... | 85 |
| Figura 45. | Diagrama de actividad: consultar listado de pruebas..... | 86 |
| Figura 46. | Diagrama de actividad: análisis predictivo | 87 |
| Figura 47. | Diagrama de actividad: reportes | 88 |
| Figura 48. | Encuesta forms parte 1 | 89 |
| Figura 49. | Encuesta forms parte 2 | 89 |
| Figura 50. | Encuesta forms parte 3..... | 90 |
| Figura 51. | Evidencia de que el correo no existe | 90 |
| Figura 52. | Encuesta forms parte 4..... | 91 |
| Figura 53. | Encuesta forms parte 5..... | 91 |
| Figura 54. | Sugerencias..... | 92 |
| Figura 55. | Evidencias de los registros | 92 |
| Figura 56. | Pregunta N°1 encuesta..... | 93 |
| Figura 57. | Pregunta N°2 encuesta..... | 94 |
| Figura 58. | Pregunta N°3 encuesta..... | 95 |
| Figura 59. | Pregunta N°4 encuesta..... | 96 |
| Figura 60. | Pregunta N°5 encuesta..... | 97 |
| Figura 61. | Descarga de dependencias | 98 |

| | | |
|------------|--|-----|
| Figura 62. | Casos de prueba..... | 99 |
| Figura 63. | Casos de prueba ejecutados | 99 |
| Figura 64. | Automatización login..... | 100 |
| Figura 65. | Automatizacion registro | 100 |
| Figura 66. | Validacion prueba once | 101 |
| Figura 67. | Automatización validación pruebas 3,5,9..... | 101 |
| Figura 68. | Resultado análisis del Proyecto en sonarcloud | 102 |
| Figura 69. | Total de archivos | 102 |
| Figura 70. | Vulnerabilidades encontradas..... | 103 |
| Figura 71. | Resultados puntaje en lenguaje saber 3° con nivel de pobreza año 2016. | 109 |
| Figura 72. | Resultados puntaje en lenguaje saber 3° con nivel de pobreza año 2017. | 110 |
| Figura 73. | Resultados puntaje en matemáticas saber 3° con nivel de pobreza año 2016 | 111 |
| Figura 74. | resultados puntaje en matemáticas saber 3° con nivel de pobreza año 2017 | 112 |
| Figura 75. | resultados puntaje lenguaje saber 5° con nivel de pobreza año 2016 | 113 |
| Figura 76. | resultados puntaje lenguaje saber 5° con nivel de pobreza año 2017 | 114 |
| Figura 77. | Resultados puntaje matemáticas saber 5° con nivel de pobreza año 2016 | 115 |
| Figura 78. | Resultados puntaje matemáticas saber 5° con nivel de pobreza año 2017 | 116 |
| Figura 79. | Resultados ciencias naturales saber 5° con nivel de pobreza año 2016 ... | 117 |
| Figura 80. | Resultados lenguaje saber 9° con nivel de pobreza año 2016 | 118 |
| Figura 81. | Resultados lenguaje saber 9° con nivel de pobreza año 2017 | 119 |
| Figura 82. | Resultados matematicas saber 9° con nivel de pobreza año 2016..... | 120 |
| Figura 83. | Resultados matematicas saber 9° con nivel de pobreza año 2017..... | 121 |
| Figura 84. | Resultados ciencias naturales saber 9° con nivel de pobreza año 2016 ... | 122 |
| Figura 85. | Puntaje global vs Indicador acceso a computador..... | 123 |
| Figura 86. | Puntaje global vs Indicador acceso a internet..... | 124 |
| Figura 87. | Puntaje global vs Indicador educación padres..... | 125 |
| Figura 88. | Puntaje global vs Indicador trabajo padres | 126 |
| Figura 89. | Puntaje global vs Indicador estrato vivienda..... | 127 |

| | |
|--|-----|
| Figura 90. Árbol j48 puntaje global vs indicador acceso a computador | 128 |
| Figura 91. Árbol j48 puntaje global Vs indicador acceso a internet..... | 129 |
| Figura 92. Árbol j48 puntaje global Vs indicador educación padres..... | 129 |
| Figura 93. Árbol j48 puntaje global Vs indicador estrato vivienda..... | 130 |

Tabla de Anexos

| | |
|--------------------------|-----|
| Manual de usuarios | 150 |
| Manual técnico | 170 |
| Artículo 1 | 185 |
| Artículo 2 | 198 |

Introducción

El desempeño académico en la educación básica primaria y secundaria es un factor crítico en el desarrollo de cualquier sociedad. La calidad de la educación es fundamental para el crecimiento social, económico y científico de un país. A lo largo de los años, se ha reconocido que existen diferencias significativas en el rendimiento académico de los estudiantes, y estas diferencias pueden estar influenciadas por diversos factores socioeconómicos, demográficos y familiares.

En Colombia, como en muchos otros lugares del mundo, la educación desempeña un papel crucial en la formación de las generaciones futuras. Sin embargo, se han identificado disparidades en los métodos de enseñanza y aprendizaje, que a menudo están vinculados al entorno familiar y social en el que crecen los estudiantes. Identificar estos factores y comprender cómo afectan al rendimiento académico es esencial para diseñar estrategias efectivas que mejoren la calidad de la educación.

Este proyecto tiene como objetivo principal analizar los factores que inciden en el desempeño académico de los estudiantes en la educación básica primaria y secundaria en Colombia. Para lograrlo, se utilizarán técnicas de minería de datos educativos para identificar indicadores que influyen en el rendimiento estudiantil. Estos indicadores serán fundamentales para comprender mejor los procesos de formación de los jóvenes y para desarrollar estrategias que ayuden a cerrar las brechas en el desempeño académico.

Es importante destacar el presente proyecto busca identificar por medio de técnicas de minería de datos educativa que factores o indicadores influyen en la relación entre el desempeño académico y una serie de factores que van entorno a lo socioeconómico. Además, se utilizará un aplicativo web desarrollado en JAVA para presentar de manera gráfica los resultados de los análisis realizados

I. Informe De Investigación

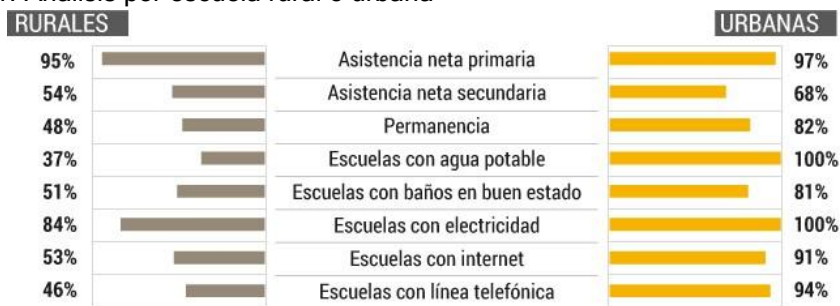
I.1 Estado del arte

El instituto colombiano para la educación (ICFES), es el encargado de ofrecer servicios para la evaluación de la educación en los distintos niveles que tiene Colombia. Las pruebas saber 3°, 5° y 9° tienen como principal objetivo contribuir al mejoramiento de la calidad de la educación colombiana, mientras que la prueba saber 11° es la evaluación del nivel de educación media del país y tiene como objetivo proporcionar información sobre el desarrollo de las competencias básicas que un estudiante desarrolla durante su vida escolar.

El rendimiento académico es una medida de las capacidades de un estudiante, que se puede expresar con lo aprendido en un periodo de tiempo. Rodríguez, Ordoñez e Hidalgo afirman que, el acceso a herramientas tecnológicas de aprendizaje como computadoras y conexión a internet, el mayor nivel educativo de los padres de familia, la condición de ser varón y el estudiar en una institución educativa oficial urbana aumenta la probabilidad de obtener un mejor rendimiento académico.

El nivel socioeconómico es el conjunto de factores, como lo son los ingresos, el patrimonio y las condiciones generales del entorno en las que vive un ser humano. el componente social es un factor crucial para mejorar la calidad de la educación y que variables como las actividades extracurriculares apoyen la formación integral de los estudiantes para impactar positivamente en el rendimiento académico (Velásquez, Neira y Crissien. 2022). Devi, K., Ratnoo, S., & Bajaj, A. afirman, que las variables socioeconómicas de los estudiantes, como la casta, la residencia y la ocupación del padre, impactan su rendimiento académico en el sexto grado. Por otro lado, Muelle afirma, La condición social del alumno y la composición social de su escuela destacan como los factores que afectan mayormente el bajo rendimiento, asociados a factores contextuales como la repetición, la lengua materna, la matrícula oportuna, la dimensión de la escuela, el ausentismo y el género. En Colombia la educación escolar se puede dividir en colegios públicos y privados, es decir que los públicos los sostiene económicamente el estado, mientras que en la privada se sostiene gracias a los mismos alumnos, por medio de sus padres o algún acudiente legal. Por otra parte, también se puede dividir en escuelas rurales y urbanas.

Figura 1. Analisis por escuela rural o urbana



Desempeño en pruebas Piza escuelas



Fuente: (Tomado del periódico colombiano el Espectador)

Como se puede notar en la gráfica la desigualdad en recursos económicos es muy grande y se ve reflejado en el rendimiento académico de los estudiantes. Esto es algo habitual en casi todo el mundo, como es el caso en Sudáfrica el sistema educativo se caracteriza por escuelas acomodadas y otras que no lo están. Las escuelas acomodadas se desempeñan a un ritmo mucho mejor, esto conlleva a que se vuelva un factor en el rendimiento académico de los estudiantes (Adebayo, K. A., Ntokozo, N., & Grace, N. Z. 2020). Por otra parte, Sarmiento Espinel, J. A., Silva Arias, A. C., & van Gameraen, E, afirman que, la desigualdad de oportunidades educativas ha aumentado con el tiempo. La resiliencia, es la capacidad que tiene una persona para superar circunstancias traumáticas, esta se relaciona positivamente con el rendimiento académico y otros factores como la relación profesor alumno, la participación de los padres y los métodos de estudio (Bester, G., & Kuyper, N. 2020).

El rendimiento académico aumenta. Según JHU (Johns Hopkins University), aspectos como la comprensión y la fluidez lectora mejoran cuando hay participación de los padres, aún más si los papás dedican tiempo para leer con sus hijos, ya que los alumnos saben que sus papás están al pendiente, tratan de mejorar por ellos, se sienten más motivados a aprender y mejorar sus calificaciones.(Instituto para el Futuro de la Educación, 2019). Además, La relación que tienen las variables sociofamiliares y no cognitivas sobre los estudiantes en cuanto a su rendimiento académico es un elemento muy importante para el éxito en la Educación Secundaria (Rodríguez-Rodríguez, D., & Guzmán, R. 2021). En Rusia un estudio realizado por Kotomina, O. v., & Sazhina, A. afirma que, la revisión radica en la consideración de tres formas en que la familia impacta en desempeño del estudiante: el estatus socioeconómico de la familia, el capital social de la familia, y la participación de los padres en el proceso educativo. Las dos primeras formas han sido ampliamente estudiadas en la investigación, mientras que la participación de los padres a menudo se considera como un factor significativo en el rendimiento escolar.

Algo que afectó al mundo en general fue la pandemia, por temas de contagios, a la gran mayoría de los colegios el anuncio de cierre de clases los tomó por sorpresa y sin previa preparación para desarrollar su programa de educación a distancia. Fueron muy pocos los colegios los que ya tenían un programa de aprendizaje remoto listo para ser implementado (Forbes Colombia, 2020). Estudios realizados en China por Wang, Y., Xia, M., Guo, W., Xu, F., & Zhao, Y afirman que, los padres dedicaron más del doble del tiempo normal a apoyar el aprendizaje y el desarrollo de sus hijos durante el período de COVID-19. Se encontró que los factores de apoyo y motivación de los padres son la contribución más efectiva en el desarrollo de las emociones positivas de los niños y el logro del aprendizaje. Por otra parte, otro estudio en China Los resultados indicaron que el aprendizaje en línea no aumentó el rendimiento académico en el bachillerato rural y se observó una caída significativa del rendimiento en matemáticas e inglés (Zeng, L., & Luo, H. 2021).

Tabla 1. Fuentes de indicadores de desempeño estudiantil

| Indicadores | Autores |
|---|--|
| Acceso a tecnologías | Rodríguez, Ordoñez e Hidalgo |
| Conexión a internet | Rodríguez, Ordoñez e Hidalgo |
| Nivel educativo de los padres | Rodríguez, Ordoñez e Hidalgo |
| Genero | Rodríguez, Ordoñez e Hidalgo Muelle |
| Institución (Urbana o rural) | Rodríguez, Ordoñez e Hidalgo |
| Ocupación del padre | Devi K, Ratnoo S, & Bajaj A |
| La residencia | Devi K, Ratnoo S, & Bajaj A |
| Condición social | Velásquez, Neira y Crissien Muelle |
| Composición social de la escuela | Muelle Adebayo, K. A., Ntokozo, N., & Grace, N. Z. |
| Ausentismo | Muelle |
| Participación de los padres | Sarmiento Espinel, J. A., Silva Arias, A.C., & van Gameren, E Kotomina, O. v., & Sazhina |
| Desigualdad de oportunidades educativas | Sarmiento Espinel, J. A., Silva Arias, A.C., & van Gameren, E JHU (Johns Hopkins University) |
| Estatus socioeconómico de la familia | Kotomina, O. v., & Sazhina |
| Capital social de la familia | Kotomina, O. v., & Sazhina |
| COVID | Wang, Y., Xia, M., Guo, W., Xu, F., & Zhao Zeng, L., & Luo, H. |
| Variables sociofamiliares | Rodríguez-Rodríguez, D., & Guzmán, R |

Fuente: (Tomado del periódico colombiano el Espectador)

Ahora bien, frente a la medición del desempeño de las diferentes instituciones educativas, se encuentran varios estudios en el cual se relacionan los diferentes indicadores y la manera de analizar dichos datos, según la investigación de Nuankaew y Sararat (2022) realizada a 1859 estudiantes de la escuela Manchasuksa en el distrito de Mancha Khiri, provincia de Khon Kaen, Tailandia, durante el año académico 2015-2020, en donde las herramientas de investigación están separadas en 2 secciones. La primera sección es un paso básico de análisis estadístico, este se compone de análisis de frecuencia, análisis de porcentaje, análisis de media y análisis de desviación estándar. Otra sección es la fase de análisis de minería de datos, que consiste en la técnica de discretización, la técnica de clasificación XGBoost (árbol de decisión, árboles potenciados por gradiente y random forest), análisis de rendimiento de matriz de confusión y análisis de rendimiento de validación cruzada. Además, uno de los temas de enfoque en la investigación de Vasiliki Matzavela y Efthimios Alepis (2021) trata de la Minería de Datos Educativos (EDM) la cual es una aplicación de Técnicas de Minería de Conocimiento a partir de datos educativos, y su objeto es analizar datos, con el fin de resolver problemas de investigación en el campo de la Educación. Sus datos provienen de diferentes fuentes, como bases de datos de sistemas educativos, sistemas de Internet. Por otro lado, se evidencia un tema de gran importancia, los árboles de decisión pues son una de las técnicas usadas como modelo predictivo del Rendimiento Académico de los Estudiantes en entornos Inteligentes de M-Learning, en el área del aprendizaje automático y la ciencia de datos, es uno de los métodos más populares dentro de las técnicas de clasificación por ser fácil de entender e interpretar por medio de gráficas, así como también el manejo de datos numéricos y categóricos. Los sistemas M-Learning se consolidan recientemente como uno de los métodos de mayor interés para una educación mas efectiva y un aprendizaje adaptativo proporcional a las habilidades de aprendizaje de cada estudiante. Un clasificador de árbol de decisión es uno de los métodos de aprendizaje supervisado más utilizados para la exploración de datos, aproximando una función por regiones constantes a trozos, y no necesita información previa de la distribución de los datos esto según Mitra S y Acharya T. (2013). En el estudio de Witten et al. (2011) los modelos de árboles de decisión son comúnmente utilizados en la minería de datos para examinar los datos e inducir el árbol y sus reglas que serán utilizadas para hacer las predicciones. Según (Rud, 2012) el verdadero propósito de los árboles de decisión es clasificar los datos en grupos distintos o ramas que generen la separación más fuerte en los valores de la variable dependiente, siendo superiores en identificar segmentos con un comportamiento deseado como la respuesta o la activación, proporcionando así una solución fácilmente interpretable. Teniendo en cuenta lo anterior, la minería de datos cuenta con diferentes sistemas que contribuyen al descubrimiento de los factores principales para un mejor rendimiento académico, en las últimas décadas se han generado avances significativos de las nuevas tecnologías en el ámbito educativo entre las cuales la Minería de Datos Educativo- EDM,

juega un papel indispensable para la búsqueda del mejoramiento pedagógico, permitiendo a los investigadores por medio de bases de datos agrupar variables que ayuden a identificar los factores que influyen en el rendimiento académico de los estudiantes, algunos de estos como la información demográfica de los estudiantes, la disposición o voluntad de aprendizaje y la interacción familiar, usando EDM como regresión lineal, regresión árbol, random forest y red neuronal.(Yucheng-jin y Xiaomeng Yang, 2021) Con los conceptos tratados por Han, J. and Kamber, M. (2010) en los últimos años, el campo de la minería de datos se vuelve muy importante para diferentes industrias, corporaciones y empresas, etc. debido a su capacidad para utilizar una gran cantidad de datos que antes no tenían uso y respecto de los cuales se pueden realizar análisis, predicción de tendencias y patrones. Para determinar los indicadores de desempeño se tendrán presentes las técnicas de minería de datos implementado la herramienta WEKA como lo hicieron los investigadores Sadiq Hussain y Neama Abdulaziz (2018) para la selección de los atributos o factores y de este modo permitiendo clasificar la información donde los autores determinaron mediante los resultados que el algoritmo de clasificación random forest destaca en precisión. Para confirmar que tecnología es más óptima en el estudio del desempeño académico de los estudiantes se pondrán a prueba varios métodos de selección de características para así extraer los indicadores fundamentales tomando como ejemplo el artículo de Talha Mahboob, Mubbashar Mushtaq y Kamran Shaukat (2021) donde en este estudio proponen un método novedoso para la medición del desempeño de las instituciones educativas haciendo uso de varios modelos de aprendizaje automático como los árboles de decisión, bosque de rotación, bosque aleatorio, entre otros, ya que esto permite adaptar los diferentes factores a tratar según las necesidades del estudio. Según en el artículo de Umair Shafique y Haseeb Qaiser (2014) se puede centrar la investigación haciendo uso de tres modelos de procesos de minería de datos que son muy populares y que principalmente son empleados por expertos e investigadores en minería de datos los cuales son Knowledge Discovery Databases (KDD), CRISP-DM y SEMMA. De acuerdo a ello, en la investigación de Samsudin (2021) propone hacer uso de una máquina de vectores de soporte el cual es un algoritmo de aprendizaje que sirve para determinar patrones de desempeño académico durante la pandemia de COVID-19. En el siguiente artículo se explica el proceso de aplicación de la metodología CRISP-DM para detectar factores relacionados con el rendimiento académico de estudiantes colombianos quienes presentaron las pruebas saber 11. Principalmente se construye un repositorio o base de datos con la información socioeconómica y académica disponible por el ICFES, para luego realizar una limpieza de este, se realizó un modelo de clasificación basado en arboles de decisión para predecir los patrones asociados con el buen o bajo rendimiento académico Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. (2020). En la investigación que realizaron Pandey y Pal (2015) de minería de datos utilizando clasificación de Naïve Bayes para analizar, clasificar y predecir estudiantes de alto y bajo rendimiento donde La clasificación de Naïve Bayes se usa como

técnica probabilística simple, que asume que todos los atributos dados en un conjunto de datos son independientes entre sí, de ahí el nombre "Naïve". Pandey y Pal (2015) realizaron esta investigación con una muestra de datos de estudiantes matriculados en un Diploma de Postgrado en Aplicaciones Informáticas Aplicaciones informáticas (PGDCA) en la Universidad Dr. R. M. L. Awadh, Faizabad, India. De acuerdo con el artículo de Ajbade y Bahiah Binti Ahmad (2019) para construir un modelo predictivo, se utilizan varias técnicas de la minería de datos, que son la clasificación, la regresión y la agrupación. Para la investigación de los dos autores mencionados eligieron parámetros como las notas internas, las notas de las sesiones y la puntuación de admisión además hicieron uso del algoritmo de aprendizaje SVM (máquina de vectores de soporte) desarrollado por Subaira (2016) que sirve para manejar los desafíos del reconocimiento de patrones y la predicción como también para el análisis y mapeo de funciones, asimismo se usó el algoritmo random forest igual que en el trabajo investigativo de Sabzevari M. (2018). Ahora bien, este método trata de una colección de algoritmos de árboles de decisión que no están correlacionados. Random Forest genera una gran cantidad de árboles de decisión a partir de subconjuntos del conjunto de datos a estudiar donde cada subconjunto proporciona un árbol de decisión. Ahora, cada modelo de árbol de decisión clasifica una instancia en una clase y así la clase más votada se tomada como instancia, es decir, de reiterar o ser insistente en una orden dada, todo esto es conforme a lo dicho por Amrieh E. A. (2017). En la investigación de Moisa V. (2018) utilizaron algunas técnicas de conjuntos como Bagging, Adaboosting y random forest para predecir el rendimiento académico de los estudiantes con mayor precisión. La finalidad inicial del trabajo de Contreras Leonardo y Fuentes Héctor (2022) es predecir el rendimiento académico de estudiantes mediante técnicas de aprendizaje automático donde se analizan 324 variables con métodos de selección de características, con el objetivo de determinar las variables más destacadas. El modelo de predicción del rendimiento académico universitario es estudiado por medio de algoritmos supervisados como (KNN, SVC, Naive Bayes y árbol de decisión), los cuales son optimizados mediante lenguaje Python. Además, son implementados algoritmos de ensamble que permiten mejorar la exactitud de los clasificadores previos, también se implementan métodos Bagging (CART, Random Forest), métodos Boosting (AdaBoost, GBM, XGBoost). A continuación, veremos una tabla resumiendo las técnicas de minerías de datos con sus autores:

Tabla 2. Fuentes obtenidas de minería de datos

| | | |
|------------------|-------------------|--|
| Minería de datos | Arbol de decisión | Mitra S, Acharya T. Data Mining 2013. |
| | | Witten, I. H., Frank, E., & Hall, M. A. (2011). |
| | | Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. (2020). |
| | | Rud, O. P. (2012). Data Mining Cookbook |
| | | Alam, T. M., Mushtaq, M., & Shaukat, K. (2021). |
| | | Amrieh E. A. 2017. Database Theory and Application 9 119-136. |

| | | |
|--|---|---|
| | | Nuankaew, P., & Sararat, W. (2022). Student Performance Prediction Model for Predicting Academic |
| | | Contreras Bravo, L. E., Fuentes López, H. J., & Rivas Trujillo, E. (2022). Análisis del rendimiento académico |
| | Random forest | Moisa V. 2018 Journal of Mobile Embedded and Distributed Systems 5 70-77 |
| | | Amrieh E. A. 2017. Database Theory and Application 9 119-136. |
| | | Subaira A. 2016 IEEE 8th Int. Conf. on Intelligent System and Control (ISCO) 978 274-280 |
| | | Alam, T. M., Mushtaq, M., & Shaukat, K. (2021). |
| | | Hussain, S. (2018, 1 febrero). Educational Data Mining and Analysis of Students |
| | | Sabzevari M. 2018 Cornell Uni. arXiv preprint arXiv:1802.07877 |
| | | Nuankaew, P., & Sararat, W. (2022). Student Performance Prediction Model for Predicting Academic |
| | Naive Bayes | Contreras Bravo, L. E., Fuentes López, H. J., & Rivas Trujillo, E. (2022). Análisis del rendimiento académico |
| | | Pandey, U.K. and Pal, S., 2011. Data Mining: A prediction of performer or underperformer using classification. |
| | vectores soporte de | Samsudin, N. A. M. (2021). Modeling Student's Academic Performance. |
| | | .Subaira A. 2016 IEEE 8th Int. Conf. on Intelligent System and Control (ISCO) 978 274-280 |
| Ajibade, S. S. M., & Bahiah Binti Ahmad, N. (2019). Educational Data Mining: Enhancement of Student Performance. | | |
| Metodos Bagging | .Moisa V. 2018 Journal of Mobile Embedded and Distributed Systems 5 70-77 | |
| Minería de datos educativos | Random forest | Matzavela, V., & Alepis, E. (2021, 5 octubre). |
| | | [Jin, Y., & Yang, X. (2021). Educational Data Mining |
| Metodologías | WEKA | Hussain, S. (2018, 1 febrero). Educational Data Mining and Analysis of Students |
| | KDD, CRISP-DM y SEMMA | Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models. |
| | | Han, J. and Kamber, M. "Data Mining: Concepts and Techniques. Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. (2020). |

Fuente: (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

I.II. Línea de investigación

El presente proyecto “Aplicativo web para la construcción de indicadores de desempeño en las pruebas saber”, se realiza bajo la línea traslocal de investigación de aprendizaje, conocimiento, tecnologías, comunicación y digitalización; la cual se orienta a través de líneas de investigación especialmente por los grupos de investigación busca desarrollar proyectos investigativos que den soluciones tecnológicas a problemas y necesidades del

entorno social, a partir del conocimiento y la innovación. a fin de orientar, articular y generar conocimiento.

I.III. Planteamiento del problema y pregunta de investigación

La educación es un elemento fundamental para el desarrollo próspero y equitativo de cualquier sociedad. En el contexto colombiano, se han identificado diversos desafíos y obstáculos que afectan el acceso a una educación de calidad y, en última instancia, el desempeño académico de los estudiantes.

En primer lugar, es importante destacar que la educación en Colombia está vinculada de manera significativa a consideraciones económicas. Estudiar en colegios privados en el país implica costos considerables, lo que limita severamente el acceso a esta opción educativa para una parte significativa de la población. Esta disparidad en los recursos disponibles para la educación crea una brecha sustancial en las oportunidades de aprendizaje entre aquellos que pueden pagar una educación privada de alta calidad y aquellos que dependen de la educación pública.

Por otro lado, la inversión y la infraestructura en las instituciones educativas públicas son insuficientes y, en algunos casos, precarias. Esto se traduce en la falta de acceso a recursos y ambientes educativos adecuados para muchos estudiantes, lo que a su vez influye en su capacidad para aprender y alcanzar un desempeño académico óptimo.

El Objetivo de Desarrollo Sostenible número 4, centrado en la educación de calidad, subraya factores que explican la insuficiencia de la educación en Colombia. Estos factores incluyen la formación deficiente de los docentes, las deficientes condiciones de infraestructura en las instituciones educativas y la inequidad en el acceso a la educación, especialmente en las áreas rurales. Estos problemas estructurales impactan negativamente en la calidad de la educación que se ofrece en el país.

En este contexto, el Estado colombiano lleva a cabo las pruebas Saber en los grados tercero, quinto y noveno con el propósito de recopilar información que permita mejorar la calidad de la educación. Además, el ICFES administra pruebas en el grado once para evaluar las aptitudes de los estudiantes en materias clave como lectura crítica, matemáticas, estudios sociales, ciencias e inglés.

El presente proyecto se enfoca en abordar la problemática de cuáles son los factores o indicadores que inciden en el desempeño de los estudiantes en estas pruebas estandarizadas del Estado colombiano, específicamente durante los años 2017, 2018, 2019, 2020 y 2021. El objetivo principal es identificar estos indicadores y comprender cómo influyen en el rendimiento académico de los estudiantes en la educación primaria y secundaria del país. Este análisis se llevará a cabo mediante la aplicación de técnicas de minería de datos, lo que permitirá obtener una visión más precisa de las variables que inciden en el desempeño estudiantil. La finalidad última es proporcionar una base de conocimiento que pueda utilizarse para diseñar estrategias y políticas educativas más efectivas y orientadas a mejorar la calidad de la educación en Colombia.

¿Cuáles son las variables que explican la variación en el rendimiento de los estudiantes en las pruebas Saber, en la educación primaria y secundaria del país?

I.IV. Objetivo general y objetivos específicos

Desarrollar un aplicativo web para el análisis y la construcción de indicadores de desempeño en el país, de acuerdo, a los datos dados por el ICFES de las pruebas saber (tercero, quinto, noveno y once)

- Realizar los requerimientos para el funcionamiento del aplicativo de tal manera que se adapten para el análisis de los resultados de las pruebas.
- Implementar un modelo descriptivo y predictivo de los indicadores de desempeño, por medio de técnicas de minería de datos.
- Diseñar y desarrollar un aplicativo web para dar a conocer los resultados del análisis basado en los indicadores obtenidos en las pruebas de estado (tercero, quinto, noveno y once).
- Ejecutar pruebas de verificación del software para determinar la correcta funcionalidad.
- Analizar escenarios de prueba del modelo predictivo usando las técnicas de minería de datos.

I.V. Alcance e impacto del proyecto

En el marco de la investigación de brechas de desempeño académico en el país, se destaca el Objetivo de Desarrollo Sostenible planteados por la ONU tal como:

Objetivo 4 “Educación de calidad”. Dentro de este objetivo se plantea comometa principal para el 2030, eliminar las disparidades de género en la educación y asegurar el acceso igualitario a todos los niveles de enseñanza y formación profesional.

Por consiguiente, con ayuda del portal de datos abiertos del ICFES se desarrollará un modelo descriptivo que mida el desempeño académico e identifique los factores asociados a la problemática, posteriormente incorporar la información obtenida en un modelo predictivo, el cual ayude en la toma de decisiones para la mejora en la calidad y la equidad de la educación colombiana desde los centros educativos hasta la sociedad en general.

I.VI. Metodología

Este proyecto se desarrollará bajo la metodología cuantitativa, la cual se centra en aspectos observables para describir o explicar fenómenos sociales, y utiliza la estadística para el análisis de datos. Esta metodología está compuesta por un conjunto de métodos y técnicas. (La finalidad de la metodología cuantitativa, 2018) Se elige esta metodología porque permite contabilizar los datos por medio de análisis matemático arrojando

resultados con una precisión definida, además puede ser generalizable a partir de la muestra de una población determinada.

Para encontrar la relación entre ciertos indicadores con el desempeño de los estudiantes se busca realizar un modelo descriptivo, con el cual se logra distinguir relaciones o patrones en los datos, sirve como una forma de explorar las propiedades de los datos que se examinan. la agrupación en clústeres, el resumen, las reglas de asociación y el descubrimiento de secuencias son tareas de minería de datos del modelo descriptivo. (Modelos de minería de datos, 2020). El modelo predictivo se conoce como regresión estadística. Es una técnica de aprendizaje de monitoreo que incorpora una explicación de la dependencia de algunos valores de atributo sobre los valores de otros atributos en un elemento similar y el crecimiento de un modelo que puede predecir estos valores de atributo para casos recientes. (Modelos de minería de datos, 2020), durante este proceso de construcción del modelo se evaluará cual algoritmo tiene mejor porcentaje de acierto, con el fin de obtener resultados lo más precisos posibles.

En la construcción del aplicativo web se implementará elementos a metodología de desarrollo ágil SCRUM, ya que es un modelo general de gestión de entornos de producción basados en rutinas, donde el factor más importante es el conocimiento aplicado de las personas que hacen parte del proyecto, con este método de trabajo lo que se pretende es alcanzar el mejor resultado de un proyecto determinado. (Metodología Scrum, 2022)

Dentro de la metodología propuesta se incorporara el uso de las técnicas descriptivas y predictivas de la minería de datos, tales como los árboles de decisión. Para ello se tendrá en cuenta el modelo CRISP-DM “considerada la metodología más usada en el desarrollo de proyectos en los ambientes académico e industrial” (Timaran- Pereira et al., 2017), partiendo en primera instancia con el análisis del negocio el cual se enfoca en determinar los objetivos del negocio según el contexto y la situación actual, en segunda instancia la comprensión de los datos, en la cual se evalúa los atributos y las variables más importantes, en tercera instancia la preparación de los datos, aquí se hace una selección y limpieza de los datos, con el fin de obtener datos más consistentes y equilibrados para el desarrollo, en cuartainstancia el modelado, en esta fase se incorporaran técnicas de minera de datos acorde a lo datos seleccionados anteriormente, en quinta instancia la evaluación enesta fase se valida el modelo desarrollado y por último la explotación en el cual se incorpora el modelo dentro de una aplicación para la demostración al usuario final.

I.VII. Marcos de referencia

I.VII.I Marco teórico

En este apartado se explican criterios enlazados con el desarrollo del proyecto de gran importancia como los siguientes:

I.VII.I.I. Indicadores de desempeño

Los indicadores de desempeño en el contexto de este proyecto se pueden aclarar, comúnmente, como una expresión visible de un rasgo o como una propiedad de una o más variables de tendencia, sujetas a valoración, donde estas suministran información cuantitativa y cualitativa acerca de cierta característica. A partir de esta descripción el sistema o programa de indicadores se puede definir como un grupo de variables y de clases las cuales representan el funcionamiento de una unidad de análisis como puede ser una investigación (Cabeza María Alejandra, 2004).

Por otro lado, mostraremos los indicadores de desempeño que harán parte de este estudio los cuales fueron identificamos al realizar una investigación, donde los autores definen que consta el mismo.

- Estatus socioeconómico de los padres: El estatus socioeconómico describe las condiciones económicas y sociales de una familia e incluye los ingresos familiares y el nivel educativo de los padres (Kotomina, O. v., & Sazhina).
- Capital social: a través del capital social familiar, el cual proporciona un clima propicio para el aprendizaje como los es tener a los padres como a sus hermanos en la familia (Kotomina, O. v., & Sazhina).
- Pandemia o covid 19: acceso a internet y a computador (Wang, Y., Xia, M.,).
- Desigualdad de oportunidades: estratos socioeconómicos y la ocupación de los padres (Sarmiento espinel).
- Ausentismo: Los estudiantes que nunca faltaron a clase se desempeñaron mejor que sus compañeros de clase que sí lo hicieron. En términos de probabilidad, las tres razas parecen tener fortalezas similares. Es, por supuesto, un fenómeno complejo que requiere de otros incentivos sociales, culturales, económicos y sociales. (Muelle).
- Composición social de la escuela: La diferencia en los resultados entre las escuelas públicas y privadas es una indicación de que la socioeconomía y las limitaciones de recursos juegan un papel vital en el rendimiento académico de los estudiantes (Adebayo). En el caso del indicador de ausentismo y composición social de la escuela no se pueden aplicar en este estudio ya que los datos obtenidos en los resultados de las pruebas saber en la bases de datos no nos suministra la información necesaria para poder identificar la relación con el desempeño académico de los estudiantes.

I.VII.I.II. Minería de datos

Es un proceso en el cual se extrae conocimiento útil y comprensible, mediante métodos y técnicas, a partir de grandes volúmenes de datos, con el objeto de predecir de forma autorizada tendencias y comportamientos o descubrir de forma automatizada modelos previamente desconocidos. Por otra parte, una definición más concreta indica que la minería de datos se centra en la búsqueda de patrones, relaciones, reglas y excepciones útiles las cuales son obtenidas mediante árboles o reglas de clasificación, técnicas de

regresión, clusterizado, modelizado secuencial, dependencias, etc. (Rodríguez Suárez & Amador, 2009).

La minería de datos (MD) se puede definir como una técnica o mecanismo que consiste en el descubrimiento de relaciones, patrones y tendencias de este modo se está tratando de entender la cantidad de información que es posible almacenar la cual es significativamente amplia hoy en día, ya que es un paso crítico en el proceso que consta de funciones como clasificación, regresión, etc. Muestra actividades típicas en la fase de minería de datos, que pueden ser descriptivas, como descubrir patrones o relaciones interesantes, describir datos o predecir tendencias, modelar y clasificar nuevos datos basados en datos adquiridos previamente. En otras palabras, este campo se enfoca en predecir resultados y encontrar relaciones entre datos, y es interdisciplinario en su naturaleza. (Mitra & Acharya, 2010).

Ahora bien, para poder determinar la relación de los indicadores ya establecidos con el desempeño académico de los estudiantes, se hará uso de las diferentes técnicas o algoritmos como los árboles de decisión, random forest o vectores de soporte, entre otros, que nos ofrece la minería de datos, ya que como se menciona antes, estas ayudan a procesar la información hallando irregularidades o patrones en el conjunto de datos empleado en el estudio.

I.VII.I.III. Técnicas de minería de datos

En su trabajo, Rodríguez Suárez y Díaz Amador describen los métodos de recolección de datos utilizados en herramientas de minería de datos, donde destacan el Data Warehousing y el Análisis Exploratorio de Datos (EDA) como los principales. Además, mencionan las características de las redes neuronales y cómo estas pueden manejar datos incompletos, destacando dos formas de enseñanza o aprendizaje: el aprendizaje supervisado y el no supervisado (Rodríguez Suárez & Díaz Amador, 2009).

Además, los autores describen las características de las redes neuronales y cómo estas son capaces de manejar datos incompletos, lo que puede ser una ventaja o desventaja dependiendo del problema a resolver. Las redes neuronales tienen dos formas de enseñar o aprender, el primero es el aprendizaje supervisado, el cual consiste en predecir el valor de un atributo a partir de datos existentes, y el segundo es el aprendizaje no supervisado, que se enfoca en descubrir patrones y tendencias en los datos actuales (Rodríguez Suárez & Amador, 2009).

I.VII.I.IV. Árboles de decisión

Según Barrientos Rocío (2009), el árbol de decisiones es un modelo predictivo utilizado para representar y clasificar una secuencia de condiciones para resolver problemas. Además, Vanesa Berlanga Silvente y María José Rubio Hurtado (2013) explican que este modelo es una estructura analítica que permite analizar los resultados y visualizarlos para determinar el flujo del modelo. Asimismo, los autores indican que los árboles de decisión se emplean para diversas finalidades, tales como dividir y estratificar datos, predecir

resultados, reducir y filtrar variables, determinar interactivamente, fusionar categorías y discretizar variables continuas.

En el estudio realizado por Vanesa Berlanga Silvente y María José Rubio Hurtado (2013), se utilizó la creación de clasificación y árboles de decisiones para definir conjuntos, explorar relaciones entre estos conjuntos y predecir eventos futuros.

La estructura del árbol de decisión consta de las siguientes partes:

- **Nodo de decisión:** este nodo indica qué decisión debe tomarse en este punto del proceso.
- **Nodo de Probabilidad:** Este nodo representa el momento en que ocurre un evento aleatorio en el proceso. La probabilidad de un evento que puede ocurrir como resultado de una decisión.
- **Nodo Terminal:** Este es el nodo donde la variable dependiente tiene el mismo valor para todas las instancias.
- **Ramificación:** Muestra los diferentes caminos que puede tomar una decisión o evento aleatorio.

I.VII.I.V. Clustering

El clustering o el agrupamiento es una técnica de minería de datos que reúne o colecciona datos dentro de un numero de clases establecidas o no. Además, su uso proporciona resultados importantes en clasificadores o reconocimiento de patrones, como el modelado de sistemas. (Rodríguez Suárez & Amador, 2009).

Es el proceso de agrupar datos similares. El algoritmo de agregación divide el conjunto de datos en múltiples grupos de modo que la similitud dentro de un grupo sea mayor que entre los otros grupos. Además, gran parte de los datos recopilados sobre múltiples temas parecen tener características específicas que tienden a permanecer juntas.

“Pasos algorítmicos para el clustering K-means

Inicialización: se escoge el número de grupos el cual se identifica con la k, ahora se establecen k centroides en el espacio de los datos, escogiéndolos aleatoriamente.

Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroide más cercano.

$$\sqrt{(\sum_{i=0}^n [(x_i - y_i)]^2)}$$

Función objeto

$$J = \sum_{j=1}^k \sum_{i=1}^n (| [x_i]^{(j)} - c_j |)^2$$

Recalcular centroides: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Se repiten los pasos 2 y 3 hasta que los centroides no se mueven, o se mueven por debajo de una distancia umbral en cada paso.

Este algoritmo soluciona el problema de optimización minimizando las distancias de cada objeto al centroide.” (Verma et al., 2012), (Universidad de Oviedo).

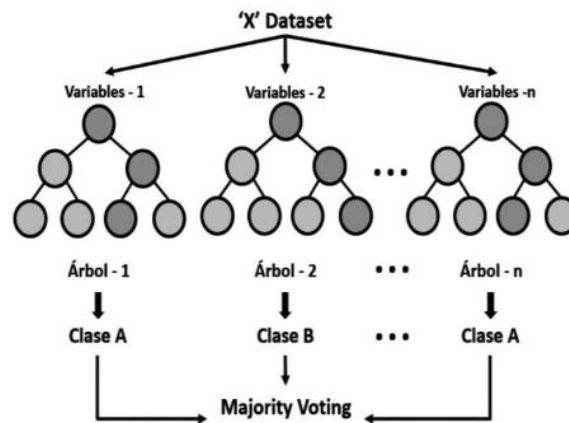
I.VII.I.VI. Random forest

Bosque aleatorio, Son bosques de decisión aleatorios, contruidos a partir de un conjunto de árboles de decisión creados por un algoritmo que introduce aleatoriedad, es decir extraer propiedades medibles (como la falta de tendencia o correlación) para reducir la consistencia entre los árboles.

Tiene un algoritmo que tiene un buen precedente por ser una técnica de aprendizaje supervisado y genera diferentes árboles de decisión basados en el conjunto de datos de entrenamiento que harán predicciones en base a lo que quieras ajustar.

Random Forest presenta la ventaja de ser una técnica de uso extendido en diversos campos, por ejemplo, en teledetección como en el trabajo investigativo de (Fulgencio Cánovas García et al., 2016), como también en las finanzas para así pronosticar comportamientos de los mercados financieros, esto quiere decir, que se puede aplicar para predecir conductas en el desempeño académico.

Figura 2. Modelo random forest



Fuente. (Patiño Pérez et al., 2020)

I.VII.I.VII. Metodologías

En las siguientes metodologías se evidenciará que son de utilidad para el desarrollo del proyecto en cuestión, ya que este requiere de las etapas o pasos a seguir que serán expuestos a continuación.

En la minería de datos, pretenden promover la implementación de nuevos proyectos con características similares (por ejemplo, esquemas de optimización), y los más visibles e implementados son KDD, CRISP-DM y SEMMA. Con esto en mente, se puede decir que

estos métodos siguen los principios de selección y preparación de datos y, por lo tanto, proceden a la fase de descubrimiento y evaluación del esquema. Por otro lado, Una fase fundamental en la minería de datos es la etapa de descubrimiento de base de conocimiento (KDD), la cual se sirve de técnicas tanto tradicionales como de inteligencia artificial para procesar grandes volúmenes de datos y descubrir conocimiento relevante para los usuarios, con el propósito de lograr sus objetivos. (Sofía J. Vallejos y Mgter. David).

También, Entre estas metodologías se destacan CRISP-DM y SEMMA, que se enfocan en los parámetros técnicos del proceso de desarrollo durante la ejecución del modelo, sin darse cuenta de los resultados, mientras que CRISP-DM es un marco de trabajo para la implementación de proyectos de minería de datos, que permite convertir problemas en cuestiones técnicas. Según Chapman et al. (2009), este modelo consta de un manual de usuario y un modelo de referencia, que proporciona un esquema general del ciclo de vida del proyecto de minería de datos. Las fases definidas en el modelo son información comercial, análisis de datos, preparación de datos, modelado, evaluación e implementación, cada una de ellas compuesta por tareas generales, desglosadas en tareas específicas y una variación del proceso de minería de datos en proyectos. La guía del usuario ofrece instrucciones y orientación paso a paso para completar cada una de las etapas del proyecto. (Cobos et al., 2010).

Ahora, por otro lado, hablaremos sobre algunos de los algoritmos que implementaremos en este estudio, para conocer de que tratan y como es su funcionamiento.

I.VII.I.VII.I. CRISP-DM

Según Moine, Gordillo y Haedo (2011) esta metodología fue creada por las empresas de SPSS, NCR y Daimler Chrysler en el 2000, es una de las metodologías más usadas en el desarrollo de proyectos de minería de datos, adicionalmente establece un conjunto de tareas y actividades para cada fase del proyecto pero no especifica cómo tratarlas. Se estructura en seis etapas:

1. Comprensión del negocio: Evaluación del problema a abordar y el contexto organizacional
2. Comprensión de los datos: Análisis de las variables y los datos, para determinar cuáles serán usadas.
3. Preparación de los datos: Limpieza, pre-procesamiento, reducción y proyección de los datos
4. Modelado: Selección de técnicas, modelos u algoritmos de minería de datos.
5. Evaluación e implementación: Interpretación, refinamiento del modelo, difusión del nuevo conocimiento dentro de la organización, planificación para el control futuro y un análisis de cierre del proyecto.

I.VII.I.VIII. WEKA

Es de libre distribución (Licencias GPL), es una herramienta destacada por la gran cantidad de algoritmos que contiene, así mismo es destacado por su eficiencia a la hora

de implementarlos, fue desarrollada por investigadores de la Universidad de Waikato, adicionalmente proporciona una gran cantidad de herramientas para la realización y el desarrollo de modelos a partir de la minería de datos, también permite la programación en JAVA de algoritmos más sofisticados para el análisis de datos, todo esto se reúne en una interfaz gráfica, la cual permite acceder fácilmente a sus funcionalidades. En ella se implementan las técnicas de clasificación, asociación, agrupamiento, y predicción existentes en la actualidad. Su sistema operativo es multiplataforma (Rodríguez Suárez & Amador, 2009).

I.VII.I.IX. JAVA

La creación del lenguaje de programación de Java se inspiró en las funcionalidades propuestas en otros lenguajes tales como C++, Objective C entre otros, el resultado obtenido es una plataforma con un lenguaje de alto nivel adecuado para el desarrollo de aplicaciones seguras, además el diseño de Java aportó una sintaxis sencilla, orientada a objetos e interpretada, que permite reducir el tiempo y el ciclo de desarrollo (Thierry Groussard).

¿Por qué java?

Java es un lenguaje de programación de propósito general, que se caracteriza por ser multiplataforma, lo que significa que puede ser ejecutado en diferentes sistemas operativos. Esto permite a los desarrolladores de software crear aplicaciones y servicios que puedan ser utilizados en una amplia variedad de dispositivos y plataformas. Gracias a su versatilidad, Java es utilizado para desarrollar proyectos de diversos tipos, desde aplicaciones web hasta servicios web que utilizan protocolos ligeros para el intercambio de información en entornos descentralizados y distribuidos; es un lenguaje de código abierto, por lo tanto, se pueden aprovechar los códigos de fuente para personalizar nuestros proyectos, también, ofrece herramientas que permiten realizar proyectos que se adapten a las necesidades y su eficiencia genera características de seguridad siendo una opción ideal para proyectos comerciales (Java, s. f.)

Tabla 3. Java vs Python vs go vs php vs c#

| Variable | Java | Python | GO | PHP | C# |
|---------------|--|---|---|--|---|
| Portabilidad. | Las aplicaciones Java pueden ejecutarse en cualquier dispositivo móvil que tenga | Para ejecutar programas escritos en Python, es necesario contar con un intérprete de Python instalado | Go es compatible con sistemas Windows, Mac OS X, Linux y FreeBSD. | PHP es compatible con diversos sistemas operativos y puede ser utilizado en diferentes tipos | Puede ser utilizado en otros IDEs como Mono o Xamarin. En resumen, C# es un lenguaje de |

| | | | | | |
|--------------------------------------|--|--|---|---|---|
| | una máquina virtual Java instalada y funcional. | en la máquina de destino que pueda leer y ejecutar el código escrito en Python. | | de servidores web, lo que lo convierte en un lenguaje de programación altamente versátil para el desarrollo de aplicaciones web y servicios en línea. | programación altamente portátil y puede ser utilizado en diferentes sistemas operativos. |
| Librerías de aprendizaje automático. | Weka, Mallet, Deeplearning4j. | Tensorflow, Pytorch. | Golearn, gorgonia. | RubixML. | ml.net |
| Mejor uso para. | Es la mejor opción para aplicaciones de escritorio, sistemas integrados, servicio de aplicaciones web y exploración de big data porque es rápido, confiable y robusto. | Es adecuado para procesamiento de datos científicos y numéricos, aplicaciones de aprendizaje automático. | es un lenguaje de programación que sirve para muchas cosas, por ejemplo, apps, compilación de plataforma cruzada, arquitectura IT | Se utiliza para crecer desde pequeños sitios web hasta grandes portales comerciales y para crear muchos tipos de aplicaciones. Se utiliza en el mundo del desarrollo web porque nos permite hacer muchas cosas. | El punto fuerte de desarrollo de c# está centrado en el desarrollo móvil y videojuegos. |
| Rendimiento. | Es más eficiente gracias a su optimización. | En Python puedes añadir implementaciones, pero estas pueden afectar negativamente a la portabilidad dentro del código. | Go compila su código directamente en el código de la máquina sin máquinas virtuales para ralentizar el proceso de compilación. | PHP proporciona una gran cantidad de marcos, es rápido y muy eficiente con resultados rentables para sitios web dinámicos. | C# es una invención de Microsoft que combina la funcionalidad básica de C++ proporcionando un marco. El problema es que es .Net y ya no es nativo/portátil. |

| | | | | | |
|----------|---|---|---------------------------|---|---|
| Ventajas | Java es directo, fuertemente tipado y tiene expectativas estrictas que rigen a los alumnos para que lo contemplen de la manera correcta, además es fácil de usar, escribir, compilar, depurar y aprender, | Python puede ejecutar una función complicada con solo unas pocas líneas de código | Sintaxis muy simplificada | permite pruebas unitarias eficientes con PHPUnit. Por otro lado, debido al desarrollo generalizado de este lenguaje, puede reducir significativamente los costos de desarrollo. | su fuerza como lenguaje, pero también su versatilidad. Admite la mayoría de los modelos, destacando el modelo funcional combinado con el modelo orientado a objetos |
|----------|---|---|---------------------------|---|---|

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

I.VII.I.X. PostgreSQL

PostgreSQL es un motor potente de bases de datos objeto-relacional, además usa multiprocesos lo cual garantiza estabilidad en el sistema esto quiere decir que si hay un fallo en alguno de los procesos no afectara el resto y el sistema seguirá funcionando. Otro punto para destacar es su accesibilidad a los sistemas operativos como Windows, Linux y Unix, también se puede definir PostgreSQL como un sistema que permite la manipulación de la información la cual se almacena en tablas y se rige por las reglas del algebra relacional (Mariuxi Ordóñez, Jimmy Ríos).

¿Por qué PostgreSQL?

Aparte de ser la aplicación de bases de datos más avanzada de código abierto, se puede emplear por su funcionalidad y por su capacidad, una de las ventajas fundamentales es que tiene gran capacidad de soportar enorme cantidad de datos.

Se caracteriza por la capacidad de manejar bastante cantidad de clientes sin bloqueos, permite trabajar con los datos como si fueran objetos, es decir, se pueden heredar tablas, tiene compatibilidad con el lenguaje de programación Java, y por último, tiene licencia abierta para cualquier uso sin costo alguno, todo esto lo hace una alternativa ideal para este proyecto (¿Qué es PostgreSQL y por qué llevarlo a Cloud?, s. f.).

Tabla 4. PostgreSQL vs Oracle vs MySQL

| Variable | PostgreSQL | Oracle | MySQL |
|----------|------------|--------|-------|
|----------|------------|--------|-------|

| | | | |
|--------------------------|---|--|--|
| Estructura | Es un servidor de bases de datos relacional libre, liberado bajo la licencia BSD lo que significa que cualquiera puede acceder a su código fuente y modificarlo a su voluntad, está diseñada como una base de datos orientada a objetos, es decir las tablas son objetos y las tuplas son instancias de esos objetos. | utiliza un modelo de base de datos relacional que permite almacenar y representar los datos de la empresa y los clientes en forma de conjuntos de datos organizados. Las cantidades de datos se estructuran en columnas, tablas y filas, y los puntos de datos se relacionan con la ayuda de atributos. La gran ventaja de la base de datos de Oracle es que organiza y presenta volúmenes de datos de manera intuitiva y eficiente. | Tiene una extensiva reutilización del código dentro del software y un acercamiento minimalista para producir características funcionales, dando lugar a un sistema de administración de las bases de datos con muy buena velocidad, capacitación estabilidad y facilidad de despliegue |
| Rendimiento | Es un gestor de bases de datos muy ágil que para el caso de proyectos grandes y complejos con operaciones de datos de alto volumen, que requieren alta velocidad, fiabilidad y escalabilidad tanto en lectura como escrituras de datos cumple estas necesidades si problema. | La base de datos de Oracle posee escalabilidad, protección y alto rendimiento para la actividad empresarial. | es un motor muy ágil rápido en el caso de proyectos de páginas o aplicaciones web escalables, que hacen uso de una base de datos únicamente para transacciones de datos sencillas, y en los cuales la velocidad en lectura de datos es lo más importante. |
| Costo Total de Propiedad | Es de código abierto, no hay precio para adquisición y soporte de productos ya que son absolutamente gratuitos. | El precio para adquirir y tener el soporte de productos para Oracle Database es elevado y cuando haya la necesidad de mejores características se deberá pagar un valor adicional | MySQL es de uso libre y gratuito posee Software con Licencia GPL, además tiene un Bajo costo en requerimientos para la elaboración y ejecución del programa, no se necesita disponer |

| | | | |
|------------------------------|--|---|--|
| | | | de Hardware o Software de alto rendimiento para la ejecución del programa. |
| Soporte | ofrece atención al cliente de forma gratuita, pero lleva tiempo resolver un problema, ya que lo lleva a cabo la propia comunidad de desarrolladores. La comunidad de desarrolladores ofrece activamente soporte en línea gratuito. | El soporte al cliente para Oracle se puede utilizar a un costo. El costo es casi una cuarta parte del costo de la licencia | Posee soporte aproximadamente en el 95% de los sistemas operativos actuales, además, la comunidad mediante foros se apoya a solucionar problemas comunes. |
| replicación y compatibilidad | garantiza una alta disponibilidad a través de la replicación de transmisión. Es compatible con el modelo Maestro-Eslavo para la replicación que proporciona un rendimiento impecable durante la asignación de tareas, la copia de seguridad y la agrupación en clústeres. | Oracle garantiza una alta disponibilidad a través de DataGuard. es compatible con los modelos Maestro-Eslavo y Maestro-Maestro. | MySQL soporta replicación unidireccional asíncrona, es decir, las consultas de actualización ejecutadas en el maestro son replicadas en los servidores esclavos. |
| Seguridad | Permite a los usuarios establecer permisos al ofrecer roles heredados y controles de acceso adicionales a través de SE-PostgreSQL. También es compatible con SSL nativo que ayuda a cifrar las comunicaciones del | Ofrece características de seguridad más robustas en comparación con PostgreSQL. Proporciona excelentes soluciones de aislamiento entre la gestión de cifrado de claves independiente y las bases de datos conectables | Por el lado de la seguridad, se trata con un sistema de privilegios y contraseñas siendo flexible y seguro, además, permite verificación basada en el host. Las contraseñas están encriptadas al |

| | | | |
|--|-----------|--|-------------------------|
| | servidor. | | conectarse al servidor. |
|--|-----------|--|-------------------------|

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

I.VII.II. Marco legal

Este proyecto se desarrollará en términos de software libre, es decir que no requiere de licencias privadas, por tal motivo se le permite a los usuarios realizar estudios y modificaciones las cuales permitan mejorar sus capacidades y adaptarlo a necesidades particulares. De acuerdo a Rodríguez (1992) el autor otorga las siguientes libertades a los usuarios:

- La libertad de ejecutar el programa para cualquier propósito.
- La libertad de estudiar la manera en que el programa opera y adaptarlo a sus necesidades particulares.
- La libertad para redistribuir copias del programa (incluido su código fuente) a quien desee.
- La libertad de mejorar el programa y distribuir sus mejoras al público bajo las mismas condiciones del programa original.

A continuación, se describen aspectos legales y normativos que influyen en el proyecto en cuanto a minería de datos, derechos de autor y protección de obras.

La constitución política de Colombia en el artículo 61, establece que “El estado protegerá la propiedad intelectual por el tiempo y mediante las formalidades que establezca la ley”.

I.VII.II.I. Ley 23 de 1982

Artículo 1:

Autores de obras literarias, científicas y artísticas gozarán de protección para sus obras en la forma prescrita por la presente ley y, en cuanto fuere compatible con ella, por el derecho común. También protege esta ley a los intérpretes o ejecutantes, a los productores de fonogramas y a los organismos de radiodifusión, en sus derechos conexos a los del autor.

Artículo 2:

Los derechos de autor recaen sobre las obras científicas, literarias y artísticas las cuales se comprenden todas las creaciones del espíritu en el campo científico, literario y artístico, cualquiera que sea el modo o forma de expresión y cualquiera que sea su destinación, tales como: los libros, folletos y otros escritos; las conferencias, alocuciones,

sermones y otras obras de la misma naturaleza; las obras dramáticas o dramático musicales; las obras coreográficas y las pantomimas; las composiciones musicales con letra o sin ella; las obras cinematográficas, a las cuales se asimilan las obras expresadas por procedimiento análogo a la cinematografía, inclusive los video-gramas; las obras de dibujo, pintura, arquitectura, escultura, grabado, litografía; las obras fotográficas a las cuales se asimilan las expresadas por procedimiento análogo a la fotografía; las obras de arte aplicadas; las ilustraciones, mapas, planos, croquis y obras plásticas relativas a la geografía, a la topografía, a la arquitectura o a las ciencias, y, en fin, toda producción del dominio científico, literario o artístico que pueda reproducirse, o definirse por cualquier forma de impresión o de reproducción, por fonografía, radiotelefonía o cualquier otro medio conocido o por conocer.

Artículo 30:

El autor tendrá sobre su obra un derecho perpetuo, inalienable, e irrenunciable para:

Reivindicar en todo tiempo la paternidad de su obra y, en especial, para que se indique su nombre o seudónimo cuando se realice cualquiera de los actos mencionados en el artículo 12 de esta ley.

A oponerse a toda deformación, mutilación u otra modificación de la obra, cuando tales actos puedan causar o acusen perjuicio a su honor o a su reputación, o la obra se demerite, y a pedir reparación por éstos.

A conservar su obra inédita o anónima hasta su fallecimiento, o después de él cuando así lo ordenase por disposición testamentaria.

A modificarla, antes o después de su publicación, y

A retirarla de la circulación o suspender cualquier forma de utilización aunque ella hubiese sido previamente autorizada

I.VII.II.II. El Decreto 1360 de 1989

“Por el cual se reglamenta la inscripción del soporte lógico (software en el Registro Nacional del Derecho de Autor)”, incorporó el concepto del software en la normativa colombiana. Este decreto define al software “como una creación propia del dominio literario” comprendiendo el programa de computador (u ordenador), su descripción y el material auxiliar. Sin embargo, el artículo 7 señala que “la protección otorgada al soporte lógico, no excluye otras formas de protección”, es decir que otros componentes del software están protegidos.

I.VII.II.III. Decisión 351 de 1993 Acuerdo de Cartagena

De acuerdo con el Régimen Común sobre Derechos de Autor y Derechos Conexos. En su capítulo 1 indica que se protege el programa de ordenador (software) la cual define como “Expresión de un conjunto de instrucciones mediante palabras, códigos, planes o en cualquier otra forma que, al ser incorporada en un dispositivo de lectura automatizada, es capaz de hacer que un ordenador (un aparato electrónico o similar capaz de elaborar informaciones) ejecute determinada tarea u obtenga determinado resultado. El programa de ordenador comprende también la documentación técnica y los manuales de uso”.

I.VII.II.IV.Ley 565 de 2000

Por medio de la cual se aprueba el Tratado de la OMPI sobre Derechos de Autor, adoptado en Ginebra el 20 de diciembre de 1996. Artículo 4: “Los programas de ordenador están protegidos como obras literarias en el marco de lo dispuesto en el artículo 2 del Convenio de Berna. Dicha protección se aplica a los programas de ordenador, cualquiera que sea su modo o forma de expresión.”

I.VII.II.V. Ley 1712 de 2014

Se podrá hacer uso, aprovechamiento y/o transformación de forma libre y sin restricciones, para hacer aplicaciones por parte de terceros y contenidos de su propia creación

II. Limpieza De Datos

La limpieza de los resultados de las pruebas es un paso esencial en cualquier proceso de análisis de datos. Cuando se trabaja con conjuntos de datos reales, es común encontrar datos incompletos, ruidosos o mal formateados que pueden afectar negativamente la calidad y precisión de los análisis posteriores. La limpieza de datos implica la identificación y corrección de estos problemas para garantizar que los resultados sean confiables y significativos.

En el contexto de este proyecto, la limpieza de los resultados de las Pruebas Saber desempeña un papel fundamental en la preparación de los datos para su análisis posterior. Los datos pueden contener errores tipográficos, valores faltantes o inconsistencias que deben abordarse antes de que se puedan extraer conclusiones válidas.

II.I. Script de limpieza

Figura 3. Función de limpieza de datos

```
function limpiarTexto(texto) {  
  // Reemplazar "-" por punto y coma (;)  
  texto = texto.replace(/~/g, ';');  
  
  // Eliminar caracteres no ASCII  
  texto = texto.normalize('NFD').replace(/[\u0300-\u036f]/g, '');  
  
  // Reemplazar ;; por ";?" y ",," por ",?"; y eliminar comillas dobles  
  texto = texto.replace(/;;/g, ';?').replace(/"/g, '');  
  
  // Eliminar caracteres ``  
  texto = texto.replace(/['`]/g, '');  
  
  return texto;  
}
```

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Esta función toma un texto como entrada y lo limpia aplicando una serie de transformaciones:

- Reemplaza "-" por punto y coma (;).
- Elimina caracteres no ASCII.
- Reemplaza ";;" por ";?" y ",," por ",?";.
- Elimina comillas dobles.
- Elimina caracteres "`". Finalmente, devuelve el texto limpio.

II.II. Selección del archivo

El usuario puede hacer clic en el campo de selección de archivo para cargar un archivo de texto que contiene los datos que se desean limpiar. El atributo accept en el elemento input limita la selección de archivos a archivos de texto (.txt).

II.III. Procesamiento de datos

Cuando el usuario hace clic en el botón "Limpiar Datos", se inicia la función limpieza(numVeces). Esta función realiza varias tareas:

- Comprueba si se ha seleccionado un archivo de texto. Si no, muestra una alerta.
- Muestra el indicador visual de carga.
- Utiliza el objeto FileReader para leer el contenido del archivo seleccionado.
- Aplica la función limpiarTexto(texto) al contenido del archivo para realizar la limpieza de datos. Esta función lleva a cabo acciones como reemplazar

caracteres, eliminar caracteres no ASCII y realizar otras transformaciones en el texto.

- El bucle for permite repetir el proceso de limpieza varias veces (según el valor de numVeces).
- Finalmente, convierte el contenido limpio en un archivo CSV y crea un enlace de descarga para que el usuario pueda descargar el resultado limpio.

II.IV. Descarga de los datos limpios

Cuando se completa la limpieza de datos y se genera el archivo CSV limpio, se muestra un enlace de descarga ("Descargar CSV") que permite al usuario descargar el archivo resultante.

III. Selección De Indicadores

III.I. Configuración Weka

Para el análisis de los indicadores se hizo uso del aplicativo Weka desde allí se realizó el análisis de cada una de las variables cargadas en el archivo CSV, inicialmente se realizó la siguiente configuración:

- Instalar WEKA
- Descargar el controlador de PostgreSQL, versión "42.2.9"
- Adicionar a la carpeta donde quedo instalado el WEKA
- Buscar el archivo "DatabaseUtils.props" y configurarlo de la siguiente manera

Figura 4. Configuración del archivo “DatabaseUtils.props”

```
# url: http://www.mysql.com/
# jdbc: http://www.mysql.com/products/connector/j/
# author: Fracpete (fracpete at waikato dot ac dot nz)
# version: $Revision: 15255 $

# JDBC driver (comma-separated list)
jdbcDriver=org.postgresql.Driver

# database URL
jdbcURL=jdbc:postgresql://localhost:5432/basededatos

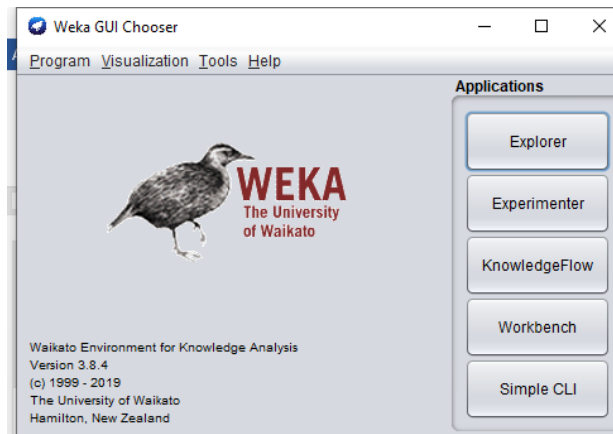
# specific data types
string, getString() = 0; --> nominal
boolean, getBoolean() = 1; --> nominal
double, getDouble() = 2; --> numeric
byte, getByte() = 3; --> numeric
short, getByte() = 4; --> numeric
int, getInteger() = 5; --> numeric
long, getLong() = 6; --> numeric
float, getFloat() = 7; --> numeric
date, getDate() = 8; --> date
text, getString() = 9; --> string
time, getTime() = 10; --> date
timestamp, getTime() = 11; --> date

# other options
```

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Una vez configurado, abrir weka y seleccionar la opción “Explorer”

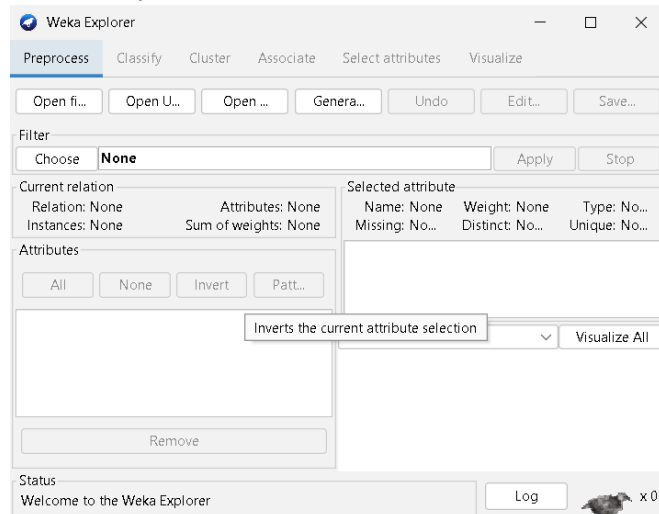
Figura 5. Interfaz inicial de weka



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Al ingresar a la interfaz de explorer nos abre la siguiente interfaz donde procedemos a cargar el archivo csv de los resultados para realizar el análisis.

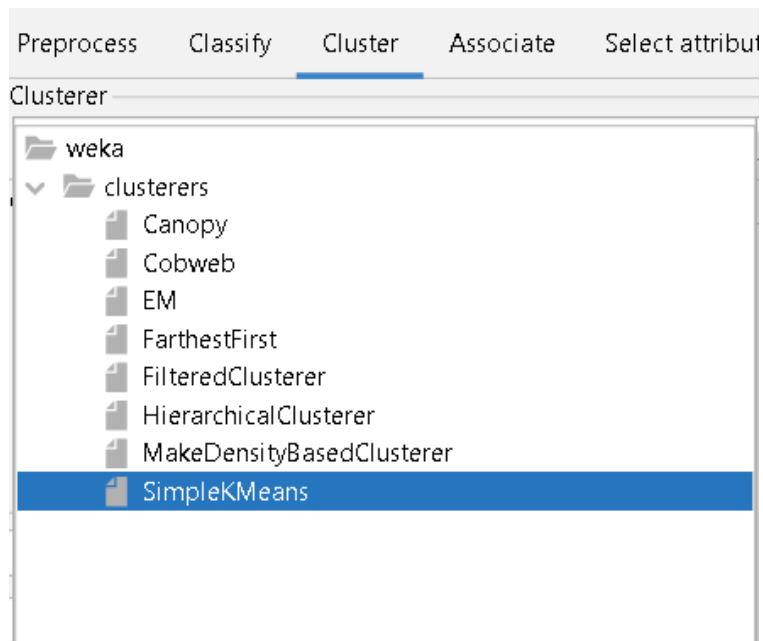
Figura 6. Interfaz explorer



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- A partir de los datos obtenidos se debe seleccionar el tipo de análisis a realizar. En este caso se seleccionó la metodología de “Cluster” y como algoritmo “SimpleKMeans”.

Figura 7. Menu clasificadores weka



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Una vez seleccionado el algoritmo, se procede a ejecutar el análisis.

Figura 8. Analisis algoritmo simpleKmeans

The screenshot shows the Weka Clusterer window with the SimpleKMeans algorithm selected. The interface includes a menu bar (Preprocess, Classify, Cluster, Associate, Select attributes, Visualize), a toolbar (Choose, Start, Stop), and a main area divided into 'Cluster mode' and 'Clusterer output'.

Cluster mode:

- Use training set (selected)
- Supplied test set: Set...
- Percentage split: % 66
- Classes to clusters evaluation: (Nom) PUNT_GLOBAL
- Store clusters for visualization (checked)

Clusterer output:

```

Instances: 15528
Attributes: 3
          FAMI_TIENEINTERNET
          FAMI_TIENECOMPUTADOR
          PUNT_GLOBAL
Test mode: evaluate on training data

==== Clustering model (full training set) ====

kMeans
=====

Number of iterations: 2
Within cluster sum of squared errors: 2514.0

Initial starting points (random):

Cluster 0: 81,81,``(258.5-376.75)``
Cluster 1: 81,81,``(140.25-258.5)``

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute          Full Data          Cluster#          1
                   (15528.0)         (11950.0)         (3578.0)
=====
FAMI_TIENEINTERNET          81          81          81
FAMI_TIENECOMPUTADOR        81          81          81
PUNT_GLOBAL          *(258.5-376.75)* *(258.5-376.75)* *(140.25-258.5)*
  
```

Result list (right-click for options):

- 16:03:39 - SimpleKMeans
- 16:07:31 - SimpleKMeans
- 16:08:38 - SimpleKMeans
- 16:09:04 - SimpleKMeans
- 16:10:52 - SimpleKMeans
- 16:16:30 - SimpleKMeans
- 16:25:04 - SimpleKMeans
- 11:15:32 - SimpleKMeans (highlighted)

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

III.II. Selección De Indicadores

III.II.I Indicador Acceso A Internet

Para el caso del indicador Acceso a internet se obtienen dos grupos identificados por el algoritmo k-means parecen estar diferenciados por los atributos PUNT_GLOBAL y FAMI_TIENEINTERNET.

Se pudo evidenciar que el primer caso presenta un 77% que se caracteriza por tener una puntuación PUNT_GLOBAL más alta al tener en el grupo de FAMI_TIENEINTERNET acceso a internet. El segundo caso, que contiene el 23% de los datos, se caracteriza por tener una puntuación PUNT_GLOBAL más baja y en el grupo FAMI_TIENEINTERNET no tienen acceso a internet.

Estos resultados sugieren que el primer caso puede representar hogares con niveles educativos altos pudiendo deberse a que estos tienen acceso a internet facilitando la entrada a recursos académicos, mientras que el segundo caso puede representar hogares con niveles educativos más bajos.

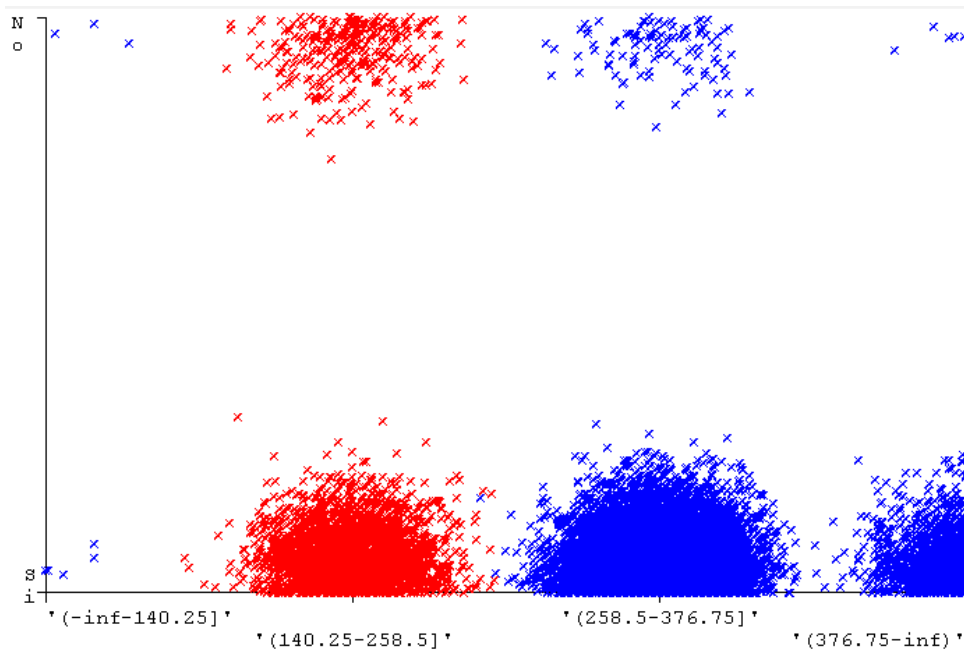
Figura 9. Ejemplo de clasificacion Acceso a internet

Final cluster centroids:

| Attribute | Full Data | | Cluster# | |
|--------------------|------------------|------------------|------------------|------------------|
| | (15528.0) | (11950.0) | 0 | 1 |
| FAMI_TIENEINTERNET | si | si | si | si |
| PUNT_GLOBAL | '(258.5-376.75]' | '(258.5-376.75]' | '(140.25-258.5]' | '(140.25-258.5]' |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Figura 10. Clusters del indicador de acceso a internet



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

III.II.II Indicador Acceso A Computador

El análisis de los resultados del algoritmo k-means aplicado a los datos indica que se han identificado dos grupos diferenciados principalmente por el atributo PUNT_GLOBAL:

- Grupo 1: Representa el 77% de los datos y se caracteriza por tener una puntuación PUNT_GLOBAL más alta.
- Grupo 2: Representa el 23% de los datos y se caracteriza por tener una puntuación PUNT_GLOBAL más baja.

La diferencia en la puntuación PUNT_GLOBAL sugiere que el Grupo 1 puede representar hogares con niveles educativos más altos con acceso a computador, mientras que el Grupo 2 puede representar hogares con niveles educativos más bajos sin computador. Esto se debe a que la puntuación PUNT_GLOBAL está relacionada con el rendimiento académico de los estudiantes, y los hogares con niveles educativos más altos suelen tener acceso a un computador dando con mejores resultados académicos.

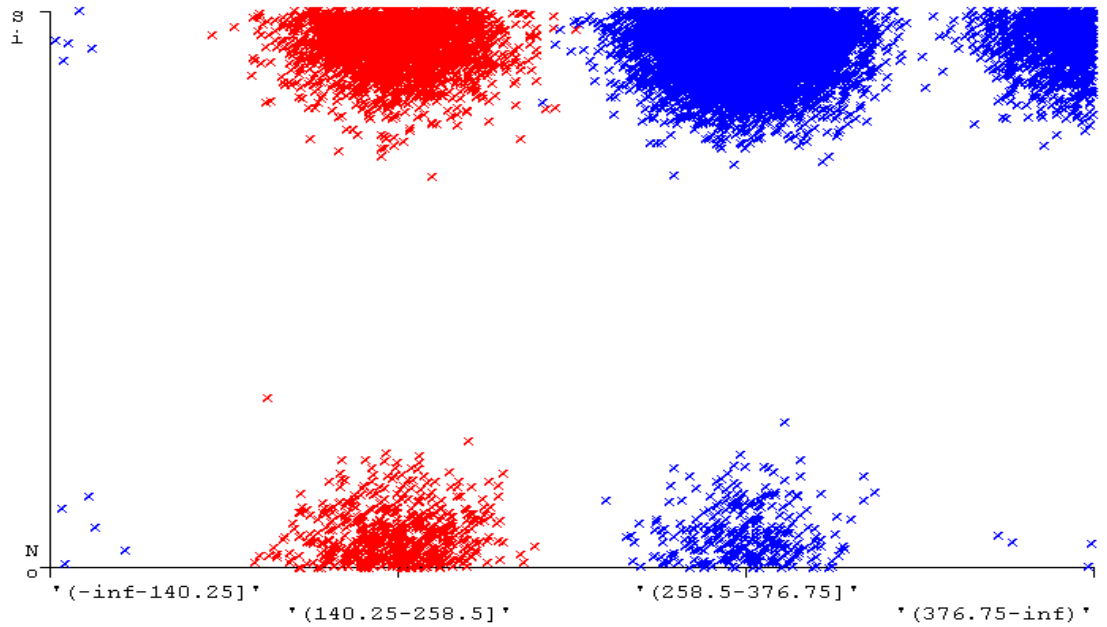
Figura 11. Ejemplo de clasificacion acceso a computador

```
Final cluster centroids:
```

| Attribute | Full Data | Cluster# | |
|----------------------|------------------|------------------|------------------|
| | (15528.0) | 0 | 1 |
| | | (11950.0) | (3578.0) |
| ===== | | | |
| FAMI_TIENECOMPUTADOR | si | si | si |
| PUNT_GLOBAL | *(258.5-376.75]* | *(258.5-376.75]* | *(140.25-258.5]* |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Figura 12. Clusters Indicador acceso a computador



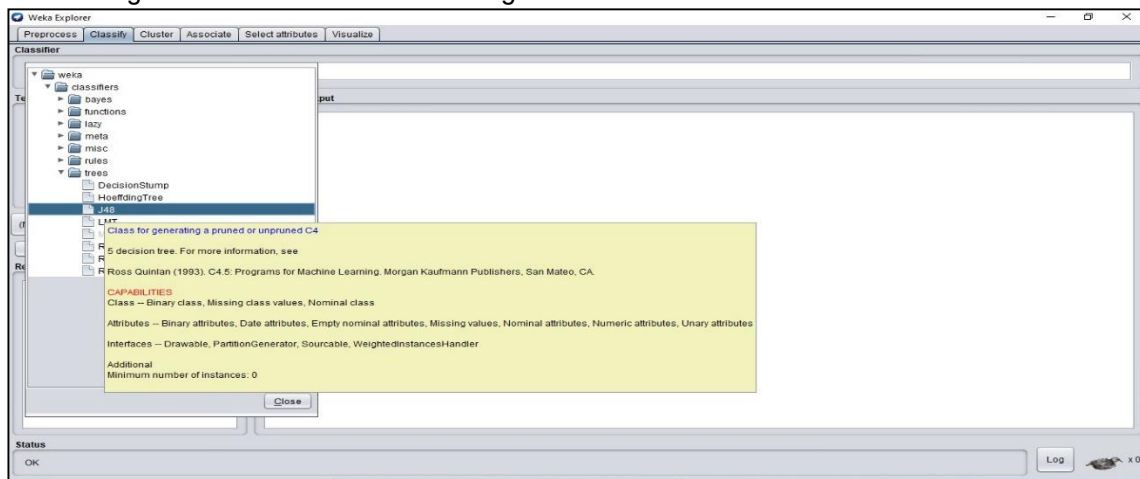
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

IV. Algoritmo De Analisis Predictivo

Para el análisis predictivo se implementó el algoritmo J48 con la ayuda de la herramienta WEKA, inicialmente se analizó el árbol generado y las matrices de confusión para determinar la validez del modelo, para este proceso se tuvieron en cuenta los siguientes pasos:

- Seleccionar de la herramienta weka la opción “Classify”, posteriormente seleccionar “trees” y “J48”

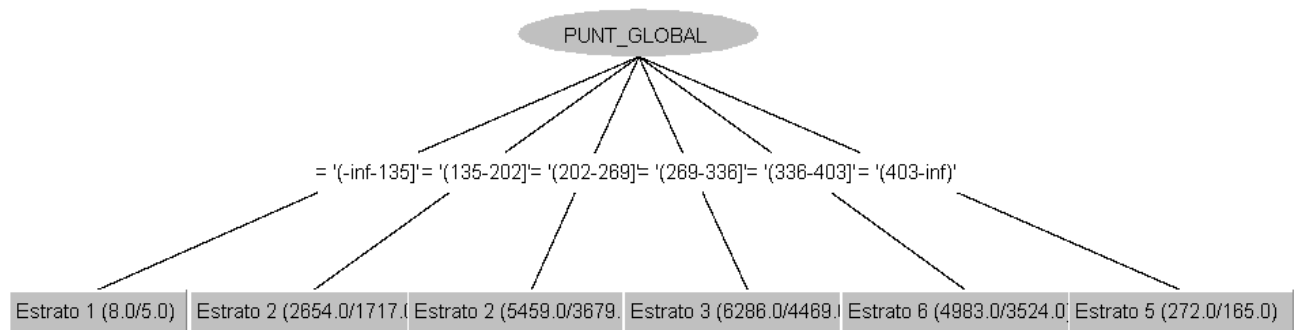
Figura 13. Selección del Algoritmo J48



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Una vez iniciado el proceso de analisis, la herramienta mostrara el arbol dedecision de la siguiente manera:

Figura 14. Arbol generado con el algoritmo J48



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Una vez generado el árbol el aplicativo procede a extraer las reglas generadas y determinar el porcentaje de clasificación para dicha regla.

IV.I. Validación algoritmo predictivo

Para determinar la validez del algoritmo, se tomó una muestra de 21066 instancias y se realizó la respectiva comparación entre el algoritmo J48 y los algoritmos de predicción RandomForest, RandomTree y REPTree, obteniendo de esta manera los siguientes resultados:

IV.I.I Prueba indicador de estrato socioeconómico

IV.I.I.I Algoritmo J48

Figura 15. Porcentaje de clasificación Algoritmo J48

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 19063 | 77.2629 % |
| Incorrectly Classified Instances | 2003 | 22.7371 % |
| Kappa statistic | 0.1995 | |
| Mean absolute error | 0.2227 | |
| Root mean squared error | 0.3325 | |
| Relative absolute error | 50.0685 % | |
| Root relative squared error | 54.5697 % | |
| Total Number of Instances | 21066 | |

Fuente: (Vergara serrato-Mahecha Hernández, 2023)

Para los resultados que el algoritmo J48 nos arroja indica que es capaz de clasificar correctamente el 77.26% de las instancias, lo que es un porcentaje relativamente alto. Además, el error absoluto medio y el error cuadrático medio son relativamente bajos, lo que indica que el modelo tiene un error promedio significativo.

- Resultados de rendimiento: El árbol J48 tiene una tasa de clasificación correcta, lo que indica que es capaz de clasificar correctamente la mayoría de las instancias de datos. Este es un resultado positivo, ya que indica que el modelo tiene un buen potencial para ser utilizado en aplicaciones de clasificación.
- Resultados de interpretabilidad: El árbol J48 es un modelo de aprendizaje automático interpretable, lo que significa que es posible comprender cómo el modelo llega a sus conclusiones. Este es un resultado positivo, ya que facilita la explicación del modelo a los usuarios finales y la identificación de posibles problemas o áreas de mejora.

IV.I.I.II. Algoritmo RandomForest

Figura 16. Porcentaje de clasificación algoritmo RandomForest

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 9003 | 42.7703 % |
| Incorrectly Classified Instances | 12063 | 57.2297 % |
| Kappa statistic | 0.1995 | |
| Mean absolute error | 0.2227 | |
| Root mean squared error | 0.3325 | |
| Relative absolute error | 90.0685 % | |
| Root relative squared error | 94.5697 % | |
| Total Number of Instances | 21066 | |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Precisión de Clasificación: El hecho de que el algoritmo haya clasificado correctamente el 42.7703% de las instancias indica que su rendimiento de clasificación no es muy alto. Esto podría deberse a varios factores, como la calidad de los datos, la elección de características, la configuración del modelo, entre otros. Una precisión del 42.7703% sugiere que el modelo no está funcionando bien en la clasificación de las instancias.

- Kappa Estadístico: El valor de Kappa estadístico de 0.1995 también es indicativo de un rendimiento de clasificación no muy fuerte. El Kappa mide la concordancia entre las clasificaciones del modelo y las clasificaciones reales, y un valor de 0.1995 sugiere una concordancia limitada.
- Errores Absolutos y Cuadráticos: El Error Absoluto Medio (MAE) de 0.2227 y el Error Cuadrático Medio (RMSE) de 0.3325 indican que las predicciones del modelo tienden a estar lejos de los valores reales. En otras palabras, las predicciones no son precisas y se cometen errores sustanciales en la estimación de las clases.

IV.I.I.III. Algoritmo RandomTree

Figura 17. Porcentaje de clasificación algoritmo randomTree

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 9003 | 31.0243 % |
| Incorrectly Classified Instances | 12063 | 68.9757 % |
| Kappa statistic | 0.1427 | |
| Mean absolute error | 0.2192 | |
| Root mean squared error | 0.3311 | |
| Relative absolute error | 93.1904 % | |
| Root relative squared error | 96.5481 % | |
| Total Number of Instances | 21066 | |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Correctly Classified Instances: El modelo clasificó correctamente 9,003 instancias. Este número representa la cantidad de casos en los que el modelo hizo predicciones precisas y acertadas.
- Incorrectly Classified Instances: El modelo clasificó incorrectamente 12,063 instancias. Esto indica que en 12,063 casos, las predicciones del modelo no coincidieron con las clases reales.
- Kappa statistic: El valor del kappa estadístico es 0.1427. El kappa es una medida que evalúa la precisión del modelo teniendo en cuenta la distribución de las clases en el conjunto de datos. Un valor de 0.1427 sugiere que el modelo tiene una precisión moderada, pero no es muy fuerte.

IV.I.I.IV. Algoritmo REPTree

Figura 18. Algoritmo de clasificación REPTree

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 6100 | 31.0243 % |
| Incorrectly Classified Instances | 13562 | 68.9757 % |
| Kappa statistic | 0.1427 | |
| Mean absolute error | 0.2192 | |
| Root mean squared error | 0.3311 | |
| Relative absolute error | 93.1904 % | |
| Root relative squared error | 96.5481 % | |
| Total Number of Instances | 19662 | |
| Ignored Class Unknown Instances | 1404 | |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- **Precisión:** La precisión es del 31,02%. Esto significa que aproximadamente el 31,02% de las instancias se clasificaron correctamente. En otras palabras, el modelo acertó en alrededor del 31,02% de las predicciones.
- **Error:** El error es del 68,98%. Esto indica que aproximadamente el 68,98% de las instancias se clasificaron incorrectamente. El alto valor de error sugiere que el modelo tiene dificultades para hacer predicciones precisas.
- **Kappa:** El valor del kappa es 0,14. El kappa es una medida que evalúa la precisión del modelo teniendo en cuenta la distribución de las clases en los datos de entrenamiento. Un valor de 0,14 sugiere que el modelo tiene una precisión moderada

IV.I.II. Prueba de indicador de acceso a internet

IV.I.II.I. Algoritmo J48

Figura 19. Porcentaje de clasificación algoritmo J48

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 17242 | 87.4873 % |
| Incorrectly Classified Instances | 2466 | 12.5127 % |
| Kappa statistic | 0 | |
| Mean absolute error | 0.2189 | |
| Root mean squared error | 0.3309 | |
| Relative absolute error | 99.9855 % | |
| Root relative squared error | 100 % | |
| Total Number of Instances | 19708 | |
| Ignored Class Unknown Instances | 1358 | |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Para los resultados que el algoritmo J48 nos arroja una tasa de clasificación correcta del 87,4873%, lo que indica que el clasificador es capaz de clasificar correctamente la mayoría de las instancias.

Lo positivo de estos resultados es que el algoritmo J48 tiene una buena precisión para clasificar los datos del conjunto de datos analizado. La tasa de clasificación correcta de 87,4873% es muy alta, lo que indica que el clasificador es capaz de clasificar correctamente la mayoría de las instancias. Esto es un buen indicador de que el clasificador es capaz de generalizar bien a nuevos datos. Además, el número de instancias clasificadas incorrectamente es relativamente pequeño.

IV.I.II.II. Algoritmo RandomForest

Figura 20. Porcentaje de clasificación algoritmo RandomForest

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 7426 | 35.2511 % |
| Incorrectly Classified Instances | 13640 | 64.7489 % |
| Kappa statistic | 0.0625 | |
| Mean absolute error | 0.2395 | |
| Root mean squared error | 0.3457 | |
| Relative absolute error | 96.8629 % | |
| Root relative squared error | 98.32 % | |
| Total Number of Instances | 21066 | |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- **Precisión:** La precisión es una medida del porcentaje de instancias que fueron clasificadas correctamente. En este caso, la precisión es del 35,25%, lo que significa que el modelo solo acertó en el 35,25% de las instancias.
- **Error:** El error es una medida del porcentaje de instancias que fueron clasificadas incorrectamente. En este caso, el error es del 64,74%, lo que significa que el modelo se equivocó en el 64,74% de las instancias.
- **Kappa:** La kappa es una medida más robusta de la precisión que tiene en cuenta la aleatoriedad. En este caso, la kappa es de 0,0625, lo que significa que el modelo no es significativamente mejor que un modelo aleatorio.

IV.I.II.III. Algoritmo RandomTree

Figura 21. Porcentaje de clasificación del algoritmo RandomTree

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 7426 | 35.2511 % |
| Incorrectly Classified Instances | 13640 | 64.7489 % |
| Kappa statistic | 0.0625 | |
| Mean absolute error | 0.2395 | |
| Root mean squared error | 0.3457 | |
| Relative absolute error | 96.8645 % | |
| Root relative squared error | 98.3203 % | |
| Total Number of Instances | 21066 | |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Los resultados dados por el algoritmo RandomTree en Weka muestran que el modelo tiene una precisión de 35.2511%. Esto significa que el modelo clasifica correctamente el 35.2511% de las instancias del conjunto de datos. El resto de las instancias, el 64.7489%, son clasificadas incorrectamente.
- El índice Kappa es un indicador de la precisión del modelo, ajustado por el azar. Un índice Kappa de 0 significa que el modelo no es mejor que una predicción aleatoria, mientras que un índice Kappa de 1 significa que el modelo es perfecto. En este caso, el índice Kappa es de 0.0625, lo que indica que el modelo es ligeramente mejor que una predicción aleatoria.

IV.I.II.IV. Algoritmo REPTree

Figura 22. Porcentaje de clasificación del algoritmo REPTree

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 6100 | 31.0243 % |
| Incorrectly Classified Instances | 13562 | 68.9757 % |
| Kappa statistic | 0.1427 | |
| Mean absolute error | 0.2192 | |
| Root mean squared error | 0.3311 | |
| Relative absolute error | 93.1904 % | |
| Root relative squared error | 96.5481 % | |
| Total Number of Instances | 19662 | |
| Ignored Class Unknown Instances | 1404 | |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

- Precisión: La precisión es del 31,02%. Esto significa que aproximadamente el 31,02% de las instancias se clasificaron correctamente. En otras palabras, el modelo acertó en alrededor del 31,02% de las predicciones.
- Error: El error es del 68,98%. Esto indica que aproximadamente el 68,98% de las instancias se clasificaron incorrectamente. El alto valor de error sugiere que el modelo tiene dificultades para hacer predicciones precisas.
- Kappa: El valor del kappa es 0,14. El kappa es una medida que evalúa la precisión del modelo teniendo en cuenta la distribución de las clases en los datos de entrenamiento. Un valor de 0,14 sugiere que el modelo tiene una precisión moderada

Analizando todas las pruebas realizadas se puede concluir que el algoritmo J48 presenta mayor precisión sobre los datos y adicionalmente, genera un árbol con resultados eficientes, lo cual corrobora el gran uso que tiene dentro de la minería de datos especialmente en el campo de la educación.

V. Documentación Del Software

V.I. Determinación de requerimientos

Tabla 5. Requerimientos de software

| | |
|----------------------------------|---|
| Identificación del Requerimiento | RF01 |
| Nombre del requerimiento | Login |
| Descripción del requerimiento | El usuario podrá acceder a la plataforma a través de un Login, se le pedirá usuario y contraseña para el Ingreso, además tendrá la opción del cambiode contraseña |
| Identificación del Requerimiento | RF02 |
| Nombre del requerimiento | Regístrese |
| Descripción del requerimiento | El cliente podrá diligenciar un formulario con el fin de solicitar su registro en el aplicativo, además de enviar cualquier solicitud de comunicación con los desarrolladores del software. |
| Identificación del Requerimiento | RF03 |
| Nombre del requerimiento | Creación usuarios |

| | |
|----------------------------------|---|
| Descripción del requerimiento | El administrador podrá diligenciar un formulario con la información pertinente para realizar el registro de nuevos usuarios al aplicativo web. La información se le enviará al correo (usuario y contraseña). |
| Identificación del Requerimiento | RF04 |
| Nombre del requerimiento | Consulta usuarios |
| Descripción del requerimiento | El administrador podrá visualizar el listado de usuarios existentes en el sistema. |
| Identificación del Requerimiento | RF05 |
| Nombre del requerimiento | Modificación usuarios |
| Descripción del requerimiento | El administrador podrá actualizar y/o modificar la información de los usuarios, para ello deberá consultar y seleccionar el usuario que desea modificar. |
| Identificación del Requerimiento | RF06 |
| Nombre del requerimiento | Eliminación usuarios |
| Descripción del requerimiento | El administrador podrá eliminar los usuarios, consultando previamente y seleccionando el usuario que desea eliminar. Antes de finalizar la operación saldrá una alerta de confirmación. |
| Identificación del Requerimiento | RF07 |
| Nombre del requerimiento | Carga de reportes |
| Descripción del requerimiento | El administrador tendrá que llenar un formulario, donde se cargara, tipo de análisis, región, año y subir el reporte correspondiente en un PDF. |
| Identificación del Requerimiento | RF08 |
| Nombre del requerimiento | Limpieza de los datos |

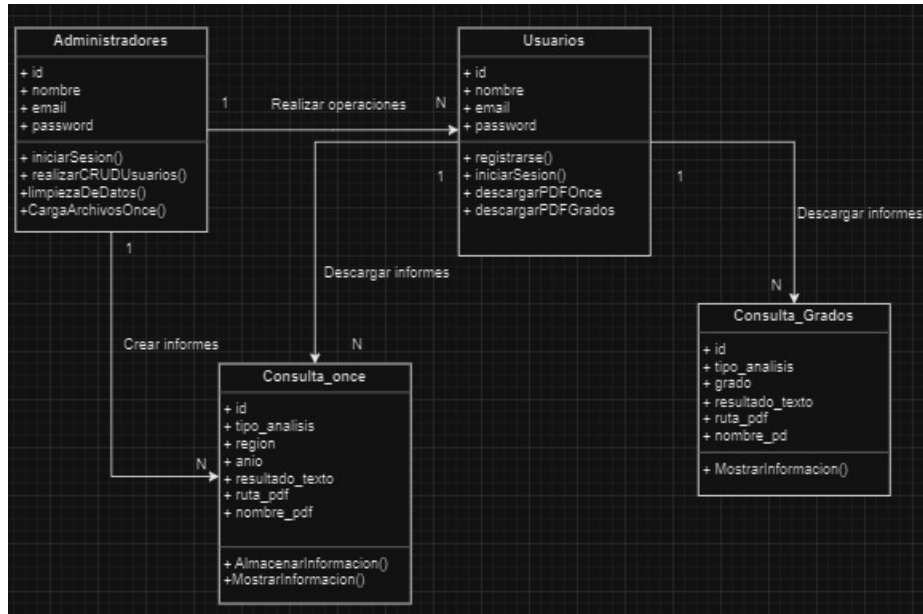
| | |
|----------------------------------|---|
| Descripción del requerimiento | El aplicativo se encargará de la limpieza de los datos, que nos proporciona el ICFES, para poder realizar los diferentes análisis en WEKA. |
| Identificación del Requerimiento | RF09 |
| Nombre del requerimiento | Visualización de los resultados |
| Descripción del requerimiento | El usuario podrá buscar en los 2 menús que reporte desea ver. |
| Identificación del Requerimiento | RF10 |
| Nombre del requerimiento | Carga de los datos |
| Descripción del requerimiento | En esta actividad el aplicativo cargará los datos transformados a la base de datos para su posterior análisis. |
| Identificación del Requerimiento | RF11 |
| Nombre del requerimiento | Reportes |
| Descripción del requerimiento | El aplicativo se encargara de proporcionar al usuario un reporte dependiendo de lo que el usuario quiera, ya sea el grado, tipo de análisis, el año y la región |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II. Especificación del diseño

V.II.I. Diagrama de clases de la base de datos

Figura 23. Modelo del diagrama de clases de la base de datos AWID



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 6. Diagrama de clases de la Base de datos resultados básico

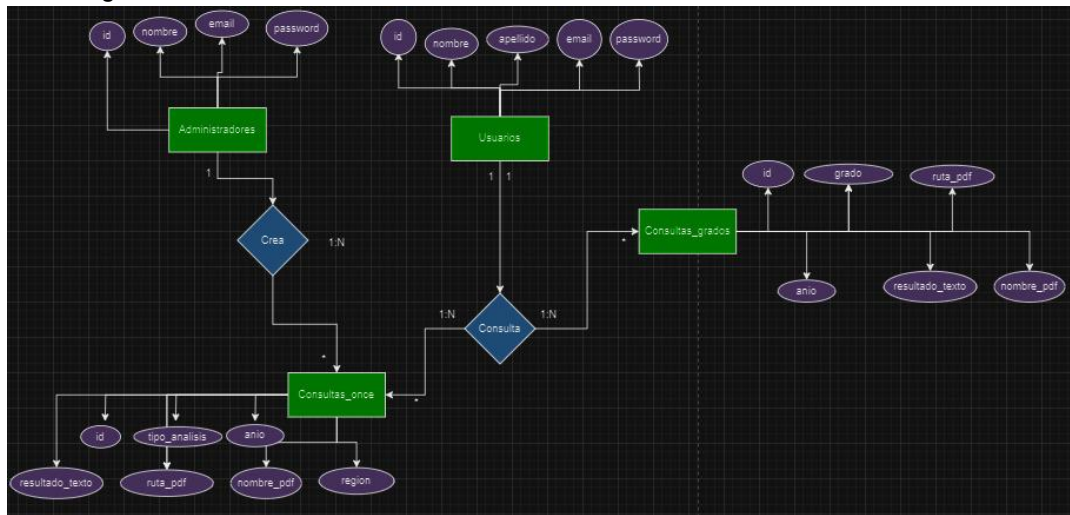
| Tabla | Descripción |
|---------------|--|
| Administrador | Representa a los administradores del sistema. Pueden iniciar sesión y tienen el poder de crear consultas. Usuario: Los usuarios registrados en el sistema. Pueden acceder a consultas_once y consultas_grados (tercero, quinto y noveno). |
| Usuarios | La tabla usuario almacena la información básica de las personas interesadas en acceder al aplicativo. Pueden realizar operaciones como descargar informes. |
| Consulta Once | Contiene información sobre las consultas relacionadas con los resultados de informes ICFES. Cada consulta_once tiene un tipo de análisis, año, región, resultado de texto, ruta de PDF y nombre de PDF. |

| | |
|-----------------|--|
| Consulta grados | Almacena los detalles de las consultas relacionadas con resultados de informes ICFES para tercero, quinto y noveno grado. Cada consulta_grados tiene un grado, año, resultado de texto, ruta de PDF y nombre de PDF. |
|-----------------|--|

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.I.I. Modelo entidad relación de la base de datos

Figura 24. MER base de datos



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

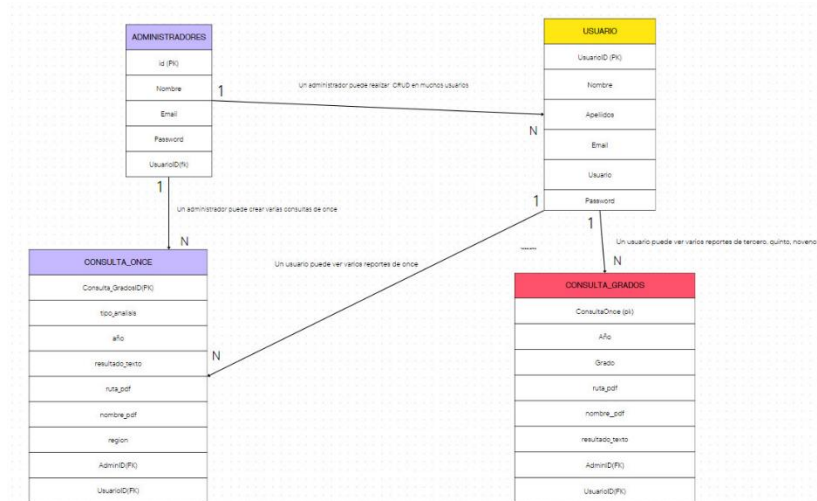
Tabla 7. Descripción relaciones

| Tabla | Descripción |
|--------------------------------|---|
| Administrador - Usuario: | Un administrador puede realizar operaciones en usuarios. (1 a muchos) |
| Administrador - Consulta_Once: | Un administrador puede crear y cargar archivos de consulta_once. (1 a muchos) |
| Usuario - Consulta_Once: | Un usuario puede descargar archivos PDF de consulta_once. (1 a muchos) |
| Usuario - Consulta_Grados: | Un usuario puede descargar archivos PDF de consulta_grados. (1 a muchos) |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.I.II Modelo relacional de la base de datos

Figura 25. Modelo relacional base de datos awid



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 8. Relaciones base de datos

| | |
|-----------------------|---|
| Tabla administrador | AdminID (PK) |
| Tabla usuario | UsuarioID (PK) |
| Tabla consulta once | ConsultaOnceID (Pk) AdminID_FK (FK) Usuario_fk (FK) |
| Tabla consulta grados | ConsultaGradosID (PK) UsuarioID_FK (FK) |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.II.Roles

Los roles son el papel que cumple cada persona a partir de sus funciones sobre el aplicativo, para así llevar a cabo un orden y ciertas responsabilidades.

V.II.II.I. Administrador

Será la persona encargada de la creación, consulta, modificación y eliminación de usuarios en el aplicativo AWID y de la actualización de los datos para la descarga de información de las pruebas, suministrada por el ICFES.

V.II.II.II. Usuario

Son las personas que accederán al aplicativo web por medio del inicio de sesión y podrán visualizar los resultados hallados y descargarlos en forma de reporte.

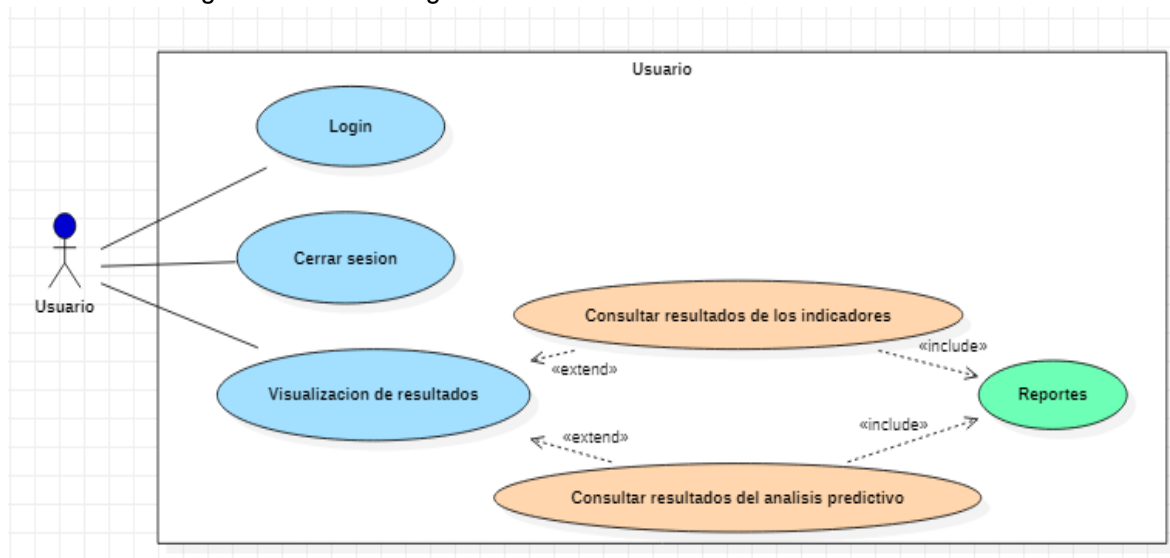
V.II.II.III. Aplicativo

Este rol se encarga de generar resultados y reportes (gráficos y estadísticas) para los usuarios, adicionalmente se encarga de la limpieza de los datos.

V.II.III. Diagramas de casos de uso

V.I.III.I. Casos de uso: Usuario

Figura 26. Diagrama casos de uso usuario



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 9. Descripción de caso de uso inicio de sesión

| CASO DE USO | |
|-------------|--|
| Nombre | Inicio de sesión |
| Actores | Usuario |
| Función | Acceder al módulo de resultados |
| Descripción | El aplicativo validara el acceso al módulo de resultados, con el usuario y la contraseña dadas por el administrador. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 10. Descripción de caso de uso reportes

| CASO DE USO | |
|-------------|---|
| Nombre | Reportes |
| Actor | Usuario |
| Función | Obtener reportes por cada resultado encontrado |
| Descripción | El usuario podrá visualizar y descargar reportes de los resultados obtenidos, allí obtendrá información mas detallada y graficas sobre los análisis hechos. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

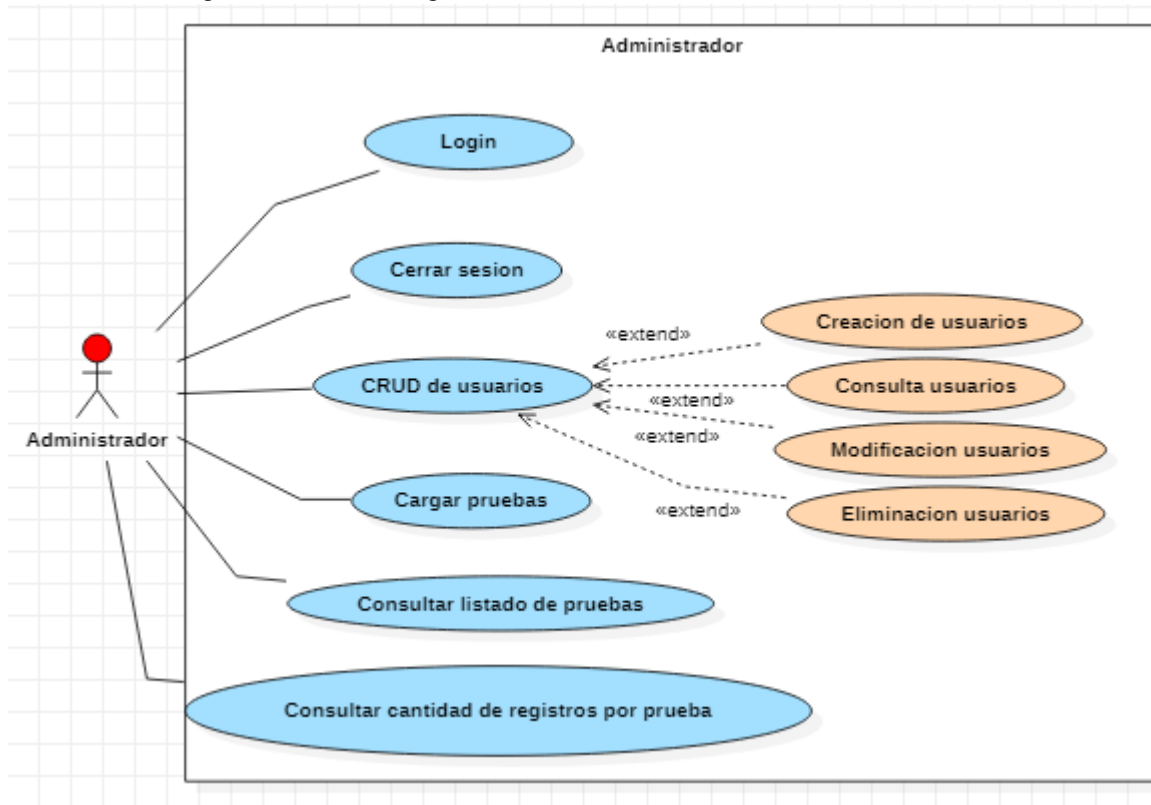
Tabla 11. Descripción de caso de uso cerrar sesión

| CASO DE USO | |
|-------------|---|
| Nombre | Cerrar Sesión |
| Actores | Usuario |
| Función | Cerrar módulo de los resultados |
| Descripción | El usuario podrá cerrar sesión del módulo de análisis descriptivo y predictivo. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.III.II. Casos de uso: Administrador

Figura 27. Diagrama caso de uso administrador



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 12. Descripción de caso de uso inicio de sesión

| CASO DE USO | |
|-------------|--|
| Nombre | Inicio de sesión |
| Actores | Administrador |
| Función | Acceder al módulo administrativo del aplicativo |
| Descripción | El aplicativo validara el acceso al módulo administrativo, con el usuario y la contraseña. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 13. Descripción de caso de uso cerrar sesión

| CASO DE USO | |
|-------------|--|
| Nombre | Cerrar Sesión |
| Actores | Administrador |
| Función | Cerrar módulo administrativo del aplicativo |
| Descripción | Se le permite al administrador cerrar sesión del módulo administrativo del aplicativo. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 14. Descripción de caso de uso creación de usuarios

| CASO DE USO | |
|-------------|--|
| Nombre | Creación usuarios |
| Actores | Administrador |
| Función | Crear usuarios |
| Descripción | El administrador podrá crear usuarios una vez le sea notificado por correo la solicitud de registro. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 15. Descripción de caso de uso consulta de usuarios

| CASO DE USO | |
|-------------|--|
| Nombre | Consulta usuarios |
| Actores | Administrador |
| Función | Obtener un listado de usuarios. |
| Descripción | El administrador podrá visualizar un listado de todos los usuarios registrados en el aplicativo. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 16. Descripción de caso de uso modificación de usuarios

| CASO DE USO | |
|-------------|------------------------------|
| Nombre | Modificación usuarios |
| Actores | Administrador |
| Función | Modificar datos del usuario. |

| | |
|-------------|--|
| Descripción | El administrador podrá modificar o actualizar datos de los usuarios registrados. |
|-------------|--|

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 17. Descripción de caso de uso eliminación usuarios

| CASO DE USO | |
|-------------|---|
| Nombre | Eliminación usuarios |
| Actores | Administrador |
| Función | Eliminar los datos registrados del usuario |
| Descripción | El administrador podrá eliminar toda la información relacionada con el usuario. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 18. Descripción de caso de uso cargar pruebas

| CASO DE USO | |
|-------------|---|
| Nombre | Cargar pruebas |
| Actores | Administrador |
| Función | Cargar pruebas al aplicativo |
| Descripción | El administrador se encargara de realizar la carga de pruebas al aplicativo, esto dependerá de los años y las pruebas que no han sido cargadas. Para la carga se debe subir un archivo “.pdf”, una vez subido el aplicativo se encargara de validar y guardar el archivo. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 19. Descripción de caso de uso consultar listado de pruebas

| CASO DE USO | |
|-------------|---|
| Nombre | Consultar listado de pruebas |
| Actores | Administrador |
| Función | Consultar listado de pruebas disponibles |
| Descripción | El administrador visualizara un listado con todas las pruebas cargadas exitosamente a la base de datos. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.III.III. Casos de uso: Aplicativo

Tabla 20. Caso de uso análisis predictivo

| CASO DE USO | |
|-------------|--|
| Nombre | Análisis predictivo |
| Actores | Aplicativo |
| Función | mostrar el análisis predictivo |
| Descripción | El aplicativo se encargará de mostrar los resultados predictivos en cuanto a la influencia de los indicadores sobre los núcleos temáticos. Para la implementación de la parte predictiva el Se hizo uso del algoritmo J48 de la librería WEKA. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 21. Caso de uso Limpieza de datos.

| CASO DE USO | |
|-------------|--|
| Nombre | Limpieza y tratamiento de los datos |
| Actor | Aplicativo |
| Función | Realizar limpieza y el tratamiento adecuado de los datos |
| Descripción | El aplicativo realiza la limpieza de los datos previamente cargados por el administrador y procede a subirlos en la Base de Datos. |

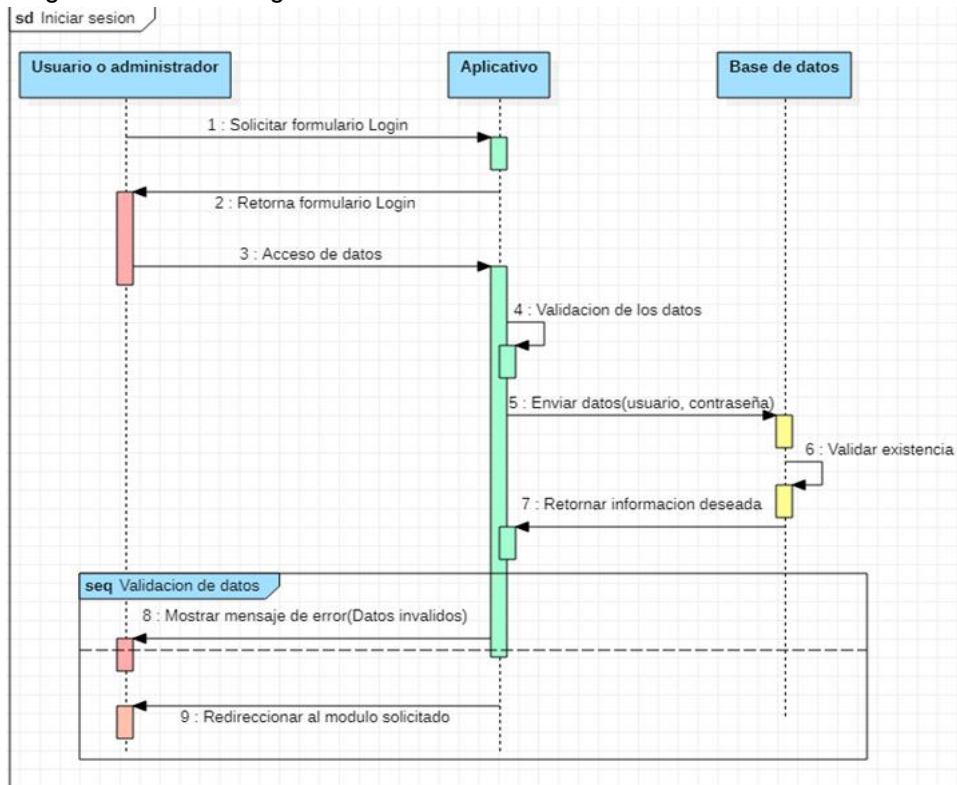
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV. Diagramas de secuencia

El diagrama de secuencia es usado para identificar el comportamiento de cada uno de los roles y la interacción de estos frente a cada acción, estos se deben elaborar por cada caso de uso.

V.I.IV.I Diagrama de secuencia de inicio de sesión

Figura 28. Diagrama de secuencia de inicio de sesión



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 22. Descripción diagrama de secuencia de inicio de sesión

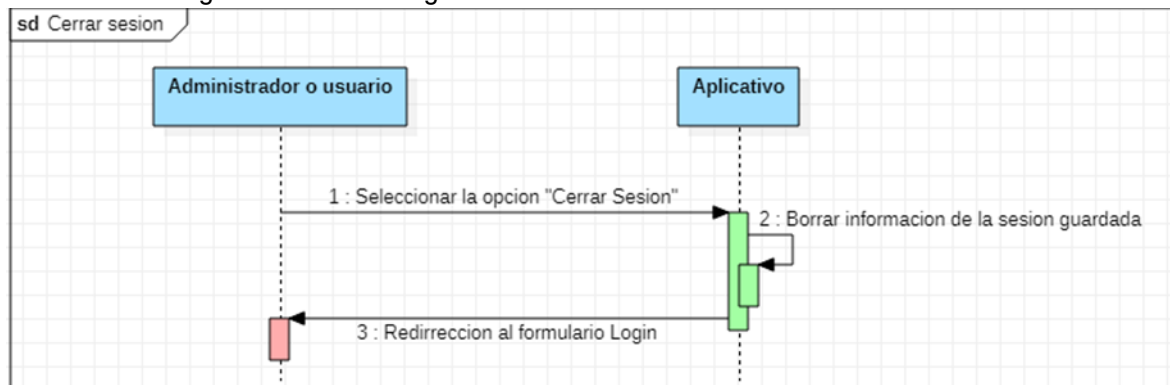
| Objeto | Descripción |
|-------------------------|--|
| Administrador o Usuario | El administrador o el usuario tiene una secuencia de actividades, la cual inicia con la solicitud del formulario de “Inicio de Sesión”, allí deberá ingresar los datos de usuario y contraseña, una vez se valide la información; si es correcta podrá ingresar al módulo solicitado de lo contrario se le mostrara un mensaje de error. |

| | |
|---------------|--|
| Aplicativo | El aplicativo se encargará inicialmente de mostrar el formulario solicitado, posteriormente realizara la validación de los datos ingresados al aplicativo y por último redireccionará al módulo solicitado o retornara un mensaje de error |
| Base de datos | La base de datos se encargará de verificar si el usuario y la contraseña ingresada coinciden con algún registro de la base de datos. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.II Diagrama se secuencia de cerrar sesión

Figura 29. Diagrama de secuencia: cerrar Sesión



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

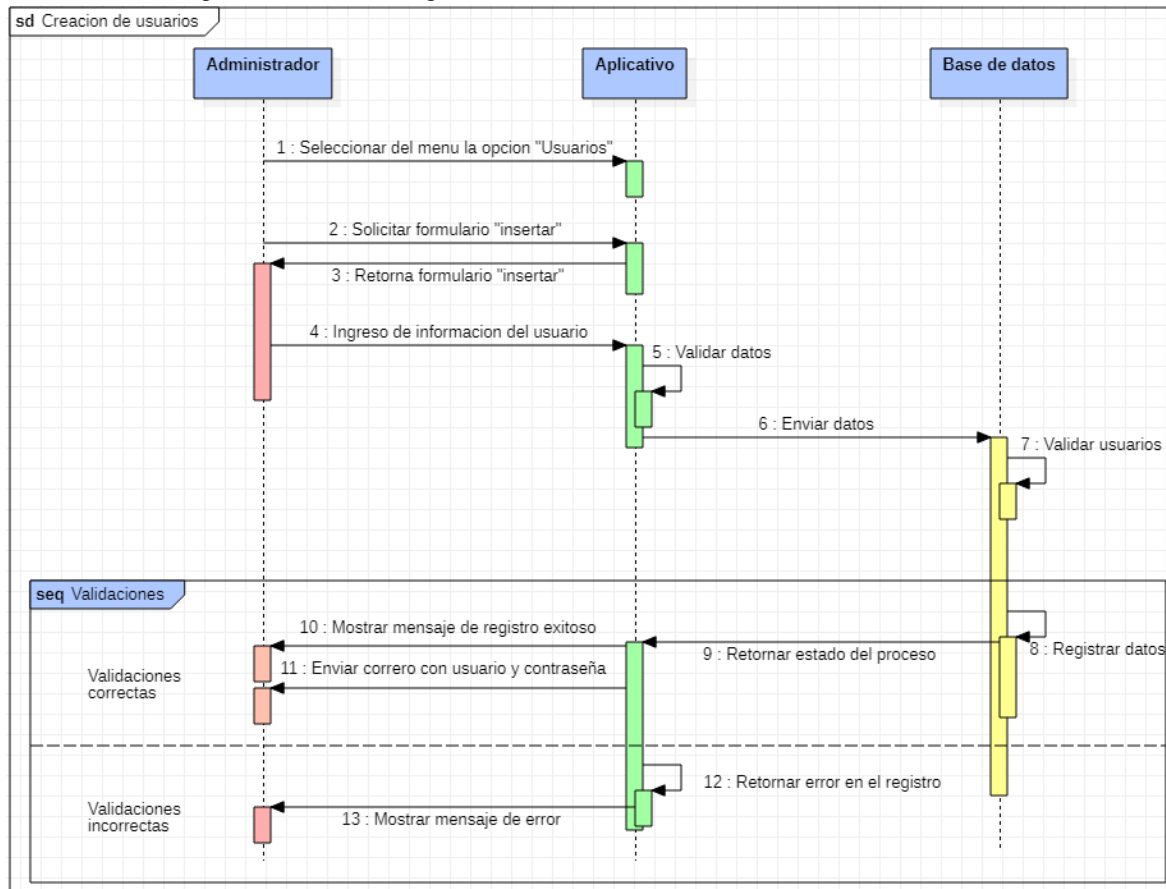
Tabla 23. Descripción diagrama de secuencia cerrar sesión

| Objeto | Descripción |
|-------------------------|---|
| Administrador o usuario | El administrador o el usuario tiene una secuencia de actividades, la cual cuenta con una sola acción que es la de "Cerrar sesión" |
| Aplicativo | El aplicativo se encargara de cerrar la sesión del usuario y borrar la sesión registrada dentro del navegador. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.III Diagrama de secuencia de creación de usuarios

Figura 30. Diagrama de secuencia: creación usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 24. Descripción diagrama de secuencias de creación de usuarios

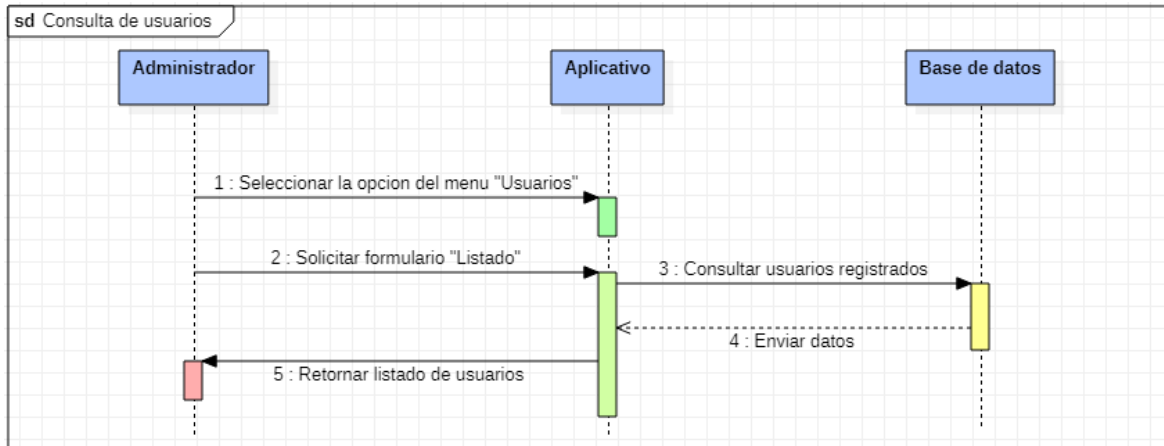
| Objeto | Descripción |
|---------------|--|
| Administrador | El administrador tiene una secuencia de actividades, la cual inicia con la selección de "Usuarios" del menú principal, posteriormente seleccionara la opción "Insertar". En el formulario deberá ingresar los datos del usuario a registrar, una vez se valide la información se le mostrara un mensaje indicándole si el proceso fue exitoso o presento inconvenientes. |

| | |
|---------------|--|
| Aplicativo | El aplicativo se encargara inicialmente de mostrar el formulario solicitado, posteriormente realizara la validación de los datos ingresados al aplicativo y por último, retornara un mensaje indicando si el proceso fue exitoso o no. Adicionalmente retornara un correo a la persona registrada indicando el usuario y la contraseña asignada. |
| Base de datos | La base de datos se encargara de validar que el usuario no exista y de insertar a la tabla de usuarios la información ingresada de la persona. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.IV Diagrama de secuencia de consultar usuarios

Figura 31. Diagrama de secuencia: consulta usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 25. Descripción diagrama de secuencia: consulta usuarios

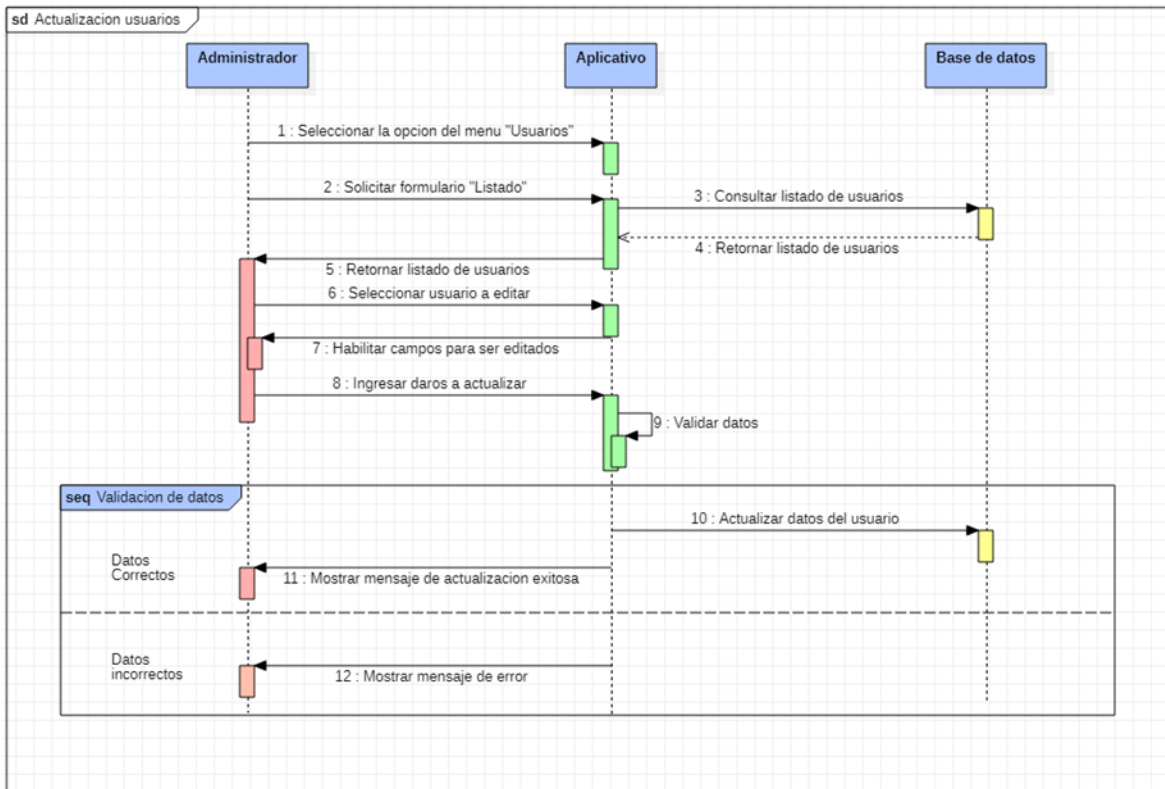
| Objeto | Descripción |
|---------------|--|
| Administrador | El administrador tiene una secuencia de actividades, la cual inicia con la selección de "Usuarios" del menú principal, posteriormente seleccionara la opción "Listado", después de ello el aplicativo le mostrara un listado con |

| | |
|---------------|---|
| | todos los usuarios registrados. |
| Aplicativo | El aplicativo se encargara inicialmente de generar un formulario con el listado de los usuarios registrados en base de datos. |
| Base de datos | La base de datos se encargara de retornar la información de los usuarios registrados. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.V Diagrama de secuencia de modificar usuarios

Figura 32. Diagrama de secuencia de la modificación de usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

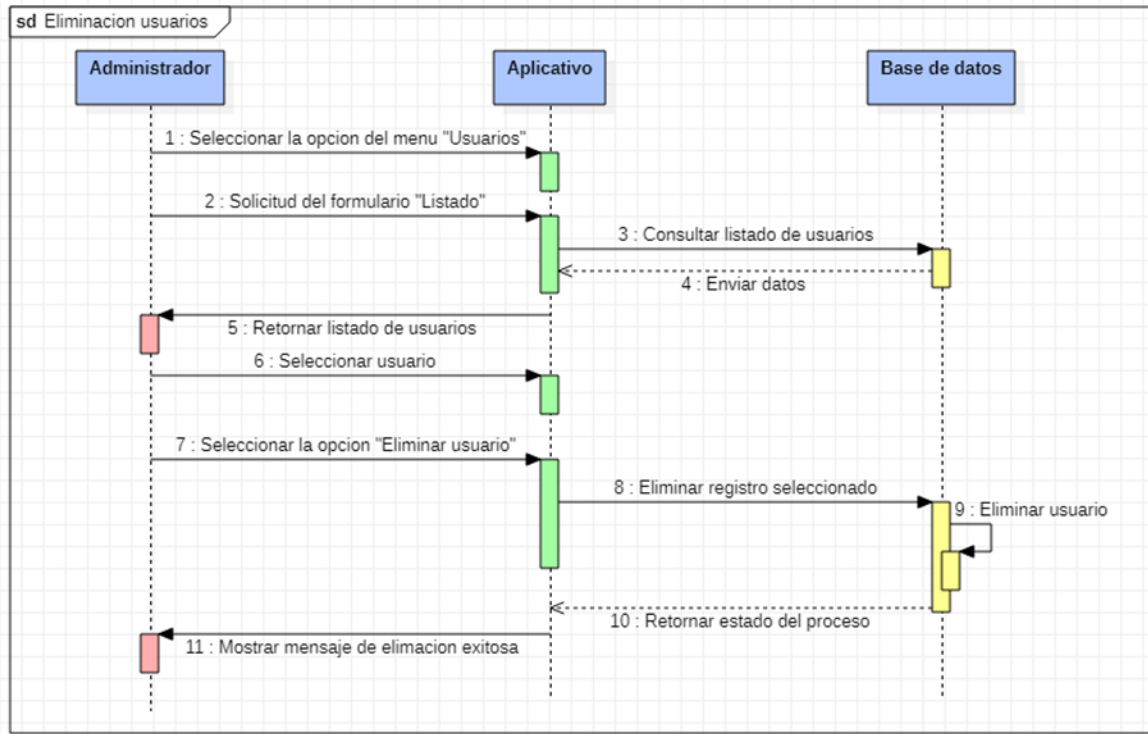
Tabla 26. Descripción diagrama de secuencia: modificación usuarios

| Objeto | Descripción |
|---------------|--|
| Administrador | El administrador tiene una secuencia de actividades, la cual inicia con la selección de "Usuarios" del menú principal, posteriormente seleccionara la opción "Listado", a partir del listado de usuarios, seleccionara el usuario a editar e ingresara los datos a modificar, una vez el aplicativo valide la información se le enviara un mensaje indicando el estado del proceso solicitado. |
| Aplicativo | El aplicativo se encargara inicialmente de generar un formulario con el listado de los usuarios registrados en base de datos, posteriormente se encargara de habilitar los campos para la edición y validar los datos ingresados, una vez realizado dicho proceso se encargara de enviar un mensaje indicando si el proceso fue exitoso. |
| Base de datos | La base de datos se encargara de consultar los usuarios registrados y de editar los datos del usuario seleccionado. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.VI Diagrama de secuencia de eliminar usuarios

Figura 33. Diagrama de secuencia de eliminar usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

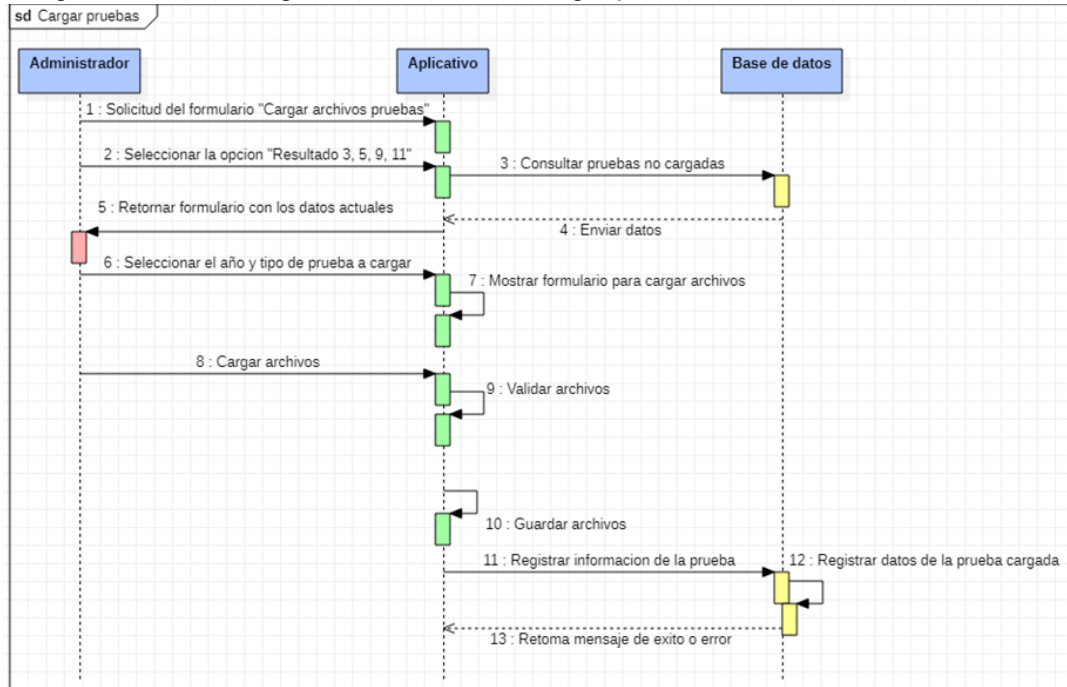
Tabla 27. Descripción diagrama de secuencia: eliminación usuarios

| Objeto | Descripción |
|---------------|---|
| Administrador | El administrador tiene una secuencia de actividades, la cual inicia con la selección de "Usuarios" del menú principal posteriormente seleccionara la opción "Listado". A partir del listado de usuarios seleccionara el usuario a eliminar. |
| Aplicativo | El aplicativo se encargara inicialmente de generar un formulario con el listado de los usuarios registrados en base de datos, posteriormente enviara un mensaje indicando la eliminación del usuario. |
| Base de datos | La base de datos se encargara de eliminar la información del usuario. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.VII Diagrama de secuencia de cargar los resultados

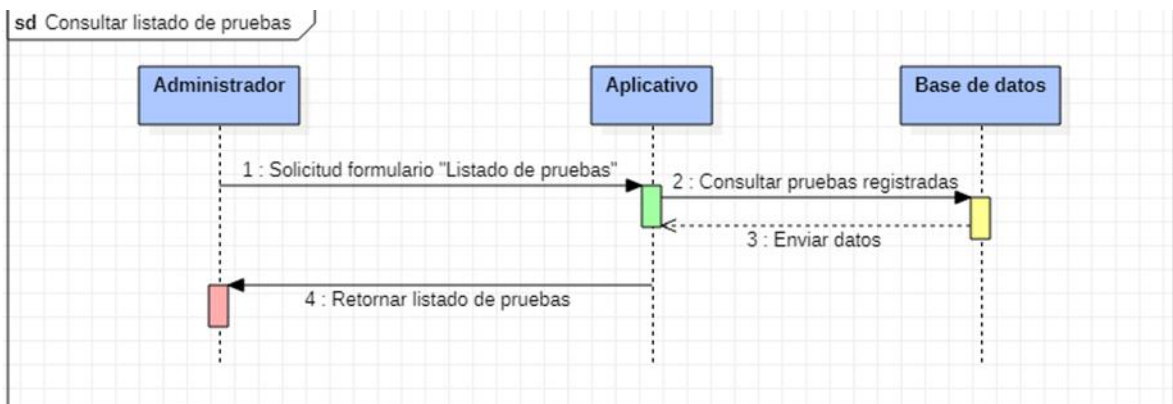
Figura 34. Diagrama de secuencia: cargar pruebas



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.VIII Diagrama de secuencia de consultar listado de pruebas

Figura 35. Diagrama de secuencia: consultar listado de pruebas



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

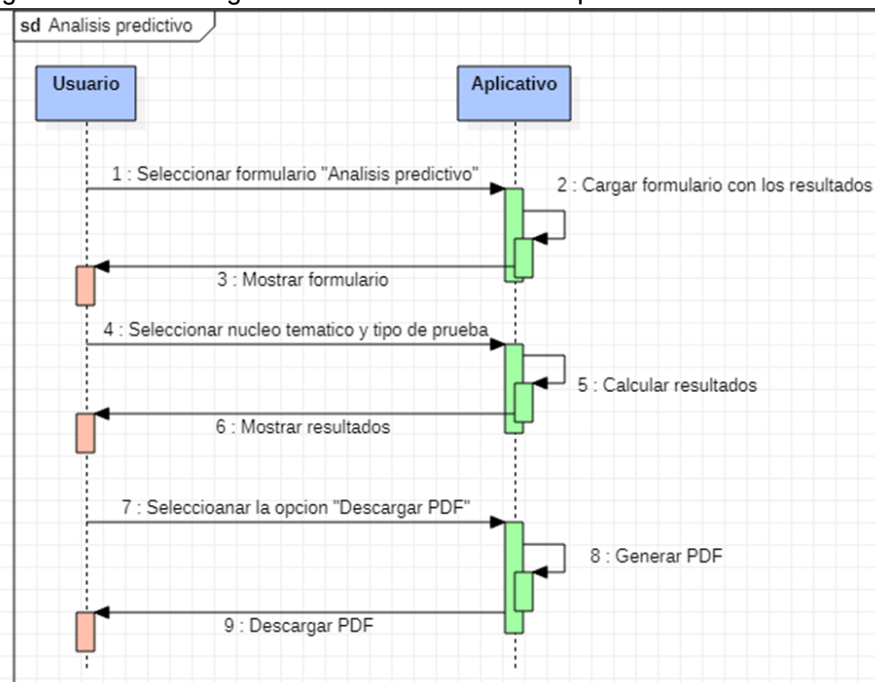
Tabla 28. Descripción diagrama de secuencia: consultar listado de pruebas

| Objeto | Descripción |
|---------------|--|
| Administrador | El administrador tiene una secuencia de actividades, la cual inicia con la solicitud del formulario "Listado de pruebas" |
| Aplicativo | El aplicativo se encargará inicialmente de solicitar a la base de datos el listado de pruebas cargadas, posteriormente se encargara de mostrarle al usuario el formulario con el listado de pruebas. |
| Base de datos | La base de datos se encargará de consultar el listado de pruebas registradas. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.IX Diagrama de secuencia del análisis predictivo

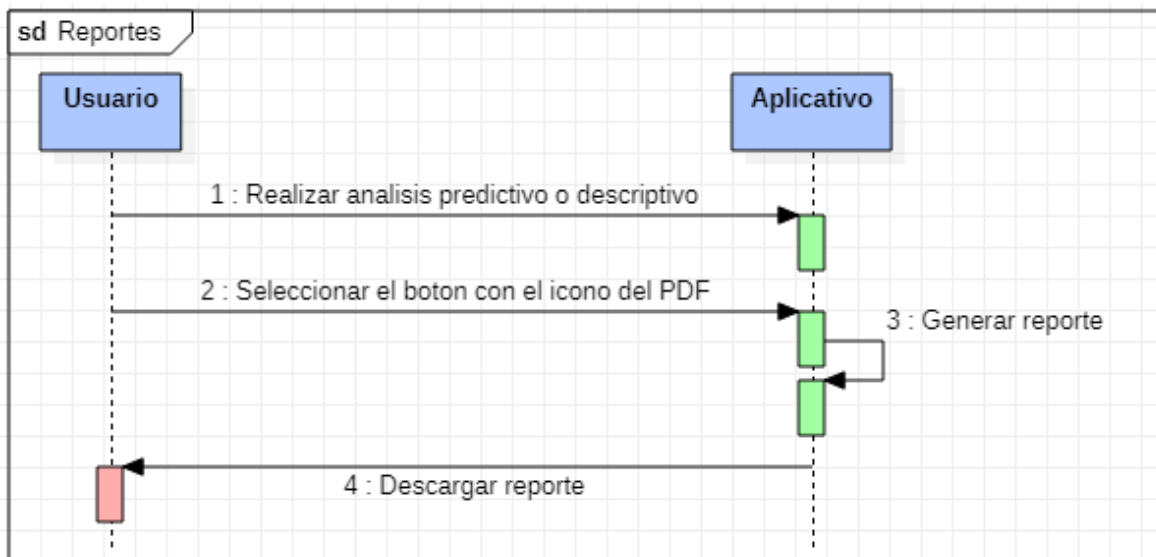
Figura 36. Diagrama de secuencia: análisis predictivo



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.I.IV.X Diagrama de secuencia de los reportes

Figura 37. Diagrama de secuencia: reportes



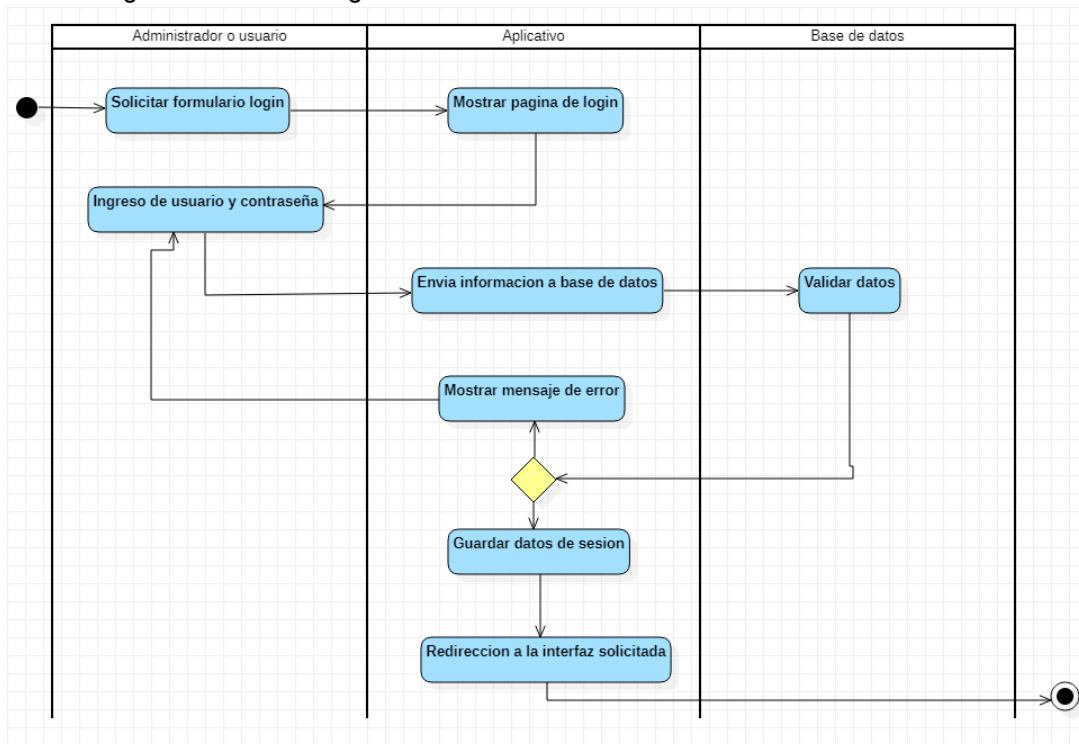
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V. Diagramas de actividad

En este diagrama se muestra un flujo de trabajo a través de una serie de acciones y roles encargados de ejecutarlas.

V.II.V.I. Diagrama de actividad de inicio de sesión

Figura 38. Diagrama de actividad: inicio de sesión



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 29. Descripción diagrama de actividad: inicio de sesión

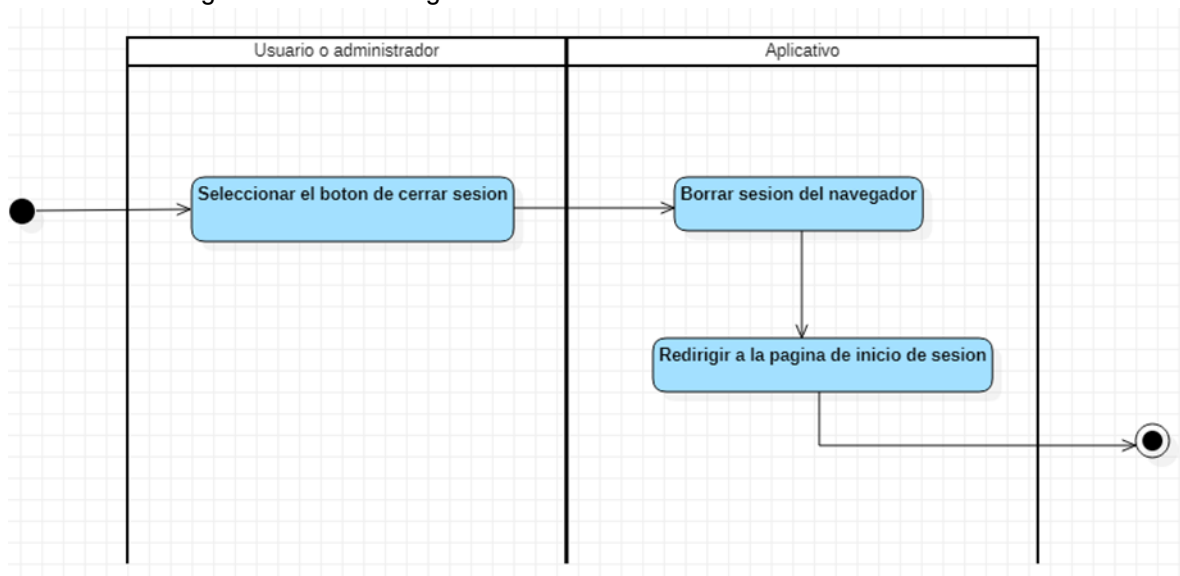
| Actividad | Descripción |
|------------------------------------|---|
| Solicitud formulario Login | El usuario o administrador ingresan al aplicativo web. |
| Mostrar página de Login | El aplicativo procede a mostrar el formulario del Login |
| Ingreso de usuario y contraseña | El usuario o administrador ingresan los datos de usuario y contraseña |
| Enviar información a base de datos | El aplicativo envía la información ingresada por el usuario a la base de datos |
| Validar datos | La base de datos se encargara de validar si los datos ingresados coinciden con algún registro. |
| Guardar datos de sesión | El aplicativo guardara los datos de sesión del usuario o administrador que se ha logueado con anterioridad. |

| | |
|--------------------------------------|--|
| Redirección a la interfaz solicitada | Si los datos son correctos el aplicativo se encargara de redireccionar al módulo solicitado. |
| Mostrar mensaje de error | En caso de que los datos sean incorrectos el aplicativo mostrara un mensaje de error. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.II. Diagrama de actividad de cerrar sesión

Figura 39. Diagrama de actividad: cerrar sesión



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

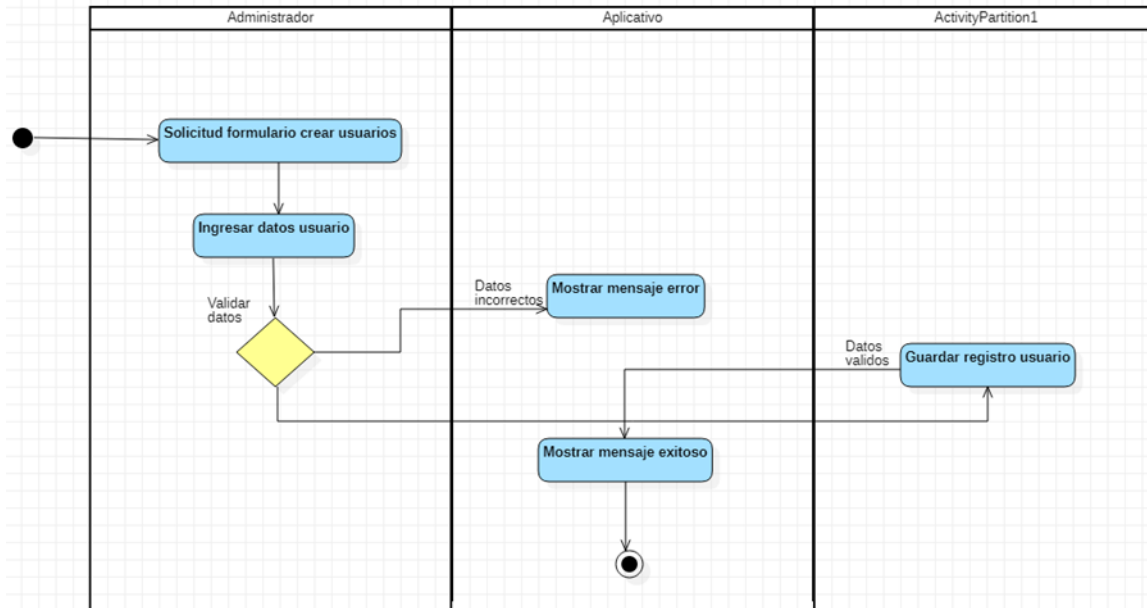
Tabla 30. Descripción diagrama de actividad: cerrar sesión

| Actividad | Descripción |
|---|---|
| Seleccionar el botón de cerrar sesión | El usuario o administrador seleccionara del menú principal el botón de cerrar sesión |
| Borrar sesión del navegador | El aplicativo se encargará de eliminar los datos de sesión registrados. |
| Redirigir a la página de inicio de sesión | El aplicativo cerrará sesión del módulo y redirigirá a la página de inicio de sesión. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.III. Diagrama de actividad de la creación de usuarios

Figura 40. Diagrama de actividad: creación usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 31. Descripción diagrama de actividad: creación usuarios

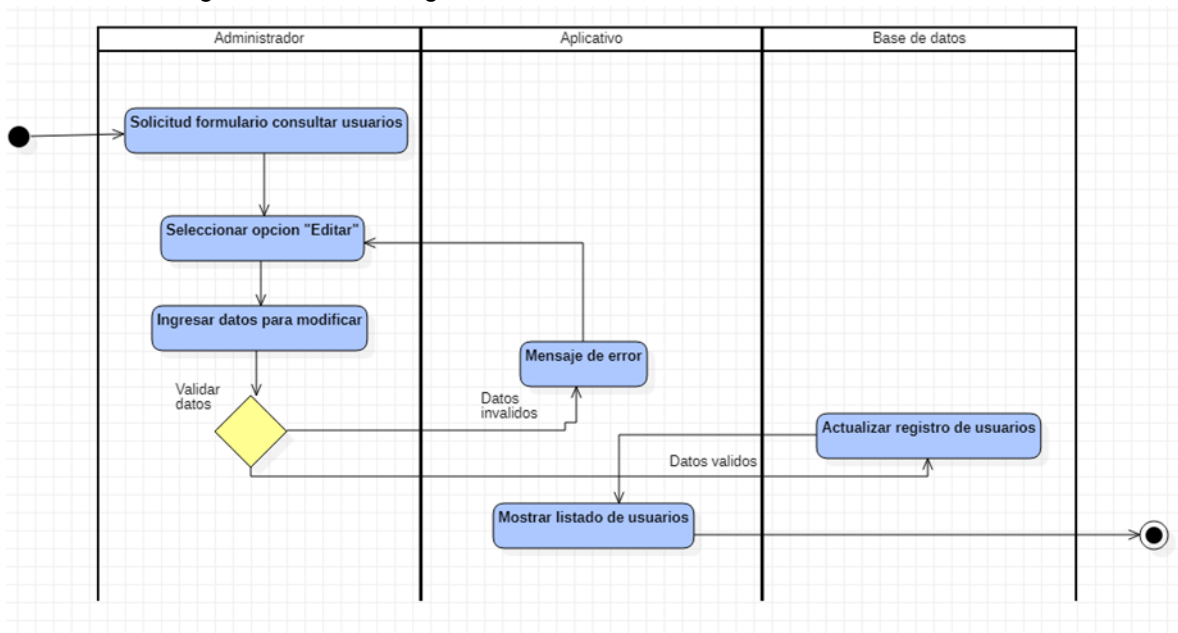
| Actividad | Descripción |
|---|--|
| Seleccionar del menú principal la opción "Usuarios" | Una vez iniciado sesión el administrador seleccionara del menú principal la opción "Usuarios" |
| Seleccionar la opción "Insertar" | El administrador seleccionara del submenú la opción "insertar", una vez seleccionado el aplicativo lo redireccionará a la página de registro de usuarios |
| Ingresar datos de usuario | El administrador llenara el formulario con los datos solicitados para el registro del usuario. |
| Validar datos ingresados | El aplicativo verifica que los datos ingresados por el administrador sean correctos y acordes a lo solicitado. |
| Verificar existencia de usuario | La base de datos se encargara de verificar que el usuario no este |

| | |
|------------------------------------|--|
| | registrado. |
| Mostrar mensaje de error | Si los datos son incorrectos o el usuario existe el aplicativo mostrara un mensaje de error indicando la falla presentada. |
| Guardar datos de usuario | Si los datos son correctos y el usuario no existe la base de datos guardara la información |
| Enviar por correo datos de usuario | Una vez guardada la información, el aplicativo enviara un correo indicando el usuario y la contraseña asignada. |
| Mostrar mensaje exitoso | Posteriormente, el aplicativo mostrara un mensaje indicando que el proceso fue exitoso |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.IV. Diagrama de actividad de modificación de usuarios

Figura 41. Diagrama de actividad: modificación usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 32. Descripción diagrama de actividad: modificación usuarios

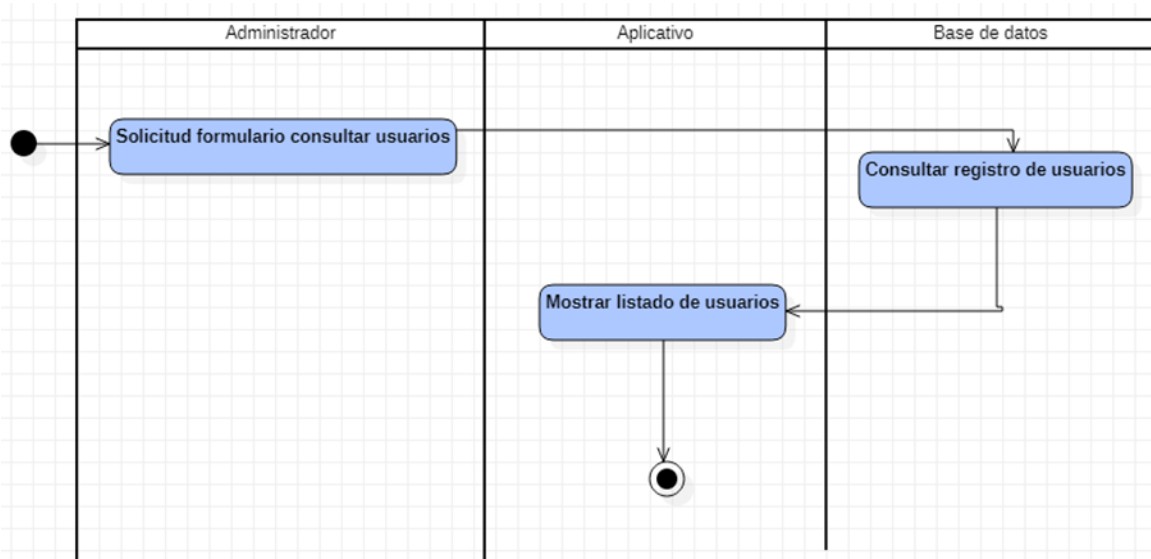
| Actividad | Descripción |
|---|---|
| Seleccionar del menú la opción "Usuarios" | El administrador una vez ha iniciado sesión, seleccionara del menú principal la opción "Usuarios" |
| Solicitud formulario "Listado" | Dentro del submenú de usuarios seleccionara la opción de "Listado". |
| Consultar datos de los usuarios | El aplicativo enviara una petición de consulta sobre los usuarios registrados a la base de datos |
| Consultar usuarios | La base de datos se encargara de buscar los usuarios registrados en ese momento. |
| Mostrar listado de usuarios | El aplicativo se encargara de mostrar los resultados de la consulta en un formulario. |
| Seleccionar usuario | El administrador se encargara de seleccionar del listado el usuario que desea modificar. |
| Seleccionar opción "Editar" | Una vez seleccionado el usuario, el administrador se encargara de seleccionar la opción editar, una vez seleccionada se habilitaran las casillas para su edición. |
| Ingresar datos para actualizar | El administrador modificara los datos que necesita actualizar. |
| Validar datos | Una vez ingresada la información el aplicativo se encargara de validar que los datos estén acorde a lo solicitado. |
| Mostrar mensaje de error | En caso de que la información ingresada presente inconsistencias el aplicativo mostrara un mensaje de error. |
| Actualizar datos | Si los datos ingresados son correctos, la base de datos se encargara de actualizar la información con los nuevos datos. |

| | |
|--|--|
| Mostrar mensaje de actualización exitosa | Una vez se registre la información en la base de datos el aplicativo mostrara un mensaje indicando que la actualización fue exitosa. |
|--|--|

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.V. Diagrama de actividad de consultar los usuarios

Figura 42. Diagrama de actividad: consulta usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 33. Descripción diagrama de actividad: consulta usuarios

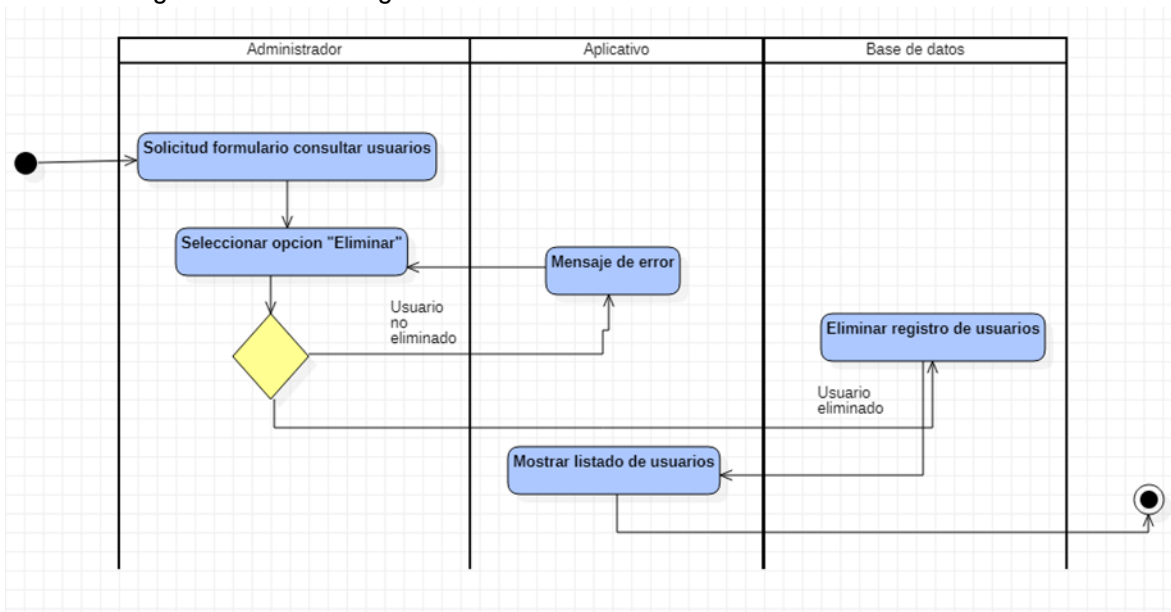
| Actividad | Descripción |
|--|---|
| Seleccionar del menú principal la opción "Usuarios" | El administrador una vez ha iniciado sesión, seleccionara del menú principal la opción "Usuarios" |
| Solicitud formulario "Listado" | Dentro del submenú de usuarios seleccionara la opción de "Listado". |
| Consultar datos de los usuarios registrados en base de datos | El aplicativo enviara una petición de consulta sobre los usuarios registrados a la base de datos |

| | |
|-----------------------------|---|
| Consultar usuarios | La base de datos se encargara de consultar los usuarios registrados en el momento. |
| Mostrar listado de usuarios | El aplicativo se encargara de mostrar los resultados de la consulta en un formulario. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.VI. Diagrama de actividad de la eliminación de usuarios

Figura 43. Diagrama de actividad: eliminación usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 34. Descripción diagrama de actividad: eliminación usuarios

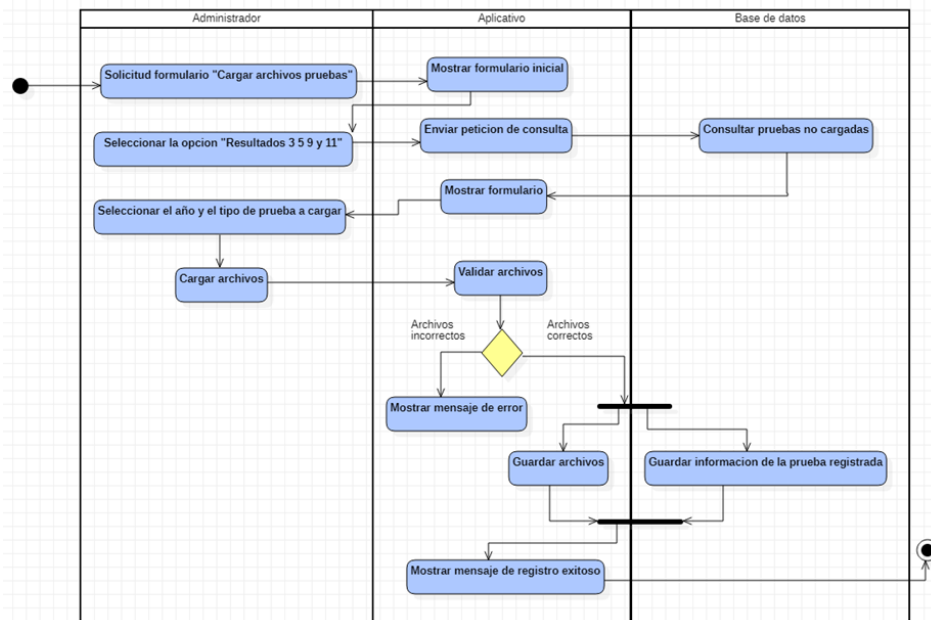
| Actividad | Descripción |
|--|---|
| Seleccionar del menú principal la opción "Usuarios" | El administrador una vez ha iniciado sesión, seleccionara del menú principal la opción "Usuarios" |
| Solicitud formulario "Listado" | Dentro del submenú de usuarios seleccionara la opción de "Listado". |
| Consultar datos de los usuarios registrados en base de datos | El aplicativo enviara una petición de consulta sobre los usuarios registrados a la base de datos |

| | |
|--|---|
| Consultar usuarios | La base de datos se encargará de buscar los usuarios registrados. |
| Mostrar listado de usuarios | El aplicativo se encarga de mostrar los resultados de la consulta en un formulario. |
| Seleccionar usuario | El administrador seleccionara del listado, el usuario que desea eliminar. |
| Seleccionar opción "Eliminar" | Una vez seleccionado el usuario, el administrador seleccionara la opción de eliminar. |
| Enviar ID del usuario a eliminar | El aplicativo enviara a la base de datos el id del usuario a eliminar. |
| Eliminar usuario | La base de datos se encargara de eliminar el usuario. |
| Mostrar mensaje de eliminación exitosa | Una vez se elimine la información del usuario el aplicativo mostrara un mensaje indicando que la eliminación fue exitosa. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.VII. Diagrama de actividad del cargue de los resultados

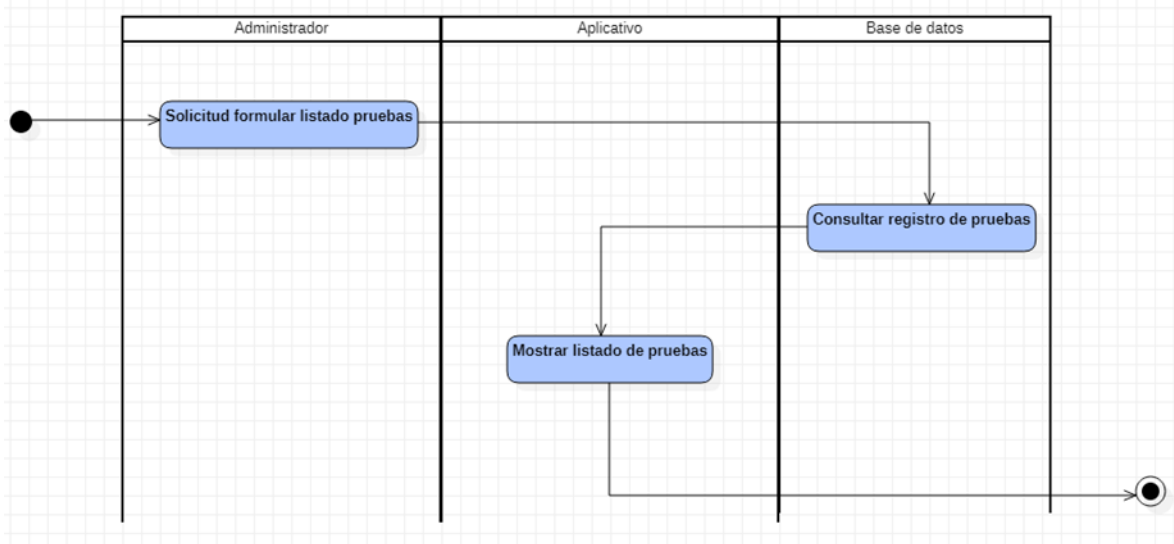
Figura 44. Diagrama de actividad: cargar resultados



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.VIII. Diagrama de actividad de consultar el listado de pruebas

Figura 45. Diagrama de actividad: consultar listado de pruebas



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

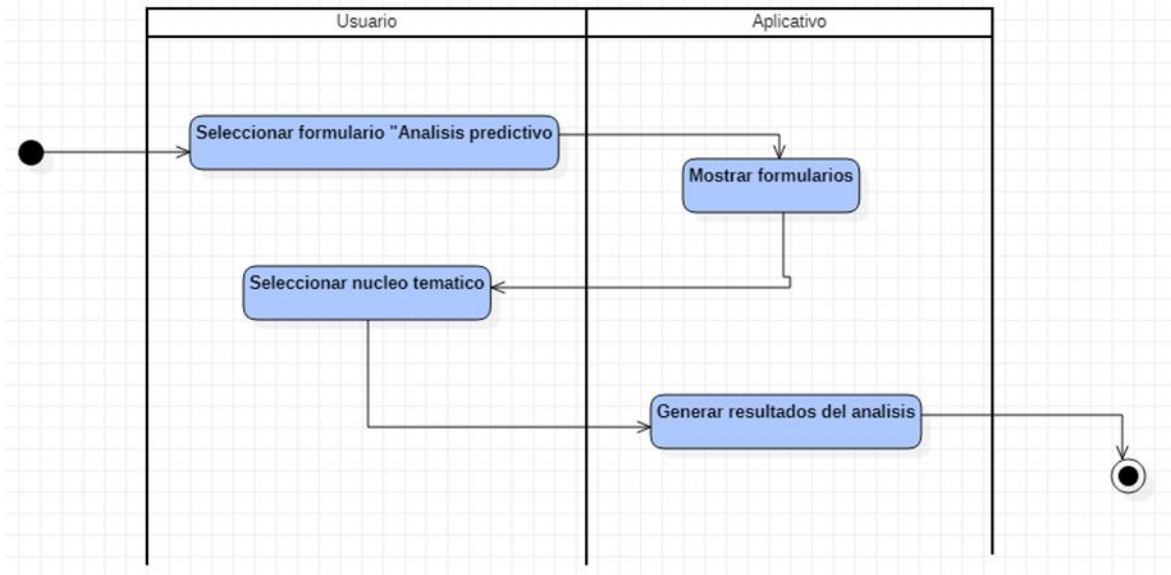
Tabla 35. Descripción diagrama de actividad: consultar listado de pruebas

| Actividad | Descripción |
|---|---|
| Solicitud formulario "Listado de pruebas" | Una vez ha iniciado sesión, el administrador seleccionara del menú principal la opción "Listado de pruebas" |
| Enviar petición de consulta de pruebas | El aplicativo enviara una petición de consulta a base de datos con respecto al listado de pruebas cargadas. |
| Mostrar listado de pruebas | El aplicativo se encargara de mostrar al usuario el listado de pruebas registradas. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.IX. Diagrama de actividad del análisis predictivo

Figura 46. Diagrama de actividad: análisis predictivo



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

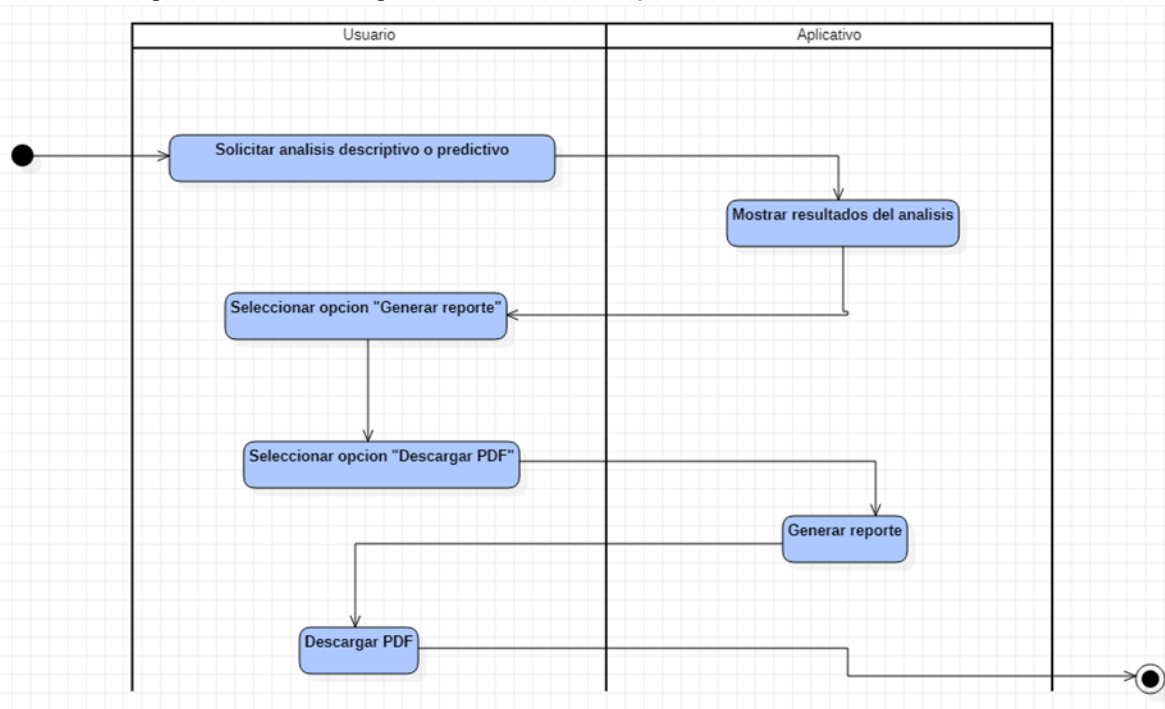
Tabla 36. Descripción diagrama de actividad: análisis predictivo

| Actividad | Descripción |
|--|--|
| Seleccionar del menú principal la opción "Análisis predictivo" | Una vez iniciada la sesión, el usuario seleccionara del menú principal la opción de "Análisis Predictivo" |
| Mostrar formulario | El aplicativo mostrara un formulario básico donde el usuario seleccionara el tipo de prueba y el núcleo temático. |
| Seleccionar el tipo de prueba y el núcleo temático | El usuario seleccionara el tipo de prueba y el núcleo temático a analizar. |
| Generar resultados | La base de datos se encargara de consultar los datos de los resultados, de acuerdo al tipo de prueba y al núcleo temático. |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.II.V.X. Diagrama de actividad de los reportes

Figura 47. Diagrama de actividad: reportes



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.III. Diseño de los casos de prueba

V.III.I. Usabilidad del aplicativo web

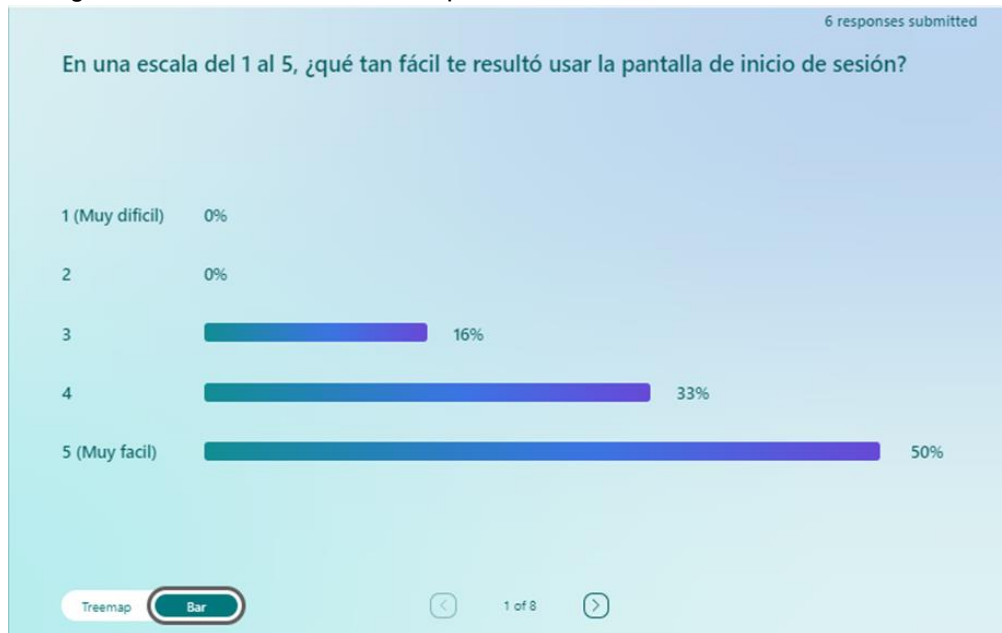
"La usabilidad web se refiere a la facilidad de uso que presenta una página o aplicación en línea. Su objetivo es permitir que los usuarios interactúen con el sitio de manera sencilla, intuitiva, agradable y segura (Rock Content, 2019)."

Para evaluar la usabilidad de nuestro aplicativo, implementamos dos enfoques distintos. En primer lugar, llevamos a cabo una encuesta en línea con la participación de 6 estudiantes de la Universidad de Cundinamarca. Estos estudiantes accedieron a nuestro aplicativo web y respondieron una serie de preguntas diseñadas en un formulario en línea. Esta encuesta nos proporcionó información valiosa sobre su experiencia de usuario.

Además, realizamos una entrevista en persona con un ingeniero de sistemas graduado de la Universidad de Cundinamarca. Durante esta entrevista, discutimos aspectos clave de la usabilidad del aplicativo y recopilamos datos cualitativos sobre su experiencia y

percepciones. Estos enfoques combinados nos permitieron obtener una visión más completa de la usabilidad de nuestro aplicativo web.

Figura 48. Encuesta forms parte 1



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El inicio de sesión, resulto ser muy fácil de entender y ninguno de los 6 estudiantes tuvo inconvenientes.

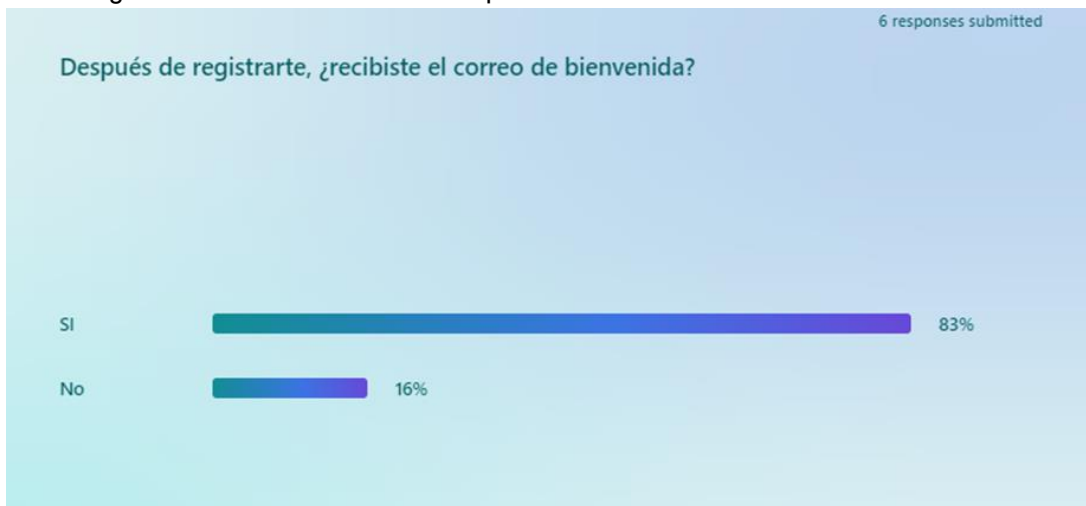
Figura 49. Encuesta forms parte 2



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El proceso de registro fue muy claro, y ninguno de los usuarios tuvo inconvenientes.

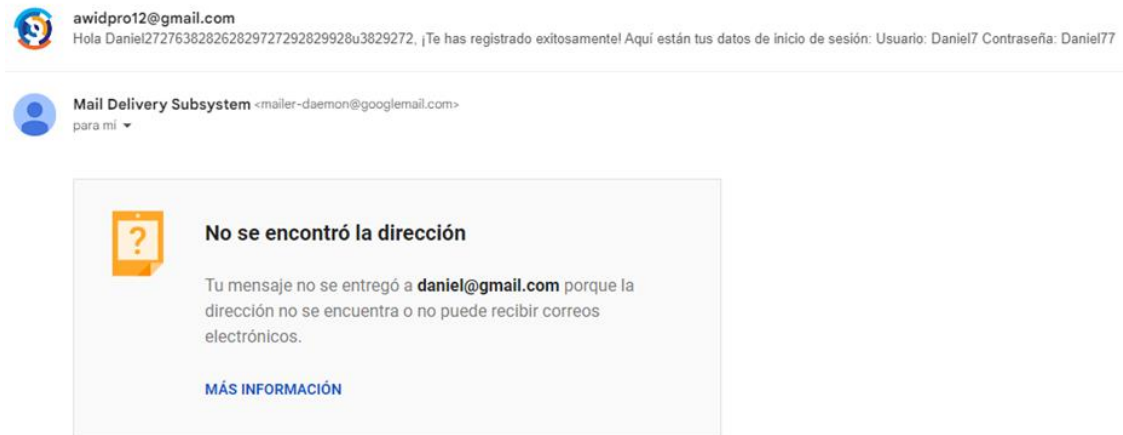
Figura 50. Encuesta forms parte 3



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Notamos que uno de los usuarios tuvo inconvenientes con el correo de bienvenida, esto se da cuando el usuario ingresa un correo no existente.

Figura 51. Evidencia de que el correo no existe



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

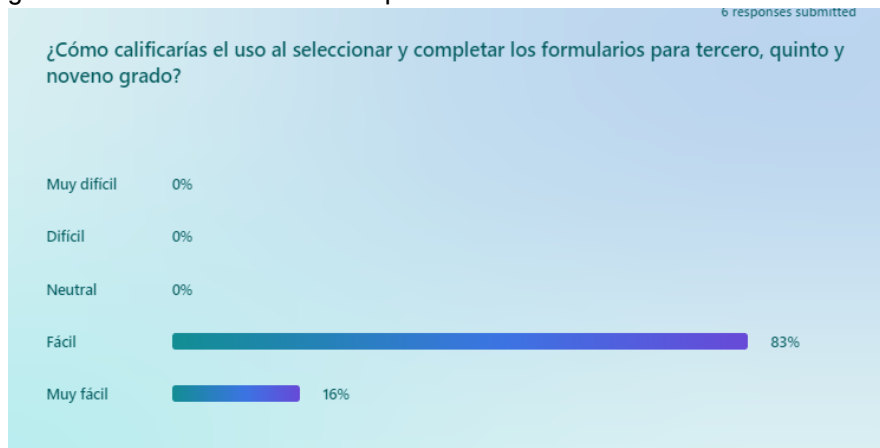
Figura 52. Encuesta forms parte 4



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Al momento de que el usuario seleccionara un reporte no hubo problema, ya que estuvo claro el proceso de selección de cada atributo.

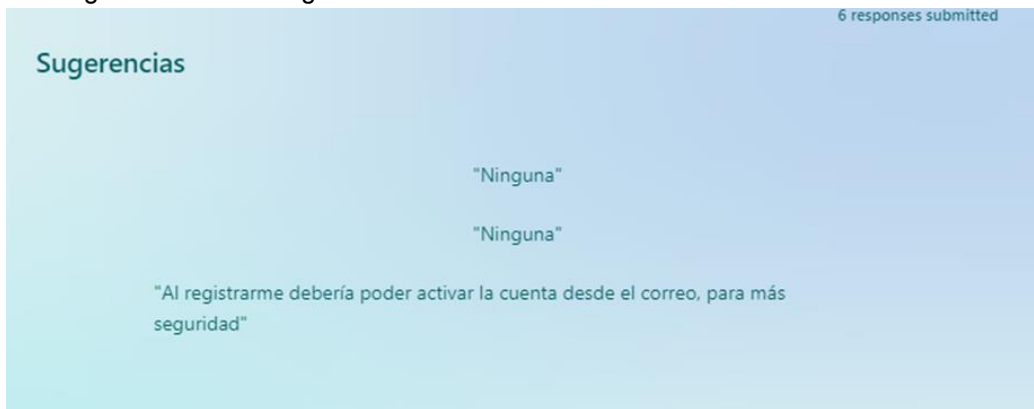
Figura 53. Encuesta forms parte 5



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Lo mismo en este caso al querer seleccionar un reporte de tercero, quinto y noveno, se evidenció que el proceso fue sencillo.

Figura 54. Sugerencias



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En el tema de seguridad un factor que se puede mejorar es la activación del usuario por medio del correo electrónico.

Figura 55. Evidencias de los registros

| Lista de Usuarios | | | | |
|--------------------|-----------------|------------------------------------|--------------------|--|
| Nombre | Apellido | Email | Usuario | Acciones |
| AutomatedTest23233 | Test | efrainvergara.udec+23233@gmail.com | AutomatedTest23233 | Editar Eliminar |
| luis | vergara | lfvergara@ucundinamarca.edu.co | luis123 | Editar Eliminar |
| johan | polania | jpolania101000@gmail.com | japolania | Editar Eliminar |
| Efrain Andrés | Vergara Serrato | efrainvergara.udec@gmail.com | eavergara96 | Editar Eliminar |
| Camilo | Mojca | motalora48@gmail.com | camilo0075 | Editar Eliminar |
| AutomatedTest19936 | Test | efrainvergara.udec+19936@gmail.com | AutomatedTest19936 | Editar Eliminar |
| Daniel | Leon | danie123975@gmail.com | ydleon | Editar Eliminar |

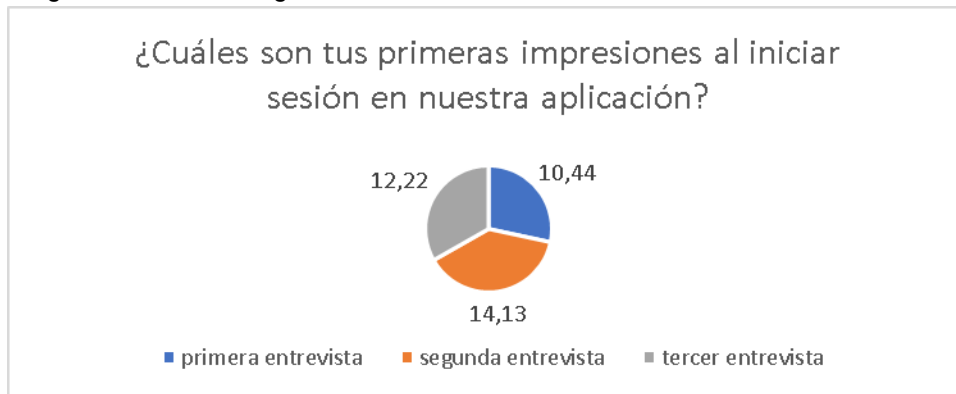
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Evidencia de los usuarios que se registraron exitosamente y realizaron la encuesta de usabilidad del aplicativo.

V.III.I.I. MAXQDA

Es una herramienta que permite a los investigadores organizar, analizar y visualizar datos textuales, audio y video. MAXQDA facilita la codificación, categorización y búsqueda de patrones en grandes conjuntos de datos cualitativos.

Figura 56. Pregunta N°1 encuesta



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

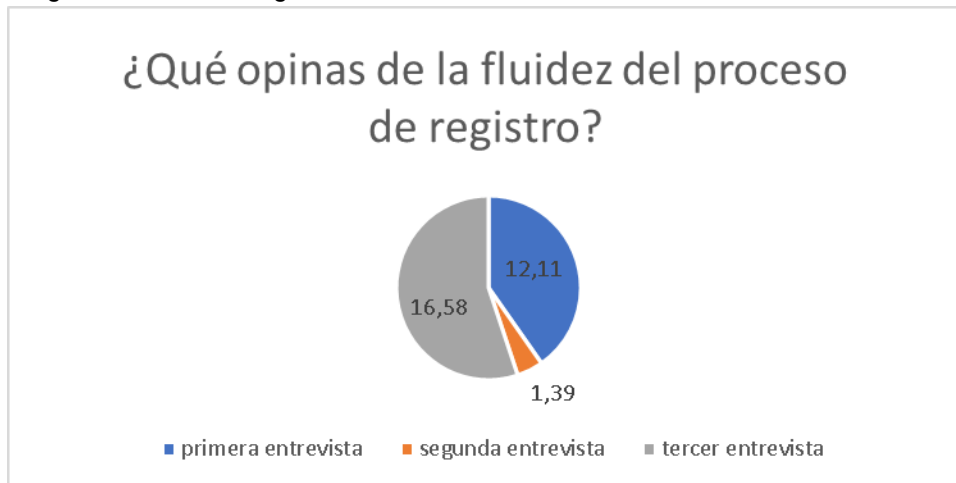
Tabla 37. Análisis de Cobertura de Segmentos en Entrevistas de Usuarios

| Document name | Segment | Coverage % |
|--------------------|---|------------|
| primera entrevista | Al iniciar sesión, noté que la pantalla de inicio es limpia y sencilla. La opción de usuario y contraseña es clara, y la disposición general parece ser amigable y fácil de entender. | 10.44 |
| segunda entrevista | Me gustó la presentación de la pantalla de inicio, muy elegante. Y encontrar la opción para poner mi usuario y contraseña fue fácil. | 14.13 |
| tercera entrevista | Pues la pantalla de entrada me llamó la atención. Colores agradables y la opción para poner mi usuario y contraseña estaba a la vista, así que me sentí bien. | 12.22 |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

La mayoría de los usuarios aprecian la simplicidad y claridad de la pantalla de inicio, junto con comentarios positivos sobre la presentación elegante y los colores agradables. Además, destaca que la ubicación de la opción para ingresar el usuario y la contraseña fue fácil de encontrar en todas las respuestas

Figura 57. Pregunta N°2 encuesta



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

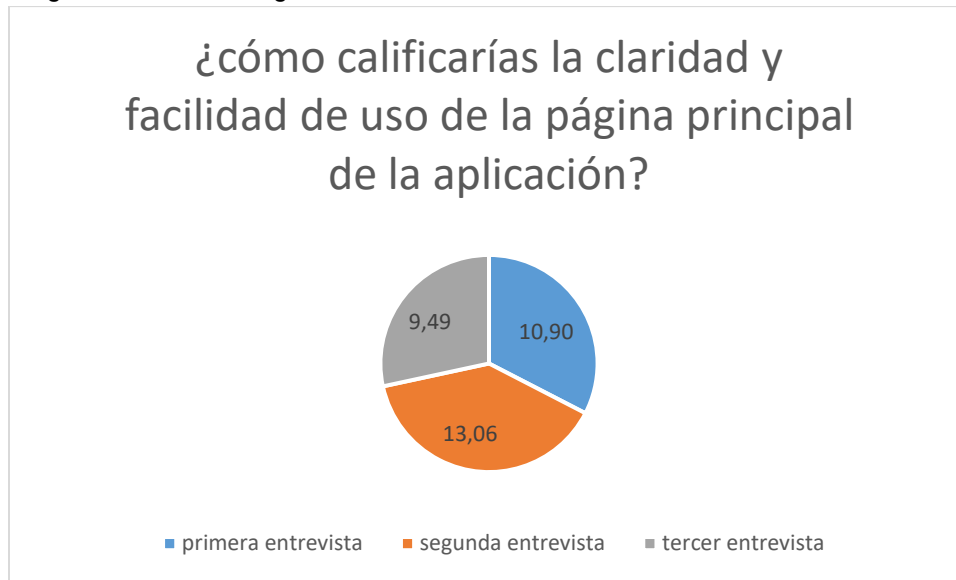
Tabla 38. Análisis de Cobertura de Segmentos en Entrevistas de Usuarios

| Document name | Segment | Coverage % |
|--------------------|---|------------|
| primera entrevista | El proceso de registro fue bastante sencillo. El formulario para ingresar datos personales, como nombre, apellido, usuario, correo y contraseña, fue intuitivo. Recibí el correo de bienvenida de manera oportuna. | 12.11 |
| segunda entrevista | Fue sencillo. | 1.39 |
| tercera entrevista | Bien, fue fácil. El formulario para poner mis datos fue como rellenar cualquier cosa y enseguida me llegó un mensaje de bienvenida. Aunque me hubiese gustado hacer el registro con Google para que fuera mas rápido. | 16.58 |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En general, la mayoría de los usuarios parece haber tenido una experiencia positiva en cuanto a la sencillez y la eficacia del proceso de registro. La tercera respuesta destaca la rapidez del proceso y expresa una preferencia por la opción de registro con Google para mayor rapidez

Figura 58. Pregunta N°3 encuesta



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

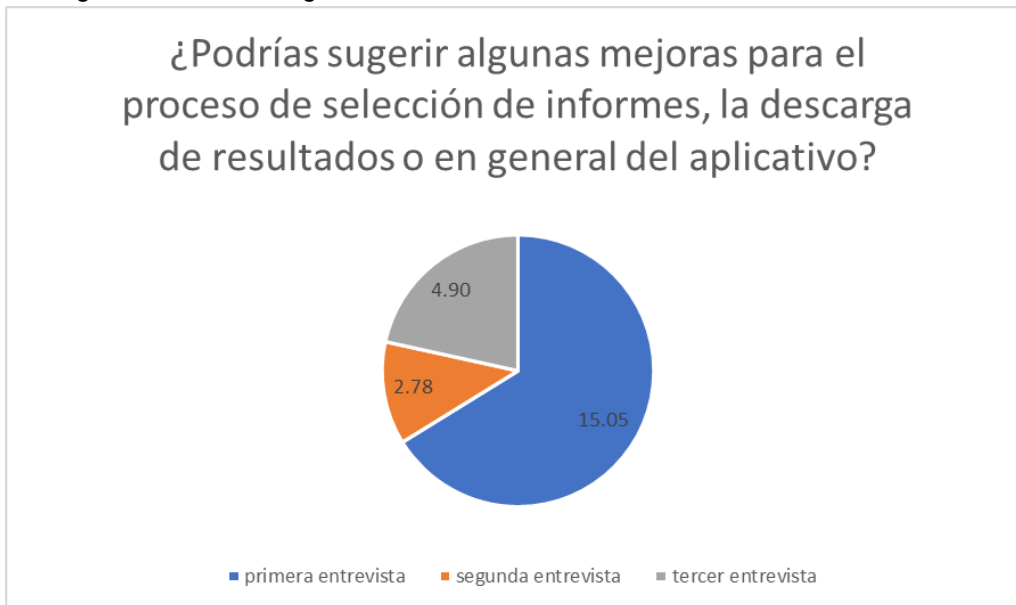
Tabla 39. Análisis de Cobertura de Segmentos en Entrevistas de Usuarios

| Document name | Segment | Coverage % |
|--------------------|---|------------|
| primera entrevista | La página principal es fácil de usar y tiene un diseño claro. Los formularios para seleccionar informes son fáciles de entender, y la navegación es fluida. En general, me resultó intuitiva. | 10.90 |
| segunda entrevista | Fácil de entender. Los botones para elegir informes eran bastante claros y el tablero dinámico me pareció muy interesante. | 13.06 |
| tercera entrevista | Fácil de usar. Los botones para elegir informes eran claros, y me moví por ahí sin problemas. En general, me sentí cómodo. | 9.49 |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

La mayoría de los usuarios perciben la página principal como fácil de usar, con botones claros y una navegación fluida. Además, se destacan aspectos específicos que contribuyen a una experiencia positiva, como la comprensión de formularios y el interés en características como el tablero dinámico.

Figura 59. Pregunta N°4 encuesta



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

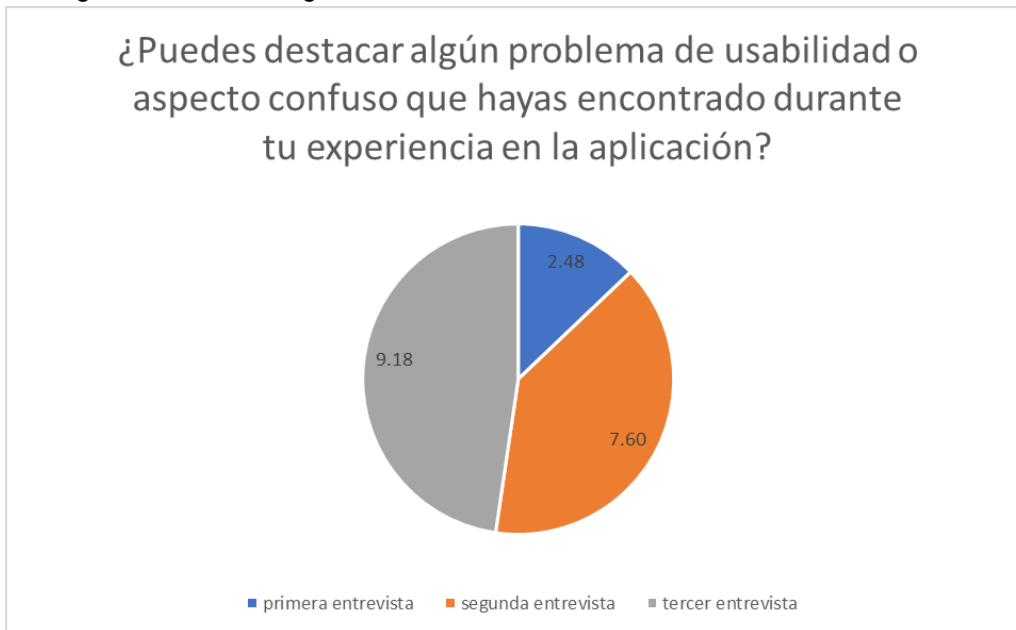
Tabla 40. Análisis de Cobertura de Segmentos en Entrevistas de Usuarios

| Document name | Segment | Coverage % |
|--------------------|---|------------|
| primera entrevista | Para mejorar, consideraría agregar una breve descripción o información adicional sobre cada informe disponible. Por mi parte, me gusta el inicio de sesión unificado por el tema de seguridad. Algo muy usado es Single Sign-On; a este aplicativo le sería muy útil. | 15.05 |
| segunda entrevista | Ninguna, me gusto tal cual | 2.78 |
| tercera entrevista | Pues quizás poner un poco más de información sobre cada informe | 4.90 |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Las sugerencias de mejora incluyen proporcionar información adicional sobre los informes disponibles, implementar un inicio de sesión unificado para mejorar la seguridad y considerar la posibilidad de incluir más detalles sobre cada informe.

Figura 60. Pregunta N°5 encuesta



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Tabla 41. Análisis de Cobertura de Segmentos en Entrevistas de Usuarios

| Document name | Segment | Coverage % |
|--------------------|--|------------|
| primera entrevista | En general, la aplicación es fácil de usar. | 2.48 |
| segunda entrevista | Me llamó la atención lo bonito que se ve todo y lo fácil que es usarlo. | 7.60 |
| tercera entrevista | Me encantó el tema del tablero dinámico, ya que podía ver diferentes resultados sin necesidad de descargar el reporte. | 9.18 |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Las respuestas indican una experiencia positiva y sin problemas de usabilidad en la aplicación. Los usuarios parecen apreciar la estética, la facilidad de uso y las funcionalidades específicas, como el tablero dinámico.

V.III.II. Pruebas de verificación del software

Para este apartado trabajamos con pruebas de automatización usando Selenium y Cucumber y para la generación de reportes usamos Allure.

"Selenium es un conjunto de utilidades que facilita la labor de obtener juegos de pruebas para aplicaciones web. Para ello nos permite grabar, editar y depurar casos de prueba, que podrán ser ejecutados de forma automática e iterativa posteriormente" (Juntadeandalucia, 2020). Por otro lado, "Cucumber es un software de testing BDD (Behavior Driven Development) que nos permitirá elaborar pruebas unitarias a partir de criterios de aceptación, fácilmente entendibles por todos los intervinientes del proceso. A través de Cucumber, el analista podrá definir un conjunto de casos de uso que permitan validar el desarrollo realizado. Estos casos de uso tendrán su correlación con escenarios de Cucumber, los cuales estarán implementados posteriormente en un lenguaje de programación dominado por el ingeniero de calidad o testing" (CHAKRAY, 2020). Por último allure es un marco de informes que se utiliza en el ámbito de pruebas de software.

Lo primero que se hizo, fue descargar las dependencias de selenium, cucumber y allure, luego de esto se cuadro el driverManager.java, con la url del proyecto.

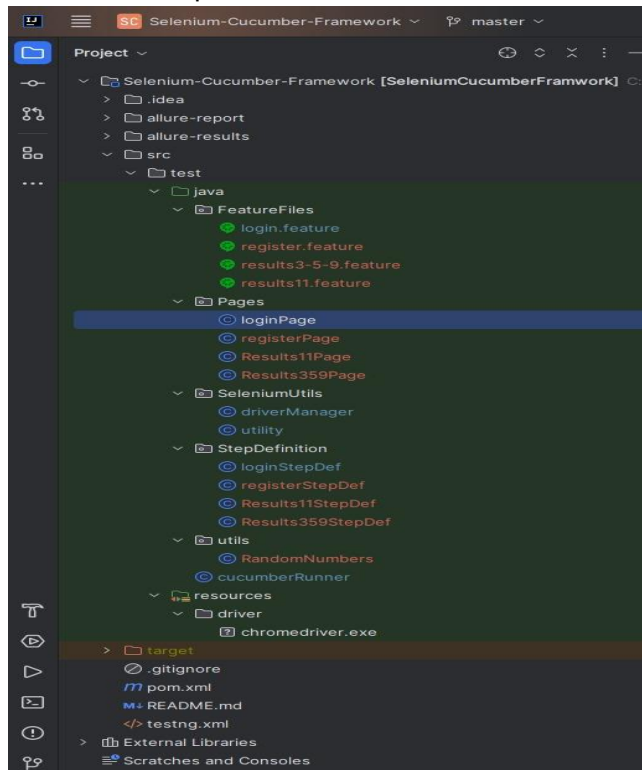
Figura 61. Descarga de dependencias

```
1 package SeleniumUtils;
2
3 > import
13
14 14 usages  sahrish.27 *
public class DriverManager {
15     10 usages
protected static WebDriver driver;
16
17
18 5 usages  sahrish.27 *
public WebDriver initializeDriver() {
19     if (driver == null) {
20         String browser = "chrome";
21         if (browser.equalsIgnoreCase("chrome")) {
22             ChromeOptions options = new ChromeOptions();
23             options.addArguments("--incognito");
24             System.setProperty("webdriver.chrome.driver", "src/test/resources/driver/chromedriver.exe");
25             driver = new ChromeDriver(options);
26             driver.manage().window().setSize(new Dimension(1440, 980));
27             driver.manage().window().maximize();
28
29             driver.get("https://2154-190-145-240-29.ngrok-free.app/proyecto/Login.jsp");
30             utility.pageLoad();
31
32         } else if (browser.equalsIgnoreCase("firefox")) {
33             System.setProperty("webdriver.gecko.driver", "geckodriver.exe");
34             driver = new FirefoxDriver();
35         } else if (browser.equalsIgnoreCase("ie")) {
36             System.setProperty("webdriver.ie.driver", "IEDriverServer.exe");
37             driver = new InternetExplorerDriver();
38         }
39         else {
40             throw new IllegalArgumentException("The Browser Type is Undefined");
41         }
42     }
43     return driver;
44 }
45
46
47
```

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Luego de eso se realizo como tal cada caso de prueba que íbamos a usar:

Figura 62. Casos de prueba



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En allure, podemos ver los casos de prueba que se realizaron:

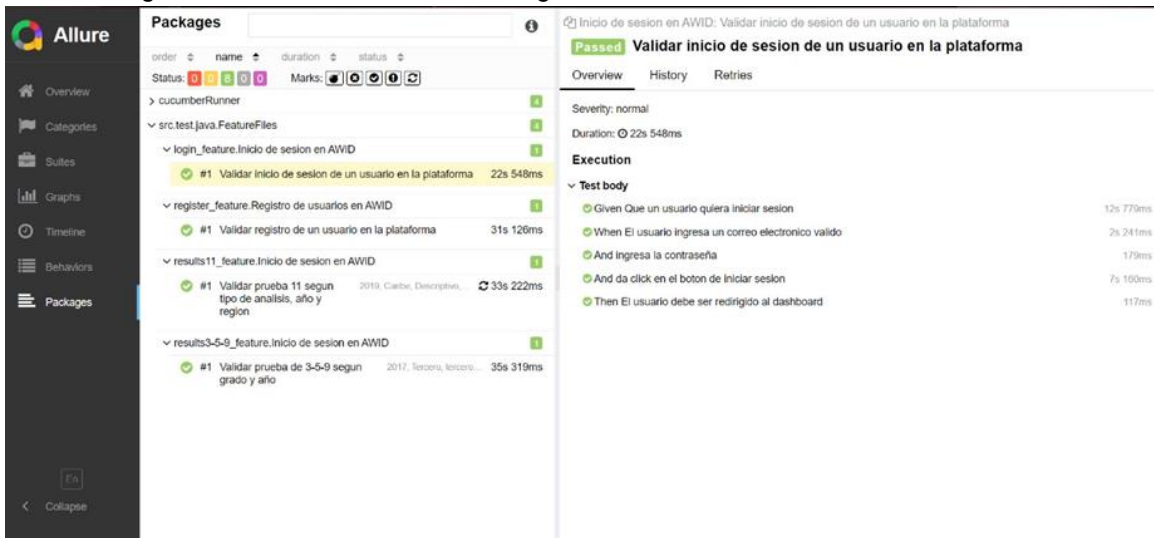
Figura 63. Casos de prueba ejecutados



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

La primera prueba automatizada fue el login de usuario:

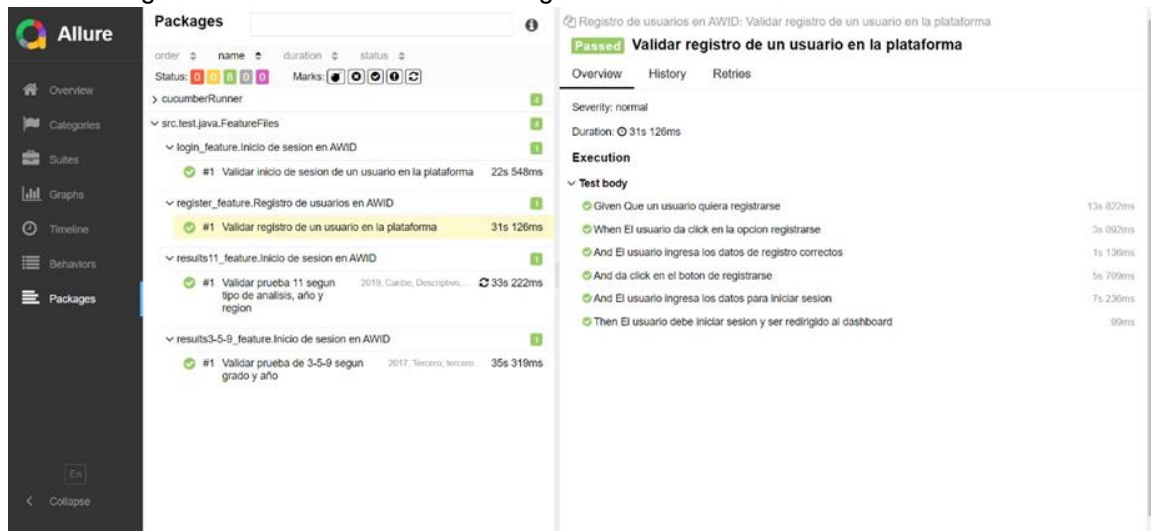
Figura 64. Automatización login



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El segundo fue el registro de usuario:

Figura 65. Automatizacion registro



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El tercero fue la validación para poder ver un reporte de once:

Figura 66. Validacion prueba once

The screenshot shows the Allure test results interface. On the left is a navigation sidebar with options like Overview, Categories, Suites, Graphs, Timeline, Behaviors, and Packages. The main area displays a list of test packages under 'src.test.java.FeatureFiles'. The selected test is '#1 Validar prueba 11 segun tipo de analisis, año y region' with a duration of 33s 222ms. The right-hand pane provides details for this test, including its severity (normal), duration, parameters (año: 2019, region: Caribe, tipoAnalisis: Descriptivo), and a detailed execution log of the test body steps.

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El último caso de prueba automatizada que realizamos fue ver la validación para poder ver un reporte de tercero, quieto y noveno:

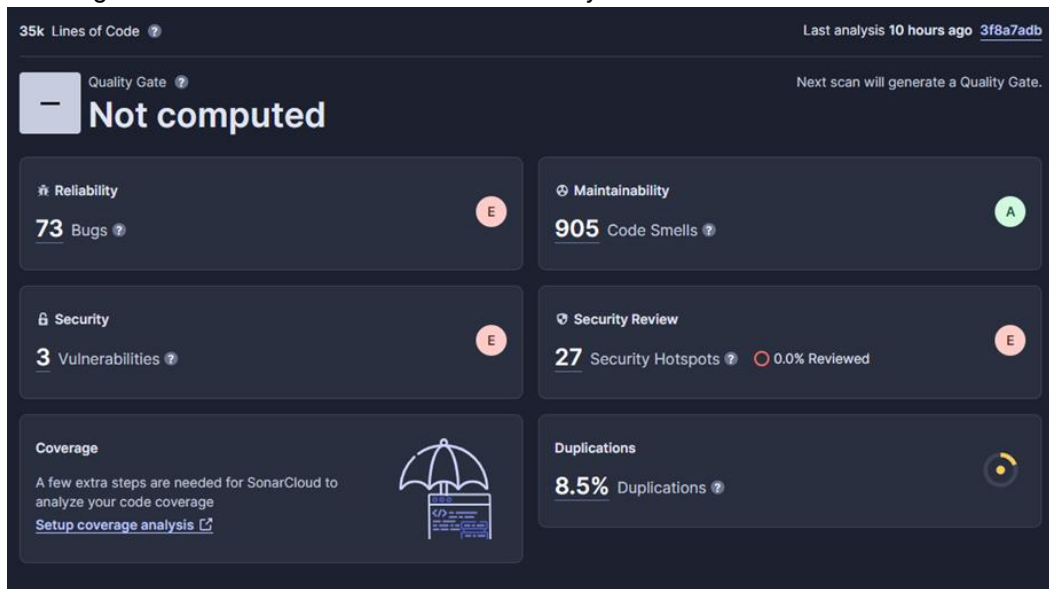
Figura 67. Automatización validación pruebas 3,5,9

The screenshot shows the Allure test results interface for a different test. The selected test is '#1 Validar prueba de 3-5-9 segun grado y año' with a duration of 35s 319ms. The right-hand pane shows details for this test, including its severity (normal), duration, parameters (año: 2017, grado: Tercero, título: tercero_2017.pdf), and a detailed execution log of the test body steps.

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

V.III.II.I. Sonarcloud

Figura 68. Resultado análisis del Proyecto en sonarcloud



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En total, SonarCloud analizó 35,366 líneas de código. En el tema de fiabilidad, la presencia de 73 bugs sugiere que hay segmentos propensos a errores, lo que podría afectar la estabilidad y confiabilidad del aplicativo. En cuanto a la mantenibilidad, se cuenta con una calificación de A; sin embargo, la presencia de 905 code smells indica áreas de mejora en la mantenibilidad del código. En seguridad, se identificaron 3 vulnerabilidades y 27 security hotspots, lo cual señala posibles riesgos en el código. Por último, en duplicidad de código contamos con un 8.5%, indicando la posibilidad de refactorizar fragmentos repetitivos para mejorar la eficiencia y reducir la complejidad.

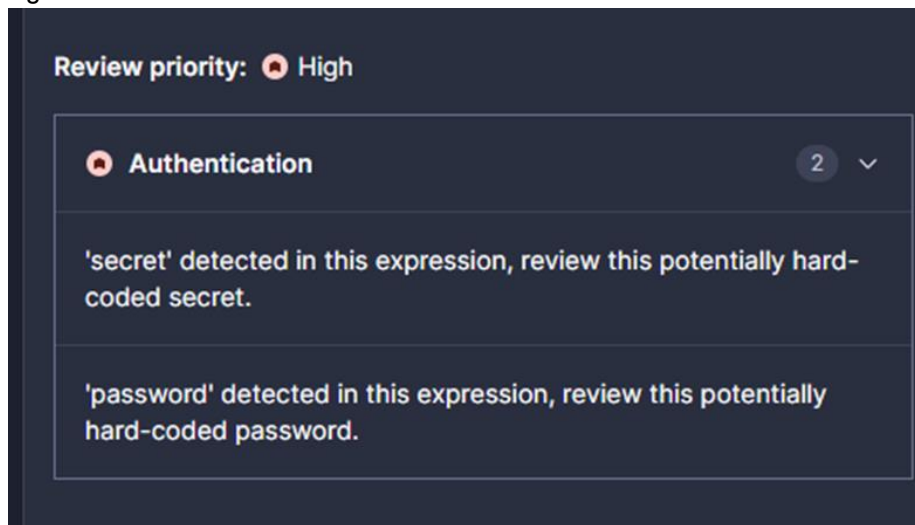
Figura 69. Total de archivos



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Se llevó a cabo el análisis en un total de 88 archivos de código fuente. Este proceso evaluó la calidad del software en términos de fiabilidad, mantenibilidad, seguridad y duplicidad de código. La identificación de bugs, code smells, vulnerabilidades y duplicaciones proporcionará una base sólida para mejorar y fortalecer la integridad del código.

Figura 70. Vulnerabilidades encontradas



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En el análisis de seguridad, se identificaron dos vulnerabilidades de alto riesgo. La primera vulnerabilidad se refiere al manejo del key de AWS, el cual está declarado una variable en el código. Este enfoque podría representar un riesgo, ya que la exposición de claves de acceso o credenciales en variables puede facilitar el acceso no autorizado a recursos de AWS.

Recomendaciones para solucionar esta vulnerabilidad:

- Usar un sistema de gestión de secretos:

Utilizar un sistema de gestión de secretos, como AWS Secrets Manager o HashiCorp Vault. Estos servicios te permiten almacenar y recuperar de forma segura información confidencial, como las claves de acceso y secretas.

- Configuración externa:

Almacenar las claves de acceso y secretas en un archivo de configuración externo que no esté incluido en tu repositorio de código. Este archivo debería ser accesible solo para las personas autorizadas y no debería ser visible públicamente.

- Variables de entorno:

Utilizar variables de entorno para pasar las claves de acceso y secretas a tu aplicación. Establece estas variables en el entorno de ejecución de tu aplicación, ya sea en el sistema operativo, en tu entorno de desarrollo, o en la plataforma de despliegue (por ejemplo, en AWS Lambda).

- Permisos mínimos necesarios:

Asegurar de que las claves de acceso utilizadas tengan los permisos mínimos necesarios para realizar las operaciones que tu aplicación requiere en Amazon S3. No otorgues más permisos de los necesarios.

- Rotación regular de claves:

Realizar rotación regular de claves para mejorar la seguridad. Cambia las claves de acceso y secretas periódicamente y actualiza la configuración en consecuencia.

- Proteger el acceso a claves:

Limitar el acceso a las claves de acceso y secretas a las personas y sistemas autorizados. No compartas estas claves de forma indiscriminada y revócalas cuando ya no sean necesarias.

La segunda vulnerabilidad está relacionada con la conexión a la base de datos, donde la contraseña está almacenada como una variable en el aplicativo. Guardar contraseñas directamente en el código fuente puede ser una práctica insegura, ya que podría exponer las credenciales en caso de acceso no autorizado.

Principales recomendaciones para esta vulnerabilidad son:

- No Almacenar Credenciales en el Código Fuente:

Evita almacenar las credenciales directamente en el código fuente. Este enfoque es inseguro y puede llevar a la exposición accidental de información sensible.

- Utilizar Variables de Entorno o Archivos de Configuración:

Almacenar las credenciales en variables de entorno o en un archivo de configuración externo. Esto permite separar la configuración de la lógica del programa y facilita la gestión de las credenciales en diferentes entornos (desarrollo, prueba, producción).

- Implementar un Gestor de Secretos:

Considerar el uso de un gestor de secretos, como AWS Secrets Manager, HashiCorp Vault o Azure Key Vault. Estos servicios están diseñados para gestionar de forma segura y centralizada las credenciales y otros secretos.

- Limitar los Permisos de la Cuenta de Base de Datos:

Asignar los permisos mínimos necesarios para la cuenta de base de datos utilizada por tu aplicación. No otorgues más privilegios de los necesarios para realizar las operaciones requeridas.

- Encriptar las Credenciales:

Considerar encriptarlas para agregar una capa adicional de seguridad. Asegúrate de manejar de manera segura las claves de cifrado.

- Realizar Auditorías de Seguridad:

Realizar auditorías de seguridad periódicas para identificar posibles vulnerabilidades y asegurarte de que las mejores prácticas de seguridad se estén siguiendo.

VI. Estimación de recursos con punto de función COSMIC

VI.I. Identificación de Requerimientos funcionales:

RF01 - Login:

Descripción: El usuario puede acceder a la plataforma mediante un login, ingresando usuario y contraseña

RF02 - Regístrese:

Descripción: Los clientes pueden completar un formulario para solicitar registro en el aplicativo.

RF03 - Creación usuarios:

Descripción: El administrador puede registrar nuevos usuarios al aplicativo, con información enviada al correo del usuario.

RF04 - Consulta usuarios:

Descripción: El administrador puede visualizar el listado de usuarios existentes en el sistema.

RF05 - Modificación usuarios:

Descripción: El administrador puede actualizar/modificar la información de los usuarios, seleccionando el usuario deseado.

RF06 - Eliminación usuarios:

Descripción: El administrador puede eliminar usuarios.

RF07 - Carga de reportes:

Descripción: El administrador llena un formulario, cargando tipo de análisis, región, año y subiendo el reporte en formato PDF.

RF08 - Limpieza de los datos:

Descripción: El aplicativo realiza la limpieza de datos provenientes del ICFES para análisis en WEKA.

RF09 - Visualización de resultados:

Descripción: Los usuarios pueden buscar y visualizar resultados en dos menús de reportes.

RF10 - Carga de los datos:

Descripción: El aplicativo carga datos transformados en reportes a la base de datos.

RF11 - Reportes:

Descripción: El aplicativo proporciona al usuario reportes según criterios como grado, tipo de análisis, año y región.

VI.II. Identificación de datos de entrada y salida

- RF01 (Login): EE (Entrada Externa)
- RF02 (Regístrese): EE (Entrada Externa)
- RF03 (Creación usuarios): EE (Entrada Externa)
- RF04 (Consulta usuarios): CE (Consulta Externa)
- RF05 (Modificación usuarios): EE (Entrada Externa)

- RF06 (Eliminación usuarios): EE (Entrada Externa)
- RF07 (Carga de reportes): EE (Entrada Externa) + SE (Salida Externa)
- RF08 (Limpieza de los datos): EE (Entrada Externa)
- RF09 (Visualización de los resultados): CE (Consulta Externa)
- RF10 (Carga de los datos): EE (Entrada Externa)
- RF11 (Reportes): CE (Consulta Externa) + SE (Salida Externa)

VI.III. Estimación de puntos de función

Dado el análisis de requerimientos y su naturaleza como Entrada Externa (EE), Salida Externa (SE), y Consulta Externa (CE), asignaremos valores en puntos de función según la siguiente lógica:

- Para cada Entrada Externa (EE): 3 puntos de función
- Para cada Salida Externa (SE): 4 puntos de función
- Para cada Consulta Externa (CE): 3 puntos de función

Tabla 42. Valor en puntos de función

| Requerimiento | Clasificación | Puntos de Función |
|--|---------------|-------------------|
| RF01 (Login) | EE | 3 |
| RF02 (Regístrate) | EE | 3 |
| RF03 (Creación usuarios) | EE | 3 |
| RF04 (Consulta usuarios) | CE | 3 |
| RF05 (Modificación usuarios) | EE | 3 |
| RF06 (Eliminación usuarios) | EE | 3 |
| RF07 (Carga de reportes) | EE + SE | 7 (3 + 4) |
| RF08 (Limpieza de los datos) | EE | 3 |
| RF09 (Visualización de los resultados) | CE | 3 |
| RF10 (Carga de los datos) | EE | 3 |
| RF11 (Reportes) | CE + SE | 7 (3 + 4) |

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Total de Puntos de Función: 42 puntos de función

VI.IV. Roles de equipo

- Investigador Principal
- Investigador Auxiliar (x2)

VI.V. Horas de Trabajo por Semana (por Investigador):

- 14 horas/semana

VI.VI. Semanas de Trabajo para 10 meses

- 14 horas/semana * 4.33semanas/mes* 10 meses = 606.2 horas por investigador

VI.VII. Tasas de Pago por Hora:

- Investigador Principal: 50,000 COP/hora
- Investigador Auxiliar: 40,000 COP/hora

VI.VIII. Determinación de Esfuerzo:

- Basados en los requerimientos, se determinó un total de 42 puntos de función COSMIC.
- Factor de Productividad (FP): Se aplicó un FP de 14 horas por punto de función, basado en la nueva carga de trabajo.
- Esfuerzo Total = Puntos de Función * FP = 42 * 14 = 588 horas

VI.IX. Distribución del Esfuerzo:

- Investigador Principal: 60%
- Investigador Auxiliar (x2): 40%
- Esfuerzo Investigador Principal = 588 * 0.6 = 352.8 horas
- Esfuerzo Investigador Auxiliar (x2) = 588 * 0.4 = 235.2 horas por investigador

VI.X. Estimación de Costos Por Rol (para 10 meses):

- Costo Investigador Principal = 352.8 * 50,000 = 17,640,000 COP
- Costo Investigador Auxiliar (x2) = 235.2 * 40,000 = 9,408,000 COP por investigador

VI.XI. Costo Total para 10 meses:

- Costo Total = Costo Investigador Principal + Costo Investigador Auxiliar (x2) = 17,640,000 + 9,408,000 (x2) = 36,456,000 COP

VI.XII. Valor Monetario Total para 10 meses (Incluyendo Gastos Asociados):

- Se consideran costos asociados como recursos informáticos, licencias de software, etc., que podrían ser un 20% adicional al costo total.
- Valor Monetario Total = Costo Total * 1.2 = 36,456,000 * 1.2 = 43,747,200 COP

VII. Resultados

Los resultados obtenidos están respaldados por un exhaustivo análisis bibliográfico realizado para el estado de arte de este proyecto, donde representan una contribución significativa para la construcción de un aplicativo web destinado a la formulación de indicadores de desempeño en las Pruebas Saber en Colombia. Este proceso se fundamentó en la revisión crítica de diversas fuentes bibliográficas, abordadas por expertos y académicos especializados en el ámbito educativo. La recopilación de indicadores de desempeño extraídos de estas fuentes proporciona una sólida base teórica y conceptual, permitiendo identificar y comprender las variables clave que impactan la evaluación educativa. Estos resultados orientarán de manera efectiva el diseño y desarrollo del aplicativo web, garantizando que los indicadores incorporados sean pertinentes, significativos y alineados con las necesidades específicas del contexto de este estudio.

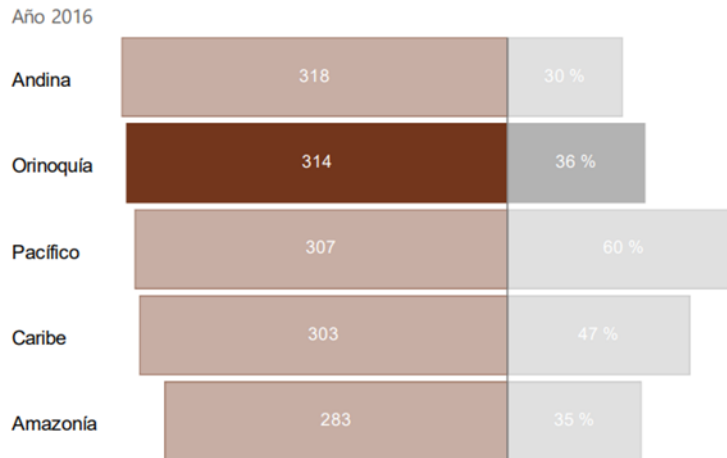
A continuación, se presentan los resultados de la investigación, los cuales se obtuvieron a partir del análisis de los datos recolectados en las pruebas saber que son representativas de estudiantes de educación básica y media en Colombia.

Los resultados se presentan de acuerdo con los indicadores de desempeño seleccionados para el estudio. Estos indicadores son los siguientes:

- Trabajo de los padres: Se refiere al tipo de ocupación de los padres de los estudiantes.
- Estrato socioeconómico: Se refiere al estrato socioeconómico de la vivienda de los estudiantes.
- Disponibilidad de computador: Se refiere a la disponibilidad de un computador en la vivienda de los estudiantes.
- Disponibilidad de internet: Se refiere a la disponibilidad de internet en la vivienda de los estudiantes.
- Educación de los padres: Se refiere al nivel educativo de los padres de los estudiantes.

VII.I. Resultados descriptivos
VII.I.I. Saber 3°
VII.I.I.I. Puntaje en lenguaje

Figura 71. Resultados puntaje en lenguaje saber 3° con nivel de pobreza año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

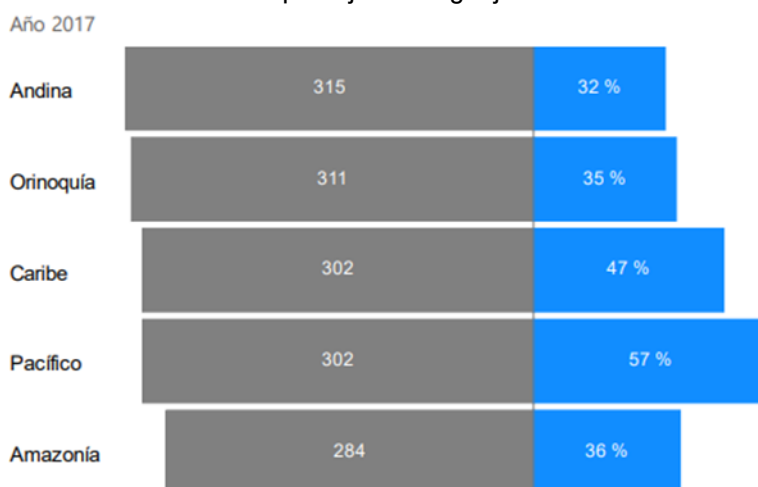
Los resultados para el año 2016 en las saber 3° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio en el área de lenguaje más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 318, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 283.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

Figura 72. Resultados puntaje en lenguaje saber 3° con nivel de pobreza año 2017



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el año 2017 en las saber 3° muestran un patrón similar ya que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio en el área de lenguaje más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 315, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 284.

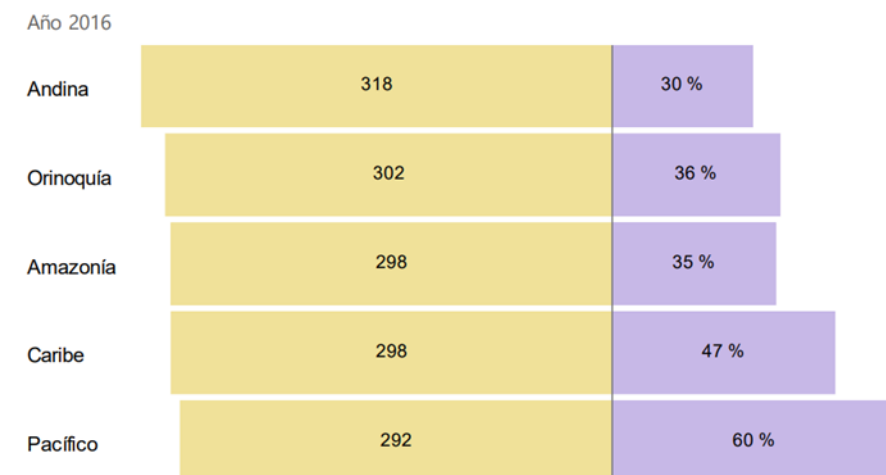
Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

VII.I.I.II. Puntaje matemáticas

Figura 73. Resultados puntaje en matemáticas saber 3° con nivel de pobreza año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

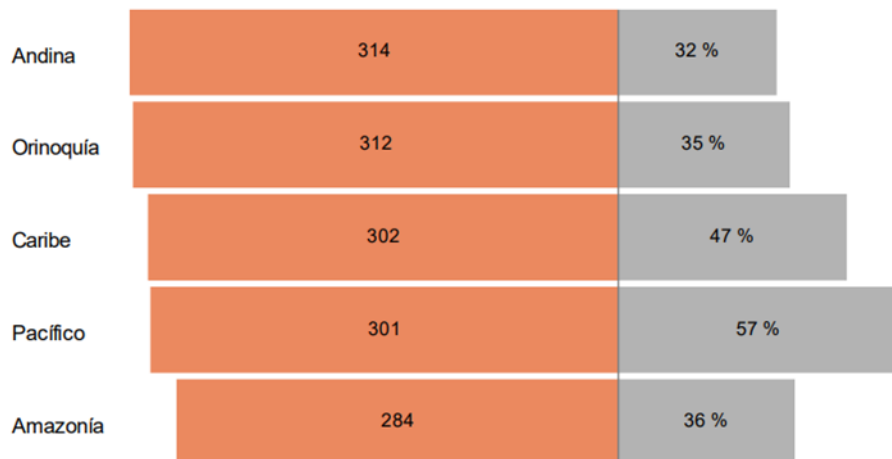
Los resultados para el área de matemáticas en el año 2016 en las saber 3° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 318, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 292.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

Figura 74. resultados puntaje en matemáticas saber 3° con nivel de pobreza año 2017
Año 2017



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de matemáticas en el año 2017 en las saber 3° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 314, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 284.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

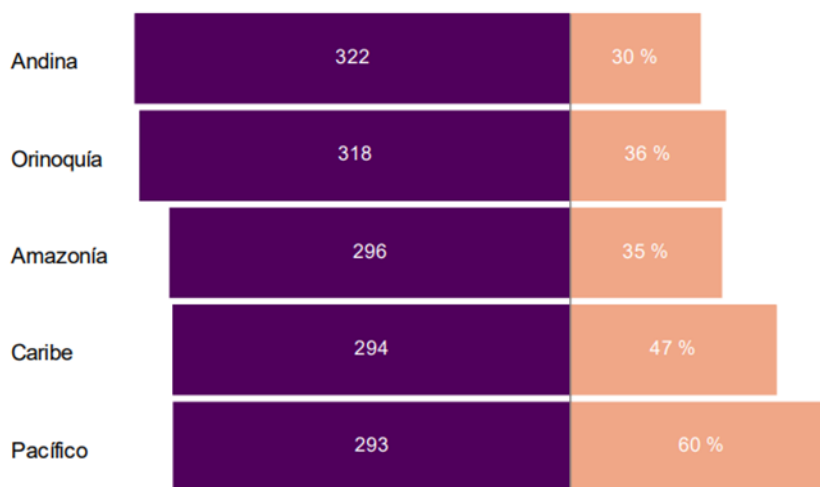
Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

VII.I.II. Saber 5

VII.I.II.I. Puntaje lenguaje

Figura 75. resultados puntaje lenguaje saber 5° con nivel de pobreza año 2016

Año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de lenguaje en el año 2016 en las saber 5° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 322, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 293.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

Figura 76. resultados puntaje lenguaje saber 5° con nivel de pobreza año 2017



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de lenguaje en el año 2017 en las saber 5° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 318, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 286.

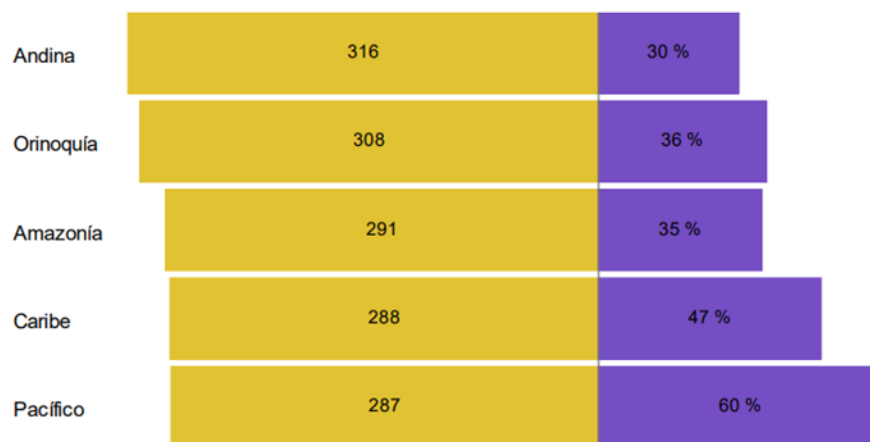
Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

VII.I.II.II. Puntaje matemáticas

Figura 77. Resultados puntaje matemáticas saber 5° con nivel de pobreza año 2016
Año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

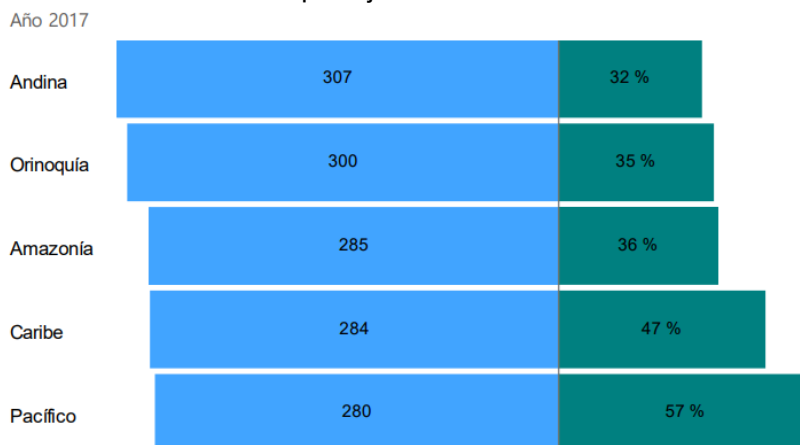
Los resultados para el área de matemáticas en el año 2016 en las saber 5° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 316, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 287.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

Figura 78. Resultados puntaje matemáticas saber 5° con nivel de pobreza año 2017



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de matemáticas en el año 2017 en las saber 5° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 307, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 280.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

VII.I.II.III.Puntaje ciencias naturales

Figura 79. Resultados ciencias naturales saber 5° con nivel de pobreza año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de ciencias naturales en el año 2016 en las saber 5° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 324, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 297.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

VII.I.III. Saber 9

VII.I.III.I. Puntaje lenguaje

Figura 80. Resultados lenguaje saber 9° con nivel de pobreza año 2016
Año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

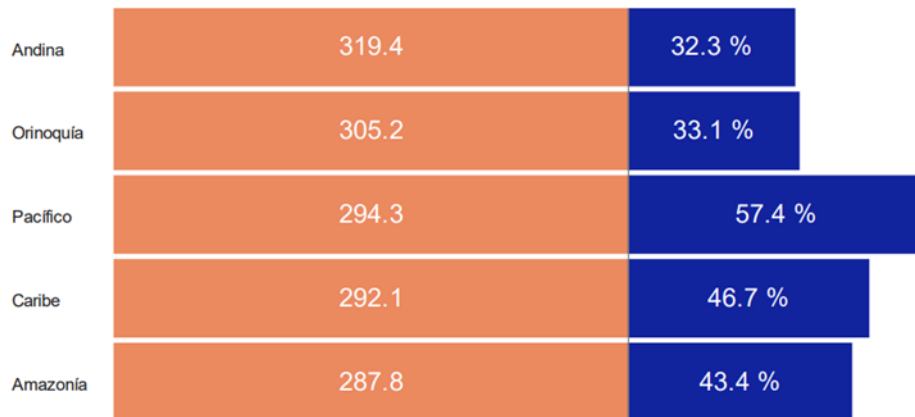
Los resultados para el área de lenguaje en el año 2016 en las saber 9° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 315, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 288.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

Figura 81. Resultados lenguaje saber 9° con nivel de pobreza año 2017
Año 2017



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de lenguaje en el año 2017 en las saber 9° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 319.4, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 287.8.

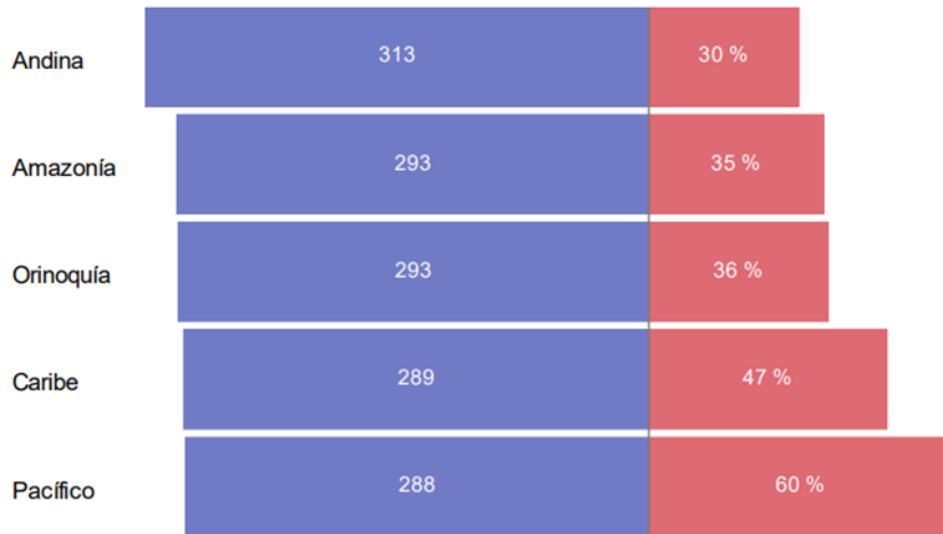
Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

VII.I.III.II.Puntaje matemáticas

Figura 82. Resultados matemáticas saber 9° con nivel de pobreza año 2016
Año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

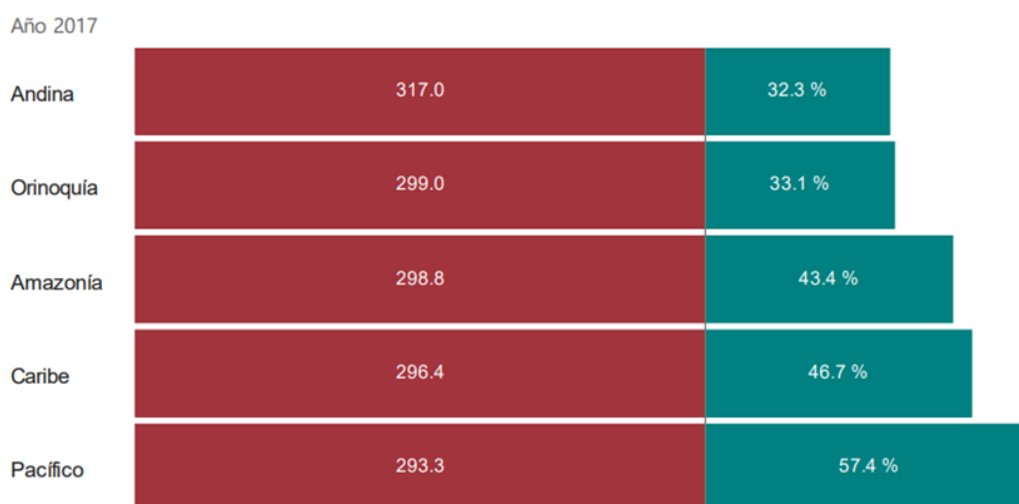
Los resultados para el área de matemáticas en el año 2016 en las saber 9° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 313, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 288.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

Figura 83. Resultados matemáticas saber 9° con nivel de pobreza año 2017



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de matemáticas en el año 2017 en las saber 9° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 317, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 293.3.

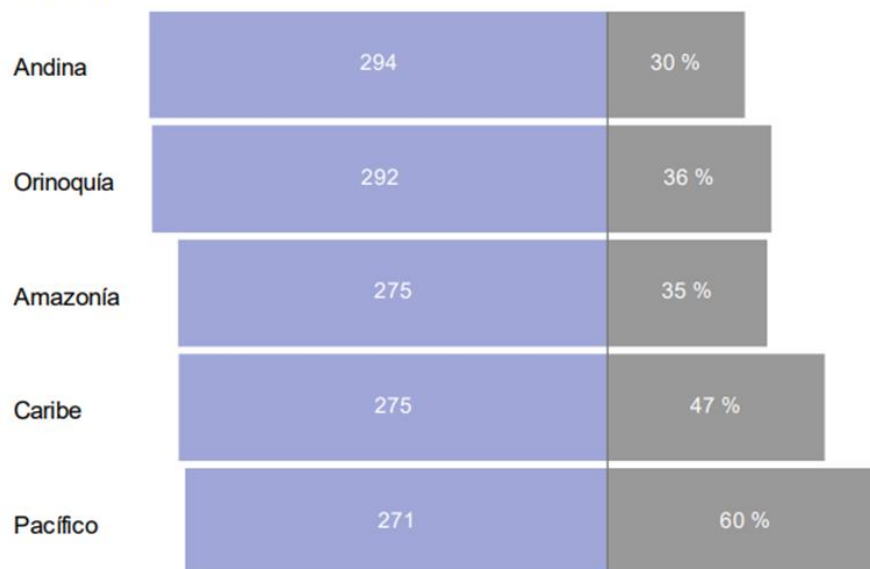
Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

VII.I.III.III. Puntaje ciencias naturales

Figura 84. Resultados ciencias naturales saber 9° con nivel de pobreza año 2016
Año 2016



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Los resultados para el área de ciencias naturales en el año 2016 en las saber 9° muestran que los estudiantes de los niveles de pobreza más alto obtuvieron los puntajes promedio más bajos. Los estudiantes con nivel de pobreza más bajo obtuvieron un puntaje promedio de 294, mientras que los estudiantes del nivel de pobreza más alto obtuvieron un puntaje promedio de 271.

Esta diferencia de puntaje es significativa y sugiere que la pobreza tiene un impacto negativo en el rendimiento académico de los estudiantes en matemáticas.

La imagen también muestra que hay diferencias regionales en los puntajes promedio en matemáticas. Los estudiantes de las regiones Andina y Orinoquía obtuvieron puntajes promedio más altos que los estudiantes de las regiones Caribe y Pacífica.

Estas diferencias regionales pueden deberse a una variedad de factores, como las diferencias en los recursos educativos, las diferencias en las políticas educativas y las diferencias en las culturas.

En lo que respecta a los puntajes en las áreas de lenguaje, matemáticas y ciencias naturales, se evidencia una relación significativa con el nivel de pobreza. Al analizar las cifras del año 2016, es notable que la región andina, caracterizada por tener un índice de pobreza más bajo, obtiene los puntajes promedio más altos en las pruebas de lenguaje. Este patrón se repite en el año 2017, donde nuevamente las regiones con un nivel de pobreza más bajo logran los mejores resultados. Estos hallazgos respaldan la noción de que el factor socioeconómico influye notablemente en el desempeño de los estudiantes en

estas áreas, subrayando la influencia del contexto económico en los resultados de las pruebas.

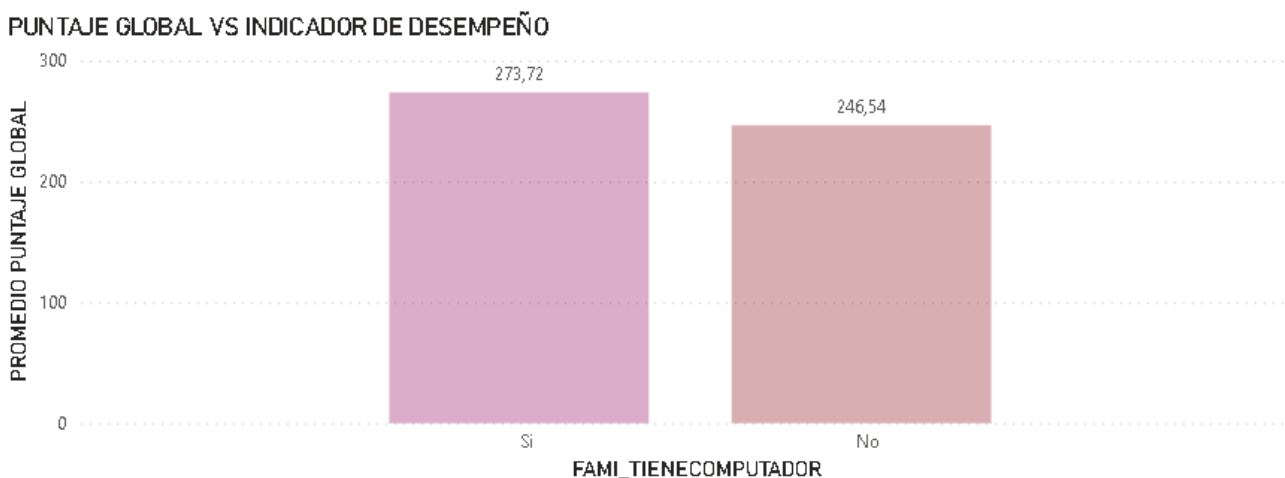
- La relación entre el nivel de pobreza por región y el rendimiento estudiantil en los grados 3, 5 y 9 es un hallazgo fundamental. Los datos demuestran que esta relación es significativa en el contexto de las Pruebas Saber en Colombia.
- Las regiones que enfrentan niveles más bajos de pobreza tienden a obtener puntajes promedio más altos en estas pruebas. Esto se traduce en un acceso más equitativo a una educación de calidad para los estudiantes en estas áreas, lo que se traduce en un mejor rendimiento académico.
- Por el contrario, las regiones con niveles de pobreza más altos pueden enfrentar desafíos adicionales en términos de acceso a recursos educativos, apoyo en el hogar y, a menudo, condiciones de vida más precarias. Estos factores se reflejan en puntajes promedio más bajos en las Pruebas Saber.

El resultado principal aquí es que el nivel de pobreza no solo es un indicador de las disparidades económicas en el país, sino que también se traduce en disparidades en las oportunidades educativas y, en última instancia, en el desempeño de los estudiantes. Esto resalta la necesidad apremiante de abordar la pobreza y sus efectos en el acceso a una educación de calidad para mejorar el rendimiento académico de los estudiantes en Colombia.

VII.I.IV. Saber 11

VII.I.IV.I. Indicador Acceso a computador

Figura 85. Puntaje global vs Indicador acceso a computador



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

La gráfica de puntaje global vs indicador de acceso a computador muestra que, en general, las familias que tienen acceso a un computador tienen un puntaje global más alto que las familias que no tienen acceso a un computador.

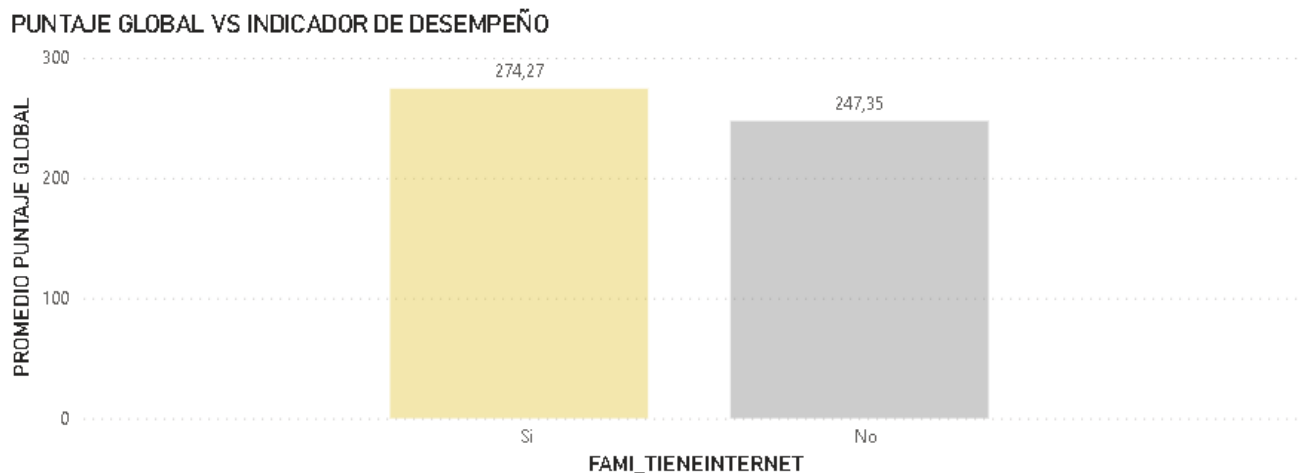
El promedio del puntaje global es de 273,72. Las familias que tienen acceso a un computador tienen un puntaje promedio de 296,54, mientras que las familias que no tienen acceso a un computador tienen un puntaje promedio de 246,54.

Esta diferencia se puede explicar por varias razones. En primer lugar, las familias que tienen acceso a un computador tienen más oportunidades de acceder a la información y los recursos educativos. Pueden usar el computador para investigar temas, aprender nuevas habilidades y conectarse con otros.

En segundo lugar, las familias que tienen acceso a un computador pueden utilizar este dispositivo para apoyar el aprendizaje de sus hijos. Pueden usar el computador para ayudar a sus hijos con las tareas escolares, proporcionarles acceso a contenido educativo y apoyar su desarrollo social y emocional.

VII.I.IV.II. Indicador Acceso a internet

Figura 86. Puntaje global vs Indicador acceso a internet



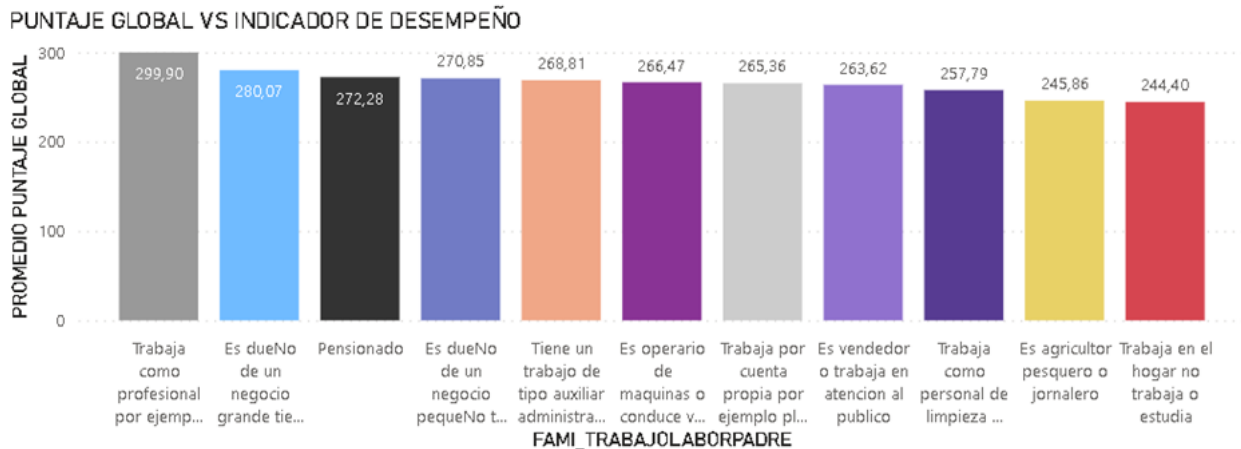
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El promedio de puntaje global para los estudiantes que no tienen acceso a internet es de 247,35. El promedio de puntaje global para los estudiantes que tienen internet es de 274,27.

Esta correlación positiva se puede explicar por el hecho de que el acceso a internet ofrece a las personas oportunidades que pueden mejorar su calidad de vida. Por ejemplo, el acceso a internet puede ayudar a las personas a obtener educación, encontrar herramientas para el aprendizaje, participar en la economía y mantenerse informadas sobre el mundo que las rodea.

VII.I.IV.III. Indicador educación padres

Figura 87. Puntaje global vs Indicador educación padres



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

La gráfica de puntaje global vs indicador de educación padres muestra una relación positiva entre ambos factores. En general, los estudiantes con padres con mayor nivel educativo tienden a obtener puntajes más altos en las pruebas estandarizadas.

Esta relación se puede explicar por varios factores. En primer lugar, los padres con mayor nivel educativo suelen tener más recursos y oportunidades para apoyar la educación de sus hijos. Por ejemplo, pueden proporcionarles libros, acceso a Internet y otros recursos educativos. Además, pueden tener más conocimientos y habilidades para ayudar a sus hijos con sus tareas escolares.

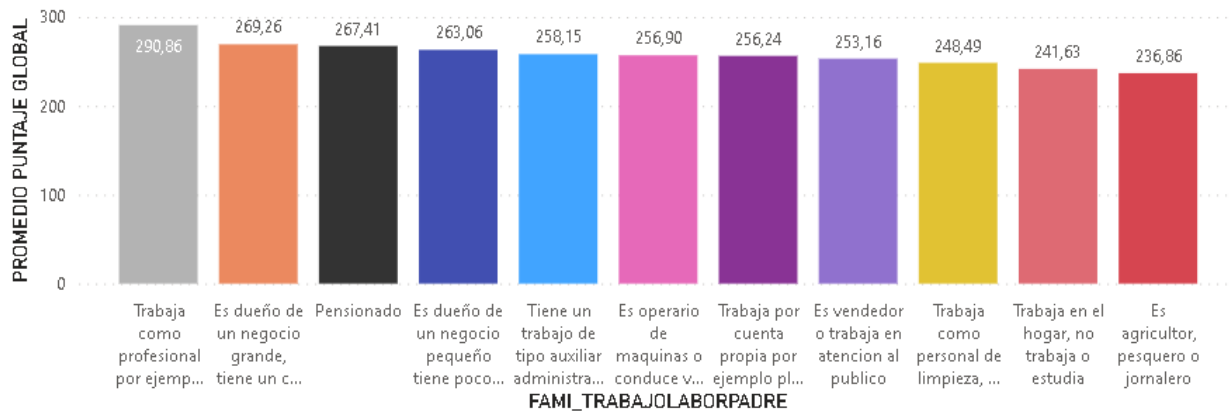
En segundo lugar, los padres con mayor nivel educativo suelen tener expectativas más altas para sus hijos. Esto puede motivar a los hijos a esforzarse más en la escuela.

En tercer lugar, los padres con mayor nivel educativo suelen tener un estilo de crianza más positivo. Esto puede ayudar a los hijos a desarrollar habilidades cognitivas y socioemocionales que son importantes para el éxito en la escuela.

VII.I.IV.IV. Indicador trabajo padres

Figura 88. Puntaje global vs Indicador trabajo padres

PUNTAJE GLOBAL VS INDICADOR DE DESEMPEÑO



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

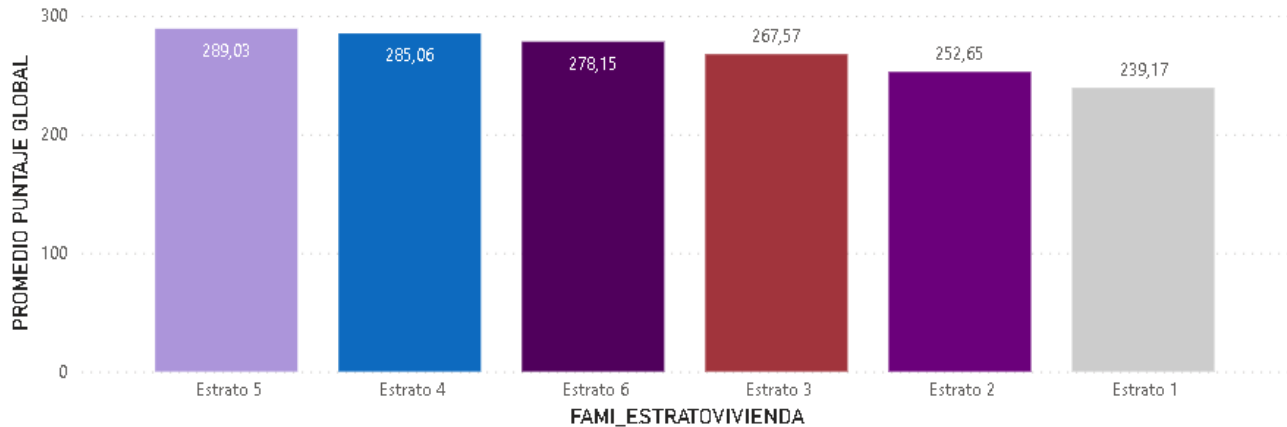
Los estudiantes con padres que tienen un trabajo como profesionales tienen un puntaje promedio más alto en las pruebas Saber que los estudiantes con padres que tienen otros tipos de trabajo. El puntaje promedio de los estudiantes con padres profesionales es de 290.86, mientras que el puntaje promedio de los estudiantes con padres que tienen otros tipos de trabajo tiende a ser menor.

Este resultado sugiere que el trabajo de los padres tiene un impacto significativo en el rendimiento académico de los hijos. Los padres con roles profesionales tienen más probabilidades de proporcionar a sus hijos un ambiente familiar propicio para el aprendizaje, así como de brindarles apoyo y motivación para que alcancen su máximo potencial.

VII.I.IV.V. Indicador estrato vivienda

Figura 89. Puntaje global vs Indicador estrato vivienda

PUNTAJE GLOBAL VS INDICADOR DE DESEMPEÑO



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

La gráfica de puntaje global vs estrato vivienda muestra que, en general, los hogares de estratos más altos tienen un puntaje global más alto que los hogares de estratos más bajos. Este resultado es consistente con la teoría de la desigualdad, que sostiene que los hogares con mayor nivel de ingresos y educación tienen más oportunidades para desarrollar su potencial y alcanzar un mejor nivel de vida.

En particular, la gráfica muestra que el puntaje global promedio de los hogares de estrato 5 es de 289,03, mientras que el de los hogares de estrato 1 es de 239,17. Esto significa que los hogares de estrato 5 tienen un puntaje global 20,86 puntos más alto que los hogares de estrato 1.

Esta diferencia de puntaje global se explica por una serie de factores, incluyendo:

- El nivel de ingresos: Los hogares de estratos más altos tienen ingresos más altos, lo que les permite acceder a mejores oportunidades educativas y laborales.
- El nivel de educación: Los hogares de estratos más altos tienen un mayor nivel de educación, lo que les permite desarrollar sus habilidades y conocimientos.
- El acceso a servicios básicos: Los hogares de estratos más altos tienen un mayor acceso a servicios básicos, como agua potable, electricidad y alcantarillado.

En el nivel de educación media (grado 11), se exploraron varios indicadores, incluyendo el acceso a internet, acceso a computadoras, educación de los padres y el trabajo de los padres, para analizar su influencia en el rendimiento de los estudiantes en las Pruebas Saber.

Los resultados del análisis revelaron la importancia significativa del acceso a internet y computadoras en el desempeño de los estudiantes. Las regiones con un mejor acceso a

estas tecnologías presentaron puntajes más elevados en las Pruebas Saber, subrayando la relevancia de la infraestructura tecnológica en la educación media.

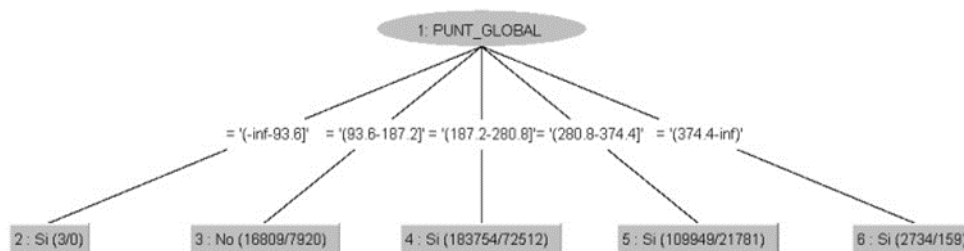
Además, se confirmó que la educación de los padres sigue siendo un factor crucial en el éxito académico de los estudiantes en esta etapa. Aquellos cuyos padres tienen un mayor nivel educativo obtuvieron resultados superiores en las Pruebas Saber, destacando la influencia de la educación parental en el logro estudiantil.

También se evidenció que el trabajo de los padres, particularmente en términos de ingresos y estabilidad laboral, está intrínsecamente relacionado con el rendimiento de los estudiantes en la educación media. Los padres que disfrutaban de empleos mejor remunerados y estables suelen tener hijos con un desempeño académico más destacado. Estos hallazgos subrayan la relevancia de la situación laboral de los padres en la educación de sus hijos y proporcionan información valiosa para la formulación de políticas destinadas a mejorar el rendimiento estudiantil en esta etapa.

VII.II. Resultados predictivos saber once

VII.II.I. Indicador acceso a computador

Figura 90. Árbol j48 puntaje global vs indicador acceso a computador



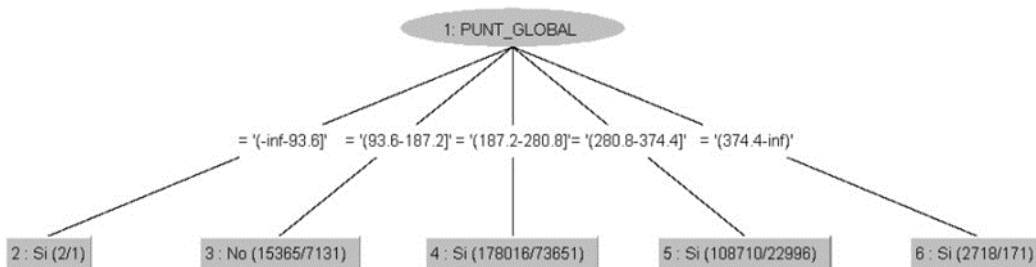
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El árbol de decisión J48 que se muestra en la imagen es un modelo de aprendizaje automático que se utiliza para predecir el rendimiento de los estudiantes en las pruebas Saber, una evaluación estandarizada de la educación en Colombia. El árbol utiliza dos características para hacer su predicción: el puntaje global de las pruebas Saber y el indicador de acceso a la computadora.

El puntaje global de las pruebas Saber es una medida del rendimiento académico del estudiante. Se calcula a partir de las puntuaciones de las pruebas en las áreas de lectura, matemáticas, ciencias y lenguaje. El indicador de acceso a la computadora es una medida de la disponibilidad de una computadora en el hogar del estudiante.

VII.II.II. Indicador acceso a internet

Figura 91. Árbol j48 puntaje global Vs indicador acceso a internet



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

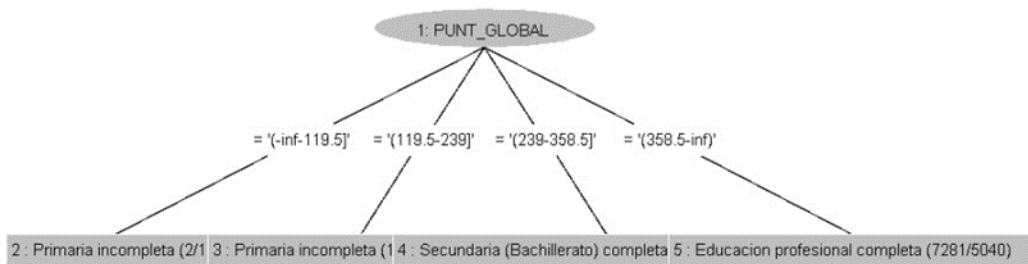
Se utiliza el árbol j48 para predecir el puntaje en función de si una persona tiene acceso o no a Internet, se puede observar que las personas con acceso a Internet tienen un puntaje global promedio más alto que las personas sin acceso a Internet.

Esto se debe a que las personas con acceso a Internet suelen tener un mayor nivel de educación y de ingresos, lo que les brinda más oportunidades para desarrollar sus habilidades y conocimientos.

En general, el árbol j48 muestra que las personas con acceso a Internet tienen más probabilidades de tener un puntaje global alto.

VII.II.III. Indicador educación padres

Figura 92. Árbol j48 puntaje global Vs indicador educación padres



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El comportamiento del Árbol j48 puntaje global vs indicador educación padres se puede analizar de la siguiente manera:

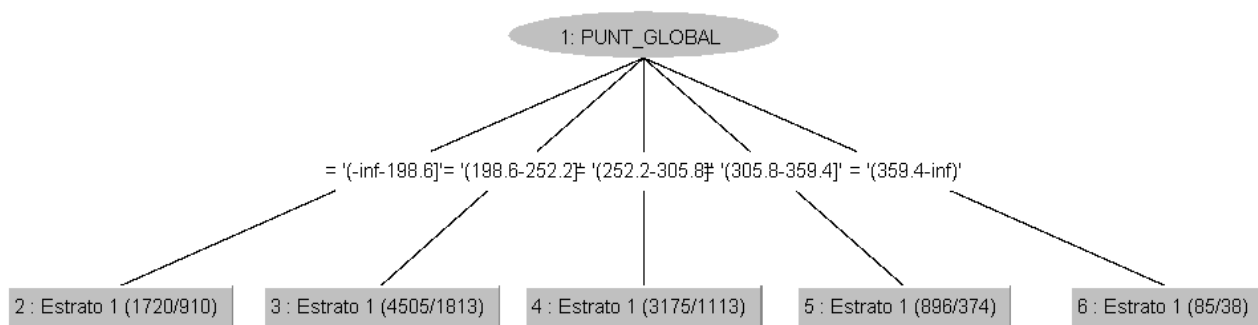
- En general, el puntaje global es mayor que el indicador de educación de los padres. Esto significa que los estudiantes con padres con mayor nivel educativo tienden a tener un mejor rendimiento académico.
- La diferencia entre el puntaje global y el indicador de educación de los padres es mayor en los niveles de educación más altos. Por ejemplo, los estudiantes con padres con educación profesional completa tienen un puntaje global promedio de 358.5, mientras que los estudiantes con padres con educación primaria incompleta tienen un puntaje global promedio de 119.5.

Estos resultados sugieren que la educación de los padres es un factor importante que contribuye al rendimiento académico de los estudiantes. Sin embargo, la influencia de la educación de los padres es mayor en los niveles de educación más altos.

Una posible explicación para esta diferencia es que los padres con mayor nivel educativo tienen más recursos y oportunidades para apoyar el aprendizaje de sus hijos. Por ejemplo, pueden proporcionarles libros, acceso a la tecnología y oportunidades de enriquecimiento extracurricular. Además, los padres con mayor nivel educativo suelen tener expectativas más altas para sus hijos y pueden brindarles un mejor apoyo emocional y académico.

VII.II.IV. Indicador estrato vivienda

Figura 93. Árbol j48 puntaje global Vs indicador estrato vivienda



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El gráfico muestra el comportamiento del Árbol j48 puntaje global vs indicador estrato vivienda.

Se evidencia en el comportamiento que los estudiantes con puntajes globales más altos tienden a provenir de estratos de vivienda más altos. Esto significa que los estudiantes de familias más adineradas tienden a tener un mejor rendimiento académico que los estudiantes de familias de bajos recursos.

Hay varias posibles explicaciones para esta relación. Una posibilidad es que los estudiantes de familias más adineradas tengan acceso a mejores recursos educativos, como clases particulares, tutorías y escuelas privadas. Otra posibilidad es que los

estudiantes de familias con más recursos tengan expectativas más altas de sus padres, lo que puede motivarlos a trabajar más duro en la escuela.

1. Acceso a Internet: El acceso a Internet se ha convertido en un componente crucial en la vida de los estudiantes. Los datos muestran que los estudiantes con acceso a Internet tienden a tener una ventaja en su rendimiento en las Pruebas Saber 11. Esto puede atribuirse a la capacidad de acceder a recursos educativos en línea, investigar y mantenerse actualizados sobre temas relevantes.
2. Acceso a Computadoras: El acceso a computadoras es otro factor importante. Los estudiantes con acceso a computadoras pueden beneficiarse al realizar investigaciones, escribir y presentar proyectos de manera efectiva, lo que puede impactar positivamente sus puntajes en las pruebas. La tecnología también les permite desarrollar habilidades digitales, que son cada vez más valiosas en el mundo actual.
3. Educación de los Padres: La educación de los padres ha demostrado ser un indicador sólido de éxito estudiantil. Los padres con un mayor nivel educativo tienden a inculcar una cultura de aprendizaje en sus hijos, brindar apoyo académico y comprender mejor la importancia de la educación. Esto se refleja en los resultados de las Pruebas Saber 11, donde los estudiantes con padres más educados tienden a obtener puntajes más altos.

Relación con el Rendimiento en las Pruebas Saber 11:

El análisis revela que existe una relación significativa entre estos indicadores y el rendimiento de los estudiantes en las Pruebas Saber 11. Los resultados muestran que el acceso a Internet y computadoras, junto con la educación de los padres, son factores que impactan directamente en el desempeño estudiantil.

Los estudiantes que tienen acceso a Internet y computadoras en sus hogares o escuelas tienden a superar a sus pares que carecen de estas ventajas tecnológicas. Esto les permite realizar investigaciones de manera más eficiente, mejorar sus habilidades de comunicación y presentación, y acceder a recursos en línea que respaldan su aprendizaje. Además, el nivel de educación de los padres sigue siendo un factor determinante en el éxito académico de los estudiantes. Los estudiantes cuyos padres han alcanzado un mayor nivel educativo generalmente obtienen puntajes más altos en las Pruebas Saber 11. Esto se debe a la influencia positiva que los padres educados pueden ejercer en sus hijos, al brindar apoyo académico y fomentar un ambiente de aprendizaje en el hogar.

Referencias

- Adebayo, K. A., Ntokozo, N., & Grace, N. Z. (2020). Availability of Educational Resources and Student Academic Performances in South Africa. *Universal Journal of Educational Research*, 8(8), 3768-3781. doi: 10.13189/ujer.2020.080858
- Así ha afectado el Covid-19 la educación en Colombia - Forbes Colombia. (n.d.). Retrieved from <https://forbes.co/2020/04/30/actualidad/asi-ha-afectado-el-covid-19-la-educacion-en-colombia/>
- Bester, G., & Kuyper, N. (2020). The Influence of Additional Educational Support on Poverty-Stricken Adolescents' Resilience and Academic Performance. <https://doi.org/10.1080/18146627.2019.1689149>, 17(3), 158-174. doi: 10.1080/18146627.2019.1689149
- Devi, K., Ratnoo, S., & Bajaj, A. (2022). Impact of Socio-Economic Factors on Students' Academic Performance: A Case Study of Jawahar Navodaya Vidyalaya. *Lecture Notes in Networks and Systems*, 419 LNNS, 774-785. doi: 10.1007/978-3-030-96299-9_73
- Efecto de la pandemia sobre el sistema educativo: El caso de Colombia | Banco de la República. (n.d.). Retrieved from <https://www.banrep.gov.co/es/borrador-1179>
- Jiang, P. (2021). Gender differences in mathematics academic performance of high school students in western China. *Journal of Physics: Conference Series*, 1978(1). doi: 10.1088/1742-6596/1978/1/012038
- Kotomina, O. v., & Sazhina, A. I. (2021). The Influence of Family Factors on the Academic Performance of Schoolchildren and University Students: Review of Foreign Studies. *Education and Self Development*, 16(4), 74-92. doi: 10.26907/esd.16.4.07
- La participación de los padres en la enseñanza – Observatorio | Instituto para el Futuro de la Educación. (n.d.). Retrieved from <https://observatorio.tec.mx/edu-news/la-importancia-de-la-participacion-de-los-padres-en-la-educacion>
- Muelle, L. (2020). Socioeconomic and contextual factors associated with low academic performance of Peruvian students in PISA 2015. *Apuntes*, 47(86), 111-146. doi: 10.21678/APUNTES.86.943

- Rocha, J. L. M., Zela, M. A. C., Torres, N. I. V., Rojas, J. T. R., Valderrama, S. O. M., & Medina, G. S. (2021). Virtual Education: Impact of Socio-emotional and Pedagogical Factors on Academic Performance based on Neural Networks and Stepwise Regression. Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology, 2021-July. doi: 10.18687/LACCEI2021.1.1.459
- Rodríguez, J. V., Rodado, D. N., Crissien Borrero, T., & Parody, A. (2022). Multidimensional indicator to measure quality in education. *International Journal of Educational Development*, 89. doi: 10.1016/j.ijedudev.2021.102541
- Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, 2. doi: 10.1016/j.caeai.2021.100018
- Rodríguez-Rodríguez, D., & Guzmán, R. (2019). Socio-familial risk factors and personal protective variables of academic performance in Secondary Education students. *Psicothema*, 31(2), 142-148. doi: 10.7334/psicothema2018.213
- Rodríguez-Rodríguez, D., & Guzmán, R. (2021). Academic performance of secondary education students in socio-familial risk contexts. *Suma Psicológica*, 28(2), 104-111. doi: 10.14349/sumapsi.2021.v28.n2.5
- Rodríguez Diego, Ordoñez Ruber, Hidalgo Mario. (2021). Academic performance determinants of high school students in the Department of Nariño, Colombia. *Lecturas de Economía*, 94, 87-126. doi: 10.17533/UDEA.LE.N94A341834
- Sarmiento Espinel, J. A., Silva Arias, A. C., & van Gameraen, E. (2019). Evolution of the inequality of educational opportunities from secondary education to university. *International Journal of Educational Development*, 66, 193-202. doi: 10.1016/j.ijedudev.2018.09.006
- Wang, Y., Xia, M., Guo, W., Xu, F., & Zhao, Y. (2022). Academic performance under COVID-19: The role of online learning readiness and emotional competence. *Current Psychology*. doi: 10.1007/s12144-022-02699-7
- Zeng, L., & Luo, H. (2021). Online Academic Performance during the COVID-19: Evidence from a Rural High School in Western China.

- Regresan las Pruebas Saber 3°, 5°, 7° y 9° para cerca de 200 mil estudiantes, de 1.300 sedes educativas de todo el país. (2022, 8 abril). Icfes. Retrieved from <https://bit.ly/38y4JWw>
- Nuankaew, P., & Sararat, W. (2022). Student Performance Prediction Model for Predicting Academic Achievement of High School Students. *European Journal of Educational Research*, 11(2), 949-963.
- Matzavela, V., & Alepis, E. (2021, 5 octubre). E-Biblioteca Ucundinamarca. ScienceDirect. Retrieved from <https://login.ucundinamarca.basesdedatosezproxy.com/login?qurl=https://www.sciencedirect.com%2fscience%2farticle%2fpii%2fS2666920X21000291%3fvia%253Dihub>
- Content, R. R. (2019, May 6). *¿Qué es la usabilidad web? Conoce este concepto y por qué es importante para la estrategia digital de las empresas.* Rock Content - ES; Rock Content. <https://rockcontent.com/es/blog/usabilidad/>
- *Selenium y la automatización de las pruebas.* (2020). Juntadeandalucia.es. from <https://www.juntadeandalucia.es/servicios/madeja/contenido/recurso/381>
- Testing con Cucumber: Cómo afianzar la fiabilidad en los desarrollos. (2020, August 10). Chakray. <https://www.chakray.com/es/testing-cucumber-afianzar-fiabilidad-desarrollos/>
- Mitra S, Acharya T. *Data Mining. Multimedia, Soft Computing, and Bioinformatics.* John Wiley & Sons, Inc., Hoboken, New Jersey; 2013.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann Publishers.
- Rud, O. P. (2012). *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management* (1.a ed.). John Wiley & Sons Inc.
- Jin, Y., & Yang, X. (2021). Educational Data Mining: Discovering Principal Factors for Better Academic Performance. 2021 the 3rd International Conference on Big Data Engineering and Technology (BDET).
- Hussain, S. (2018, 1 febrero). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA | Hussain | Indonesian Journal of Electrical Engineering and Computer Science. Indonesian Journal.

- Alam, T. M., Mushtaq, M., & Shaukat, K. (2021). A Novel Method for Performance Measurement of Public Educational Institutions Using Machine Learning Models. *Applied Sciences*, 11(19), 9296.
- Samsudin, N. A. M. (2021). Modeling Student's Academic Performance During Covid-19 Based on Classification in Support Vector Machine. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(5), 1798-1804.
- Han, J. and Kamber, M. "Data Mining: Concepts and Techniques. Second Edition", Morgan Kaufmann Publishers, San Francisco, 2010.
- Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.
- Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. (2020). Academic performance patterns of middle school students in the knowledge natural science test 11 with decision trees. *RISTI - Revista De Sistemas e Tecnologias De Informacao*, 2020(E32), 190-201.
- Pandey, U.K. and Pal, S., 2011. Data Mining: A prediction of performer or underperformer using classification. (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, Vol. 2 (2), 2011, 686- 690.
- Ajibade, S. S. M., & Bahiah Binti Ahmad, N. (2019). Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods. *IOP Conference Series: Materials Science and Engineering*, 551(1), 012061.
- Subaira A. 2016 IEEE 8th Int. Conf. on Intelligent System and Control (ISCO) 978 274-280
- Sabzevari M. 2018 Cornell Uni. arXiv preprint arXiv:1802.07877
- Amrieh E. A. 2017 Int'l. Journal of Database Theory and Application 9 119-136.
- Moisa V. 2018 Journal of Mobile Embedded and Distributed Systems 5 70-77
- Contreras Bravo, L. E., Fuentes López, H. J., & Rivas Trujillo, E. (2022). Análisis del rendimiento académico mediante técnicas de aprendizaje automático con métodos de ensamble. *Revista Boletín Redipe*, 10(13), 171-190.

- Fundación iS+D La finalidad de la metodología cuantitativa. (2018). Retrieved from <https://isdfundacion.org/2018/11/01/finalidad-metodologia-cuantitativa/>
- Modelos de minería de datos - Acervo Lima. (2020). Retrieved from <https://es.acervolima.com/modelos-de-mineria-de-datos/>
- Metodología Scrum: cómo aplicar el método Scrum | APD. (2022). Retrieved from <https://www.apd.es/metodologia-scrum-que-es>
- Más de 600.000 estudiantes presentan las pruebas Icfes en todo el país. (n.d.). Retrieved May 8, 2022, from <https://www.larepublica.co/economia/mas-de-600000-estudiantes-presentan-las-pruebas-icfes-en-todo-el-pais-2759151>
- Regresan las Pruebas Saber 3°, 5°, 7° y 9° para cerca de 200 mil estudiantes, de 1.300 sedes educativas de todo el país. (n.d.). Retrieved May 8, 2022, from <https://www.mineducacion.gov.co/portal/salaprensa/Noticias/410085:Regresan-las-Pruebas-Saber-3-5-7-y-9-para-cerca-de-200-mil-estudiantes-de-1-300-sedes-educativas-de-todo-el-pais>
- El ODS 4: El objetivo de desarrollo sostenible de la educación | Compartir Palabra maestra. (2020). Retrieved May 11, 2022, from <https://www.compartirpalabramaestra.org/actualidad/articulos-informativos/el-ods-4-el-objetivo-de-desarrollo-sostenible-de-la-educacion>
- Objetivo 4: Educación de calidad | Sustainable Development Goals Fund. (n.d.). Retrieved May 11, 2022, from <https://www.sdgfund.org/es/objetivo-4-educaci%C3%B3n-de-calidad>
- Cabeza María Alejandra. (2004). INDICADORES DE GESTIÓN EN LA EDUCACIÓN SUPERIOR COMO HERRAMIENTA DE LA PLANIFICACIÓN ESTRATÉGICA. <http://www.redalyc.org/articulo.oa?id=36410206>
- Bai, X., Zhang, F., Li, J., Guo, T., Aziz, A., Jin, A., & Xia, F. (2021). Educational Big Data: Predictions, Applications and Challenges. *Big Data Research*, 26, 100270. <https://doi.org/10.1016/j.bdr.2021.100270>
- Mitra, S., & Acharya, Tinku. (2003). *Data mining: multimedia, soft computing, and bioinformatics*. 401. <https://www.wiley.com/en-us/Data+Mining%3A+Multimedia%2C+Soft+Computing%2C+and+Bioinformatics-p-9780471460541>
- Rodríguez Suárez, Y., & Amador, A. D. (2009). *Herramientas de Minería de Datos Data Mining Tools*. 3(3), 73-80.

- Sofía J. Vallejos y Mgter. David Luis la Red Martínez. Minería de Datos. Universidad Nacional del Nordeste Facultad de Ciencias Exactas, Naturales y Agrimensura, (2006).
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., CRISP-DM 1.0: Step-by-step data mining guide: CRISP-DM Consortium., 2009.
- Sunita, B., & Lobo, Mr. (2011). Data Mining in Educational System using WEKA. Data Mining in Educational System using WEKA, 20-25.
- Witten, I. H., Frank, E., & Hall, M. (2005). Data mining: Practical machine learning tools and techniques. Miami: Elsevier
- Thierry Groussard. (n.d.). *JAVA 7: Los fundamentos del lenguaje Java - Thierry Groussard - Google Libros*. Retrieved August 17, 2022, from https://books.google.es/books?hl=es&lr=lang_es&id=JaPTzKZxbN4C&oi=fnd&pg=PA165&dq=JAVA&ots=pW5Bwhvk2b&sig=r0C1scygJio3-HS2oHIMSxFBhJQ#v=onepage&q=JAVA&f=false
- *ADMINISTRACIÓN DE BASES DE DATOS CON POSTGRESQL - Mariuxi Paola Zea Ordóñez, Jimmy Rolando Molina Ríos, Fausto Fabían Redrován Castillo - Google Libros*. (n.d.). Retrieved August 17, 2022, from https://books.google.com.co/books?id=5-mkDgAAQBAJ&pg=PA79&dq=postgresql&hl=es&sa=X&ved=2ahUKEwjm_Kf59dH5AhU8fjABHaTgBToQ6AF6BAgBEAI#v=onepage&q=postgresql&f=false
- Barrientos Rocio. (2009). *Árboles de decisión como herramienta en el diagnóstico médico*. <https://1library.co/document/zkw51d4m-%C3%A1rboles-decisi%C3%B3n-herramienta-diagn%C3%B3stico-m%C3%A9dico.html>
- Cobos, C., Zuñiga, J., Guarín, J., León, E., Mendoza, M., & English, I. (2010). CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos.
- *Ingeniería e Investigación*, 30(3), 45-56. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092010000300004&lng=en&nrm=iso&tlng=es
- Vanesa Berlanga Silvente, & María José Rubio Hurtado. (2013). *Cómo aplicar árboles de decisión en SPSS*. <http://diposit.ub.edu/dspace/bitstream/2445/43762/1/618361.pdf>
- Verma, M., Srivastava, M., Chack, N., Kumar Diswar, A., & Gupta, N. (2012). A Comparative Study of Various Clustering Algorithms in Data Mining. In *International Journal of Engineering Research and Applications (IJERA)* (Vol. 2). www.ijera.com
- Fulgencio Cánovas García, Francisco Alonso Sarría, & Francisco Gomariz Castillo. (2016). *MODIFICACIÓN DEL ALGORITMO RANDOM FOREST PARA SU EMPLEO EN CLASIFICACIÓN DE IMÁGENES DE TELEDETECCIÓN*.
- Patiño Pérez, D., Silva Bustillos, R., Munive Mora, C., & Botto-Tobar, M. (2020). Predicción de Covid19 con el uso del Algoritmo Random Forest y Redes Neuronales

- Artificiales. *Ecuadorian Science Journal*, 4(2), 101-110. <https://doi.org/10.46480/esj.4.2.41>
- Java. (s. f.). Home de DesarrolloWeb.com. <https://desarrolloweb.com/home/java>
 - ¿Qué es PostgreSQL y por qué llevarlo a Cloud? | Arsys. (s. f.). Blog de arsys.es. <https://www.arsys.es/blog/postgresql-servidores#:~:text=PostgreSQL%20es%20un%20sistema%20de,consistente%20y%20tolerante%20a%20fallos.>
 - Khoirom, S., Sonia, M., Laikhuram, B., Laishram, J., & Singh, D. (2020). Comparative Analysis of Python and Java for Beginners Cite this paper Comparative Analysis of Python and Java for Beginners. *International Research Journal of Engineering and Technology*. www.irjet.net
 - Rocha, Á., Adeli, H., Dzemyda, G., Moreira, F. y Ramalho Correia, A. M., eds., (2021). *Trends and Applications in Information Systems and Technologies* [en línea]. Cham: Springer International Publishing. [Consultado el 26 de agosto de 2022]. Disponible en: doi: 10.1007/978-3-030-72651-5
 - PostgreSQL vs Oracle: Critical Differences - Learn | Hevo [en línea]. (sin fecha). *Learn | Hevo*. [Consultado el 26 de agosto de 2022]. Disponible en: <https://hevodata.com/learn/postgresql-vs-oracle/#support>
 - kmeans [en línea]. (sin fecha). *Universidad de Oviedo - Universidad de Oviedo, la universidad de Asturias*. [Consultado el 26 de agosto de 2022]. Disponible en: https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html#:~:text=K-means%20es%20un%20algoritmo,suele%20usar%20la%20distancia%20cuadrática.
 - Gómez Moreno Antonio Steeven, Vínces Sánchez Luis Fernando, & Honores Tapia Joofre Antonio. (2018). *Comparacion De Tendencias Tecnologicas En Aplicaciones*.
 - Cambi alvarado Jessica, & Zuñiga garcia Jessica. (2016). *Comparacion de diferentes bases de datos*.
 - Ley 1266 de 2008, (2008). http://www.secretariassenado.gov.co/senado/basedoc/ley_1266_2008.html
 - LEY NÚMERO 23 DE 1982, (1982).
 - Acero Caro B. L., & Casas M., A. F. (2013). Análisis de las diferencias de género en el desempeño de estudiantes colombianos en matemáticas y lenguaje.”



UDECA
UNIVERSIDAD DE
CUNDINAMARCA



DESARROLLO DE SOFTWARE UNIVERSIDAD DE CUNDINAMARCA

www.unicundi.edu.co
unicundi@mail.unicundi.edu.co
Línea gratuita 018000 976000



Dirección de Sistemas y Tecnología
sistemasytecnologia@mail.unicundi.edu.co
PBX: 828 14 83 Ext. 110-170
Sede Fusagasugá

Contenido

| | | |
|----------------|--|----|
| I. | USUARIOS..... | 4 |
| I.I. | ADMINISTRADOR..... | 4 |
| I.II. | USUARIO..... | 4 |
| II. | REQUISITOS DE SOFTWARE Y HARDWARE..... | 4 |
| II.I. | REQUERIMIENTOS DE SOFTWARE..... | 4 |
| III. | APLICATIVO..... | 4 |
| III.I. | ICONOS GENERALES..... | 4 |
| III.II. | INGRESO AL APLICATIVO..... | 4 |
| III.II.I. | INICIO DE SESIÓN..... | 5 |
| Figura 2. | Inicio de session..... | 5 |
| III.II.II. | REGISTRARSE..... | 6 |
| Figura 3. | Pagina de registro..... | 6 |
| III.II.III. | INICO DE SESION ADMINISTRADOR..... | 7 |
| Figura 4. | Login administrador..... | 7 |
| III.II.IV. | MODULO ADMINISTADOR..... | 7 |
| III.II.IV.I. | PAGINA PRINCIPAL MODULO ADMINISTRADOR..... | 8 |
| Figura 5. | CRUD de usuarios..... | 8 |
| Figura 6. | Caracteriticas..... | 8 |
| Figura 7. | Opciones del CRUD de usuarios..... | 9 |
| III.II.IV.II. | MODULO EDITAR USUARIOS..... | 9 |
| Figura 8. | Pagina editar usuarios..... | 9 |
| III.II.IV.III. | MODULO ELIMINAR USUARIOS..... | 10 |
| III.II.IV.IV. | MODULO CREAR USUARIOS..... | 10 |
| Figura 9. | Pagina crear usuarios..... | 10 |
| III.II.IV.V. | MODULO LISTADO DE PRUEBAS..... | 10 |
| Figura 10. | Listados de pruebas/resultados..... | 11 |
| III.II.IV.VI. | MODULO LIMPIEZA DE DATOS..... | 11 |
| Figura 11. | Pagina para la limpieza de datos..... | 11 |
| Figura 12. | Resultado de la limpieza de datos..... | 12 |
| III.II.IV.VII. | MODULO DE CARGA DE REPORTES..... | 12 |
| Figura 13. | Carga de reportes..... | 12 |
| Figura 14. | Tipo de analisis..... | 13 |

| | | |
|--------------|--|----|
| Figura 15. | Selección de años analizados..... | 13 |
| Figura 16. | Regiones colombianas analizadas..... | 14 |
| Figura 17. | Selección del reporte correspondiente..... | 14 |
| III.II.V. | MODULO USUARIOS | 14 |
| Figura 18. | Página principal modulo Usuario..... | 15 |
| III.II.V.I. | PRUEBAS ONCE..... | 15 |
| Figura 19. | Menu de tipo de análisis | 15 |
| Figura 20. | Obtener resultados por reportes | 16 |
| Figura 21. | Ejemplo de reporte | 16 |
| III.II.V.II. | PRUEBAS TERCERO, QUINTO Y NOVENO..... | 17 |
| Figura 22. | Menu opciones | 17 |
| Figura 23. | Resultado consulta | 17 |
| III.II.V.III | PRUEBAS TERCERO, QUINTO Y NOVENO..... | 18 |
| Figura 24. | Elegir tablero dinámico | 18 |
| Figura 25. | Tablero dinámico | 18 |
| Figura 26. | Selección año | 19 |
| Figura 27. | Selección región..... | 19 |
| Figura 28. | Selección indicador..... | 19 |

I. USUARIOS

I.I. ADMINISTRADOR

El administrador es el usuario encargado del manejo administrativo del aplicativo web, sus funciones son:

- Registrar, modificar, consultar y eliminar usuarios
- Cargar resultados saber y llevar un control sobre los resultados.
- Ver listado de resultados cargados en base de datos

I.II. USUARIO

El usuario es el rol que tiene acceso a los resultados de los análisis hechos con las pruebas sus funciones son:

- Ver resultados descriptivos de las pruebas saber 3, 5, 9 y 11.
- Ver resultados predictivos de las pruebas saber 11

II. REQUISITOS DE SOFTWARE Y HARDWARE


II.I. REQUERIMIENTOS DE SOFTWARE

- Contar con un equipo electrónico, ya sea computador o portátil.
- Navegador: Contar con navegadores como Firefox o Google Chrome.
- Es indispensable contar con acceso a internet.

III. APLICATIVO

III.I. ICONOS GENERALES

Tabla 1. Iconos generales

| | | | |
|---|-------------|--|-----------------------------------|
|  | AWID | | Logo principal del aplicativo web |
|---|-------------|--|-----------------------------------|

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

III.II. INGRESO AL APLICATIVO

Para ingresar al aplicativo web es necesario ingresar a la siguiente dirección <https://ac24-190-145-240-29.ngrok-free.app/proyecto/>, allí se visualizará la siguiente ventana:

Figura 1. Pagina de acceso



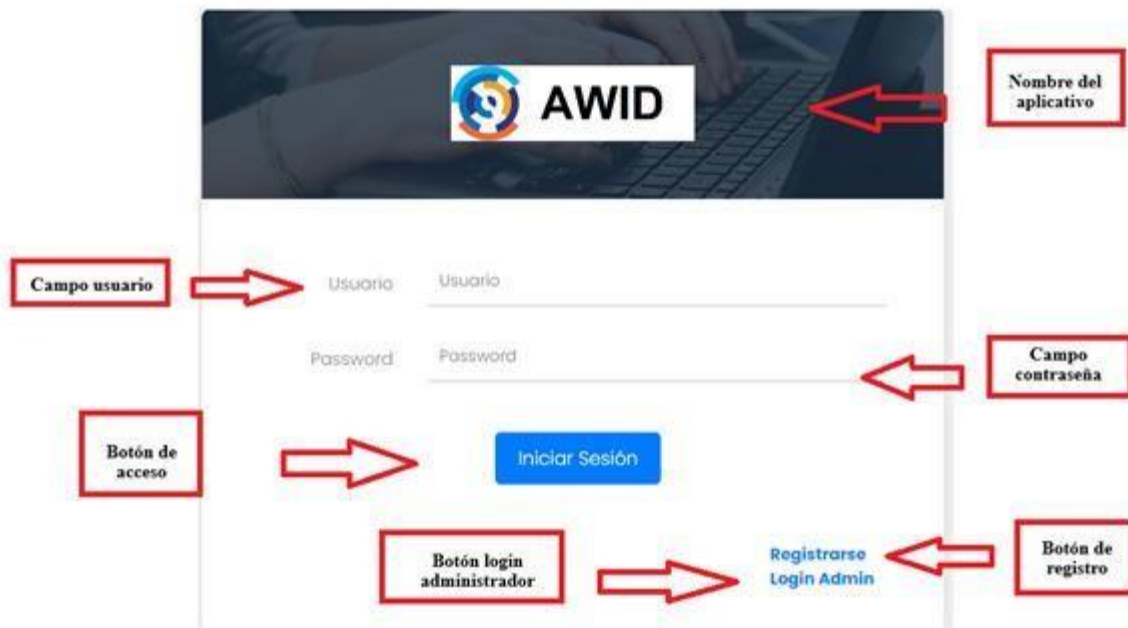
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

La imagen anterior muestra la página principal del aplicativo web, la cual corresponde al inicio de sesión.

III.II.I. INICIO DE SESIÓN

Para iniciar la persona debe acceder con el usuario y la contraseña asignados durante el registro.

Figura 2. Inicio de session



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Si la persona no está registrada dentro del aplicativo, debe dar clic en el botón registrarse.

III.II.II. REGISTRARSE

En caso de que la persona desee acceder al módulo de usuarios, debe diligenciar este formulario.

Figura 3. Pagina de registro

The image shows a screenshot of the AWID user registration page. At the top, there is a banner with the AWID logo and the text "AWID". Below the banner, the title "REGISTRO DE USUARIO" is centered. The registration form consists of five input fields: "Nombre", "Apellido", "Email", "Usuario", and "Contraseña". A red vertical line is drawn to the left of these fields, with a red arrow pointing to a box labeled "Campos del registro". Below the form, there is a blue button labeled "Registrarse" and a link labeled "Volver a la página de inicio de sesión". A red arrow points from a box labeled "Botón para registrarse" to the "Registrarse" button. Another red arrow points from a box labeled "Enlace para volver al login" to the "Volver a la página de inicio de sesión" link.

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Al momento de registrarse, al correo le llega un mensaje de bienvenida y se le muestra el usuario y la contraseña.

III.II.III. INICO DE SESION ADMINISTRADOR

Figura 4. Login administrador

The image shows a login form for the AWID administrator interface. At the top, there is a banner with the AWID logo and the text "AWID". Below the banner, the form consists of two input fields: "Email Address" (with a placeholder "E-mail") and "Password" (with a placeholder "Password"). A blue button labeled "Iniciar Sesión" is positioned below the password field. To the left of the button is a link that says "Volver a la página de inicio de sesión". Red arrows and boxes highlight specific elements: "Campo email" points to the email field, "Campo contraseña" points to the password field, "Botón inicio de sesión" points to the "Iniciar Sesión" button, and "Enlace para volver al login" points to the return link.

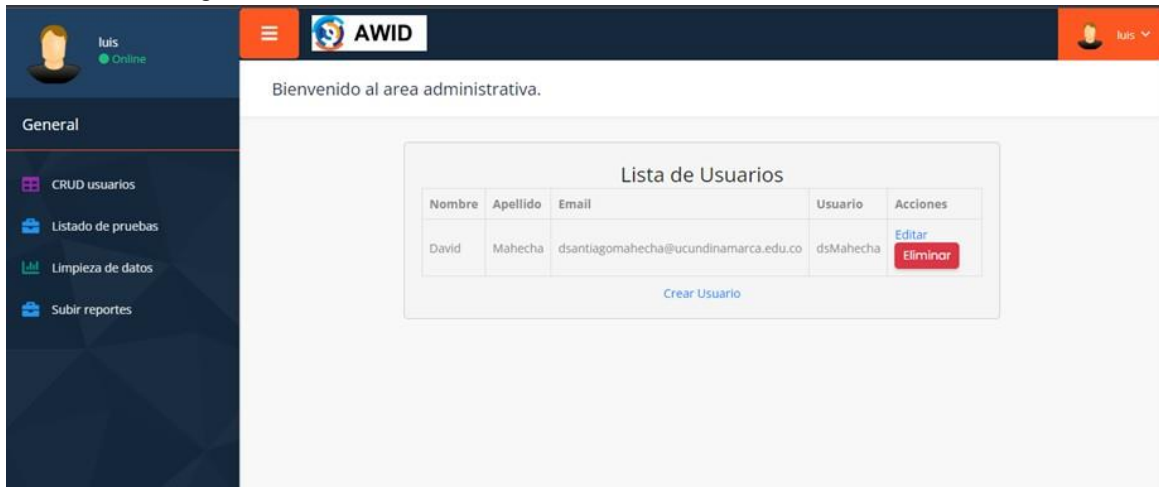
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

III.II.IV. MODULO ADMINISTRADOR

En este módulo gestiona la administración de los usuarios, el proceso de carga pruebas, consulta de las pruebas cargadas al aplicativo y la limpieza de datos.

III.II.IV.I. PAGINA PRINCIPAL MODULO ADMINISTRADOR

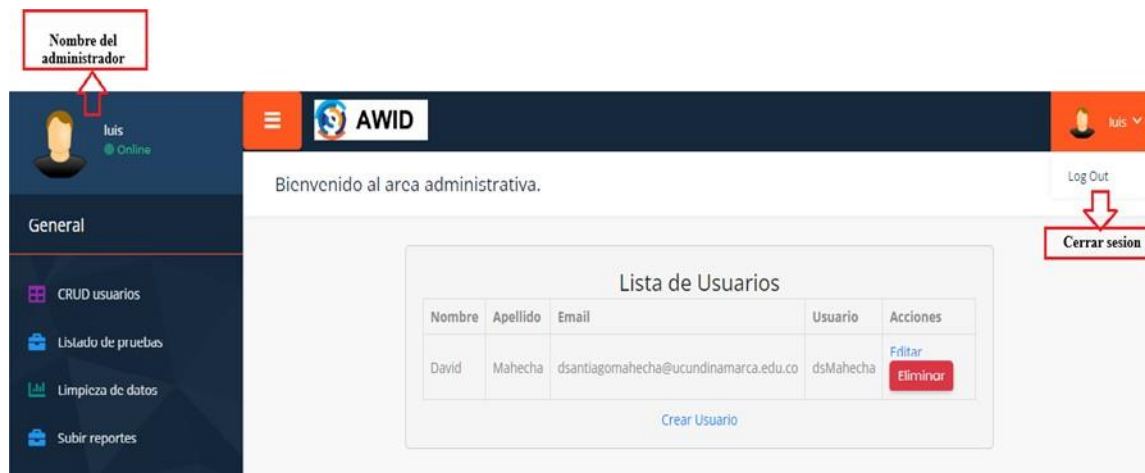
Figura 5. CRUD de usuarios



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En la parte superior vemos el nombre del administrador y la opción para cerrar sesión.

Figura 6. Características



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En el CRUD de usuarios tenemos 4 opciones la primera es visualizar que usuarios están registrados, segundo la opción de editar usuario, la tercera la opción de eliminar y por último la opción de crear usuario.

Figura 7. Opciones del CRUD de usuarios

The screenshot shows a table titled "Lista de Usuarios" with the following data:

| Nombre | Apellido | Email | Usuario | Acciones |
|--------|----------|---------------------------------------|-----------|--|
| David | Mahecha | dsantiagomahecha@ucundinamarca.edu.co | dsMahecha | Editar Eliminar |

Below the table is a link "Crear Usuario".

Annotations with red boxes and arrows:

- "Opción de crear usuario" points to the "Crear Usuario" link.
- "Opción eliminar usuario" points to the "Eliminar" button.
- "Opción editar usuario" points to the "Editar" link.

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

III.II.IV.II. MODULO EDITAR USUARIOS

Cuando el administrador seleccione esta opción se abrirá el siguiente formulario:

Figura 8. Pagina editar usuarios

The screenshot shows the "AWID" logo at the top and a form with the following fields:

- Nombre: David
- Apellido: Mahecha
- Email: dsantiagomahecha@uct
- Usuario: dsMahecha
- Contraseña:

Annotations with red boxes and arrows:

- "Campos que el administrador puede editar" points to the form fields.
- "Botón para guardar los cambios" points to the "Guardar" button.

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Cuando el administrador selecció el botón de guardar se dirigirá a la página principal del administrador.

III.II.IV.III. MODULO ELIMINAR USUARIOS

Cuando el administrador seleccione esta opción, inmediatamente el usuario queda eliminado

III.II.IV.IV. MODULO CREAR USUARIOS

Cuando el administrador seleccione esta opción se abrirá el siguiente formulario

Figura 9. Pagina crear usuarios

The image shows a web form for creating a user in the AWID system. The form is titled 'AWID' and contains the following fields: 'Nombre:', 'Apellido:', 'Email:', 'Usuario:', and 'Contraseña:'. Each field is represented by a white input box. Below these fields is a blue button labeled 'Guardar'. Red arrows point from text boxes on the right to each input field, and a red arrow points from a text box on the left to the 'Guardar' button. The text boxes on the right are labeled 'Campo de nombre', 'Campo de apellido', 'Campo de correo', 'Campo de usuario', and 'Campo de contraseña'. The text box on the left is labeled 'Boton para guardar la informacion'.

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El administrador podrá crear un usuario de manera sencilla cargando la información en cada campo. En el momento que oprima el botón de guardar se redirigirá a la pagina principal del administrador.

III.II.IV.V. MODULO LISTADO DE PRUEBAS

La primer opción es el CRUD de usuarios que se mostró anteriormente, la segunda es el listado de pruebas, la tercera la limpieza de los datos y por ultimo podrá hacer la carga de los reportes.

Figura 10. Listados de pruebas/resultados

| Tipo de Análisis | Año | Región | Resultado Texto | Acciones |
|------------------|------|-----------|--|----------|
| descriptivo | 2017 | pacifica | Información descriptivo de la región pacifica en el año 2017. | Eliminar |
| predictivo | 2017 | pacifica | Información predictivo de la región pacifica en el año 2017. | Eliminar |
| descriptivo | 2018 | pacifica | Información descriptivo de la región pacifica en el año 2018. | Eliminar |
| predictivo | 2020 | pacifica | Información predictivo de la región pacifica en el año 2020. | Eliminar |
| descriptivo | 2021 | pacifica | Información descriptivo de la región pacifica en el año 2021. | Eliminar |
| descriptivo | 2017 | andino | Información descriptivo de la región andina en el año 2017. | Eliminar |
| descriptivo | 2017 | caribe | Información descriptivo de la región caribe en el año 2017. | Eliminar |
| descriptivo | 2017 | orinoquia | Información descriptivo de la región orinoquia en el año 2017. | Eliminar |
| predictivo | 2017 | andino | Información predictivo de la región andina en el año 2017. | Eliminar |
| predictivo | 2017 | caribe | Información predictivo de la región caribe en el año 2017. | Eliminar |
| predictivo | 2017 | orinoquia | Información predictivo de la región orinoquia en el año 2017. | Eliminar |
| predictivo | 2017 | amazonica | Información predictivo de la región amazonica en el año 2017. | Eliminar |

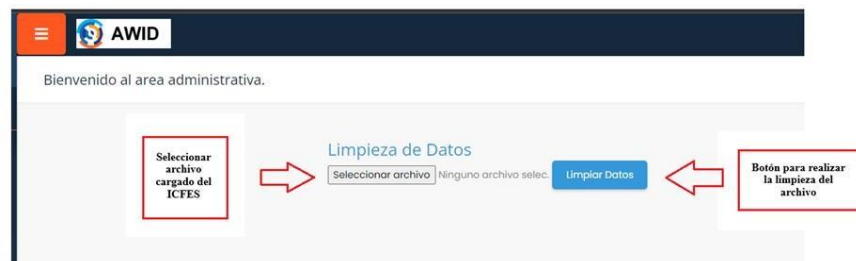
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El administrador podrá ver todas las pruebas cargadas que están en la base de datos

III.II.IV.VI. MODULO LIMPIEZA DE DATOS

Cuando el administrador entre a la limpieza de datos podrá cargar el archivo cargado del ICFES, ya que estos archivos contienen caracteres especiales los cuales no acepta weka y no se puede realizar el análisis, por ello se hace uso de este apartado.

Figura 11. Pagina para la limpieza de datos



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Al momento que el administrador seleccione y de click en el botón, el aplicativo se demora unos segundos para realizar la limpieza.

Figura 12. Resultado de la limpieza de datos



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Cuando el aplicativo realizo la limpieza, se vera la opción de un link llamado descargar CSV y ese archivo ya se puede cargar y analizar correctamente en WEKA.

III.II.IV.VII. MODULO DE CARGA DE REPORTES

Cuando el usuario este en este apartado tendrá la opción de elegir el tipo de análisis, la región y el año y por último podrá subir el reporte correspondiente.

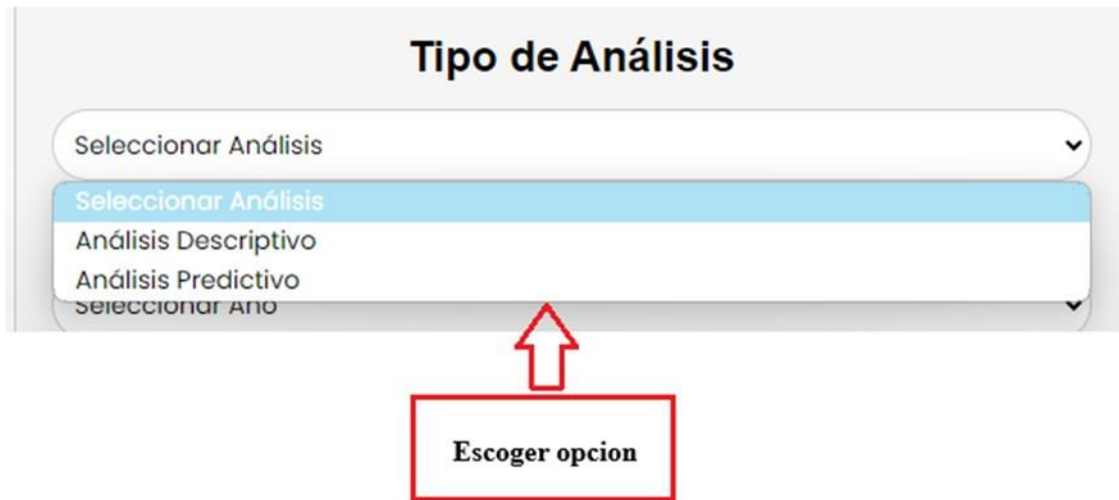
Figura 13. Carga de reportes

The image shows a web interface for uploading reports. At the top, there is a dark blue header with the 'AWID' logo. Below the header, it says 'Bienvenido al area administrativa.' The main content area is titled 'Adjunte los archivos' in blue. Below this title is a form with four sections: 'Tipo de Análisis' with a dropdown menu showing 'Seleccionar Análisis'; 'Año' with a dropdown menu showing 'Seleccionar Año'; 'Región' with a dropdown menu showing 'Seleccionar Región'; and 'Archivo PDF' with a file input field showing 'Seleccionar archivo' and 'Ninguno archivo selec.'. At the bottom of the form is a blue button labeled 'Agregar'.

Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Escogemos el tipo de análisis, ya sea predictivo o descriptivo

Figura 14. Tipo de analisis



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Seleccionamos el año del reporte

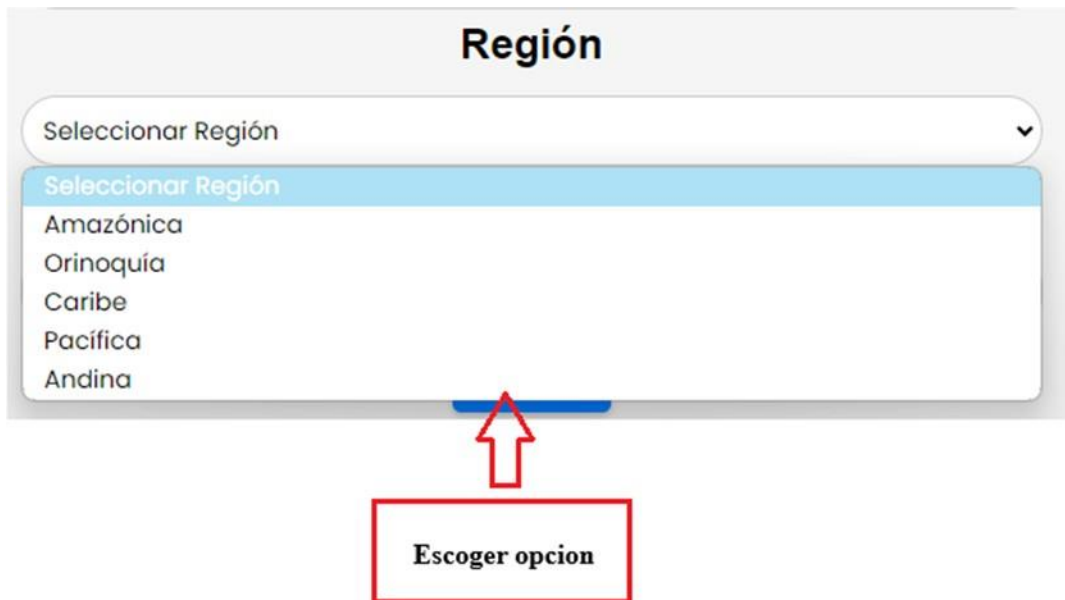
Figura 15. Seleccion de años analizados



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Seleccionamos la región del reporte

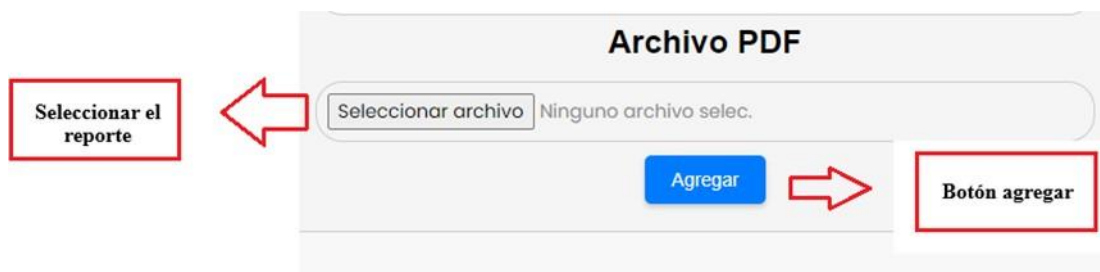
Figura 16. Regiones colombianas analizadas



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Por ultimo cargamos el PDF correspondiente y le damos en el botón agregar.

Figura 17. Seleccion del reporte correspondiente



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Con esto el administrador ya puede verificar en el listado de pruebas si cargo el archivo y además el reporte se sube a un bucket de AWS S3.

III.II.V. MODULO USUARIOS

En este módulo se tendrá acceso a los diferentes reportes de las pruebas saber tercero, quinto, noveno y once, en las pruebas once tendremos 2 análisis uno descriptivo y otro predictivo, mientras que en las demás pruebas tendremos un análisis descriptivo. Por ultimo contamos con un tablero dinamico para que el usuario pueda mirar diferentes graficas según el gusto de el sin necesidad de descargar como tal los reportes.

Figura 18. Pagina principal modulo Usuario



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

III.II.V.I. PRUEBAS ONCE

El usuario podrá seleccionar el tipo de análisis que desea ver, la región y el año.

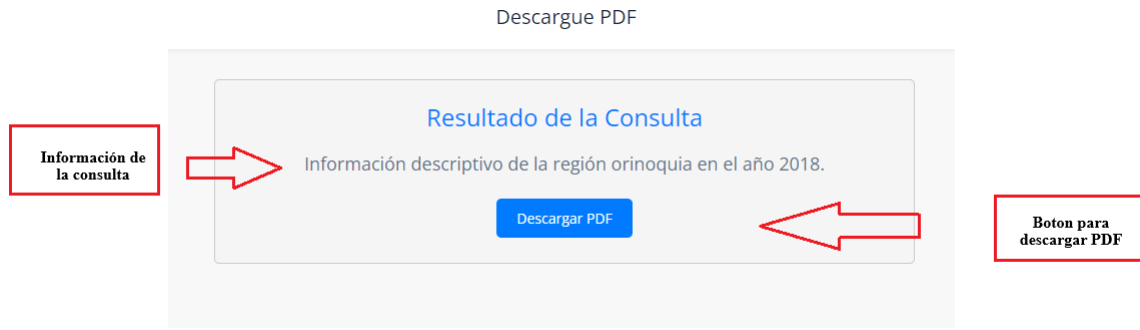
Figura 19. Menu de tipo de análisis



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Cuando el usuario seleccione cada opción y le de click en consultar aparecerá una nueva pagina.

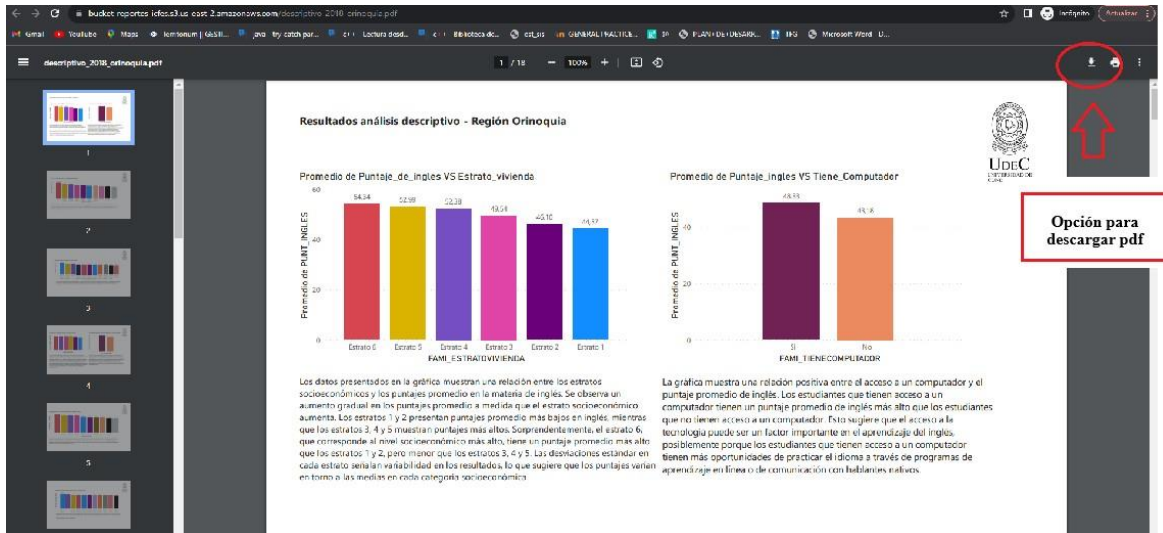
Figura 20. Obtener resultados por reportes



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El usuario observará la información seleccionada por el y posteriormente tendrá la opción de descargar el pdf, cuando el usuario de click en el botón se abrirá el pdf en el navegador.

Figura 21. Ejemplo de reporte



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

III.II.V.II. PRUEBAS TERCERO, QUINTO Y NOVENO

Figura 22. Menu opciones



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Cuando el usuario quiera ver el analisis de estas pruebas, va ser un poco diferente en comparación con las de once, aca el usuario solo deberá escoger 2 opciones el grado y el año.

Cuando el usuario escoja ambas opciones y le de click en consultar se abrirá otra pestaña.

Figura 23. Resultado consulta



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

El usuario observará la información seleccionada por el y posteriormente tendrá la opción de descargar el pdf, cuando el usuario de click en el botón se abrirá el pdf en el navegador.

III.V.II.III TABLERO DINAMICO

Figura 24. Elegir tablero dinamico.



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

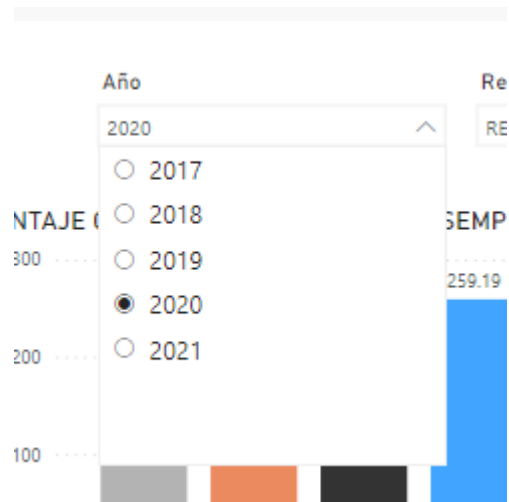
El usuario debera elegir esta opcion y se abra el siguiente apartado:

Figura 25. Tablero dinamico



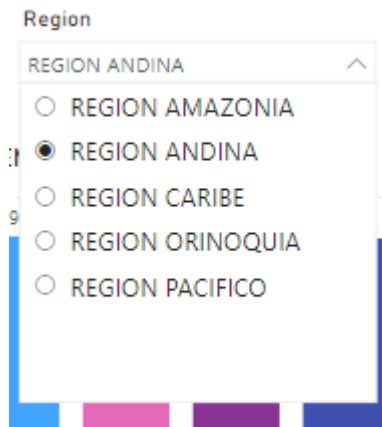
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

Figura 26. Selección año



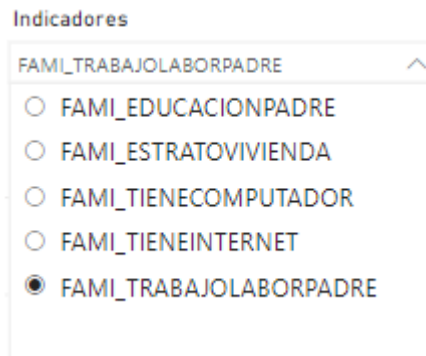
Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)
En este apartado el usuario podrá seleccionar el año.

Figura 27. Selección region



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)
En este apartado el usuario podrá seleccionar la region.

Figura 28. Selección indicador



Fuente. (Vergara Serrato, Mahecha Hernández, Barahona Rodríguez)

En este apartado el usuario podrá seleccionar el indicador.



Anexo B Manual tecnico

Contenido

| | |
|------------------------------------|----|
| 1. Especificaciones técnicas..... | 3 |
| Necesidades técnicas..... | 3 |
| Necesidades de software..... | 3 |
| 2. INSTALACION de jdk y jre..... | 3 |
| 3. Instalación eclipse | 5 |
| 4. Instalación de PostgreSQL | 9 |
| 5. Instalación de WEKA..... | 13 |

1. Especificaciones técnicas

Necesidades técnicas

- Sistema Operativo Windows 7 o superior.
- Equipo de cómputo con memoria RAM de 2GB o superior.
- Espacio disco duro de 250 Gb en adelante

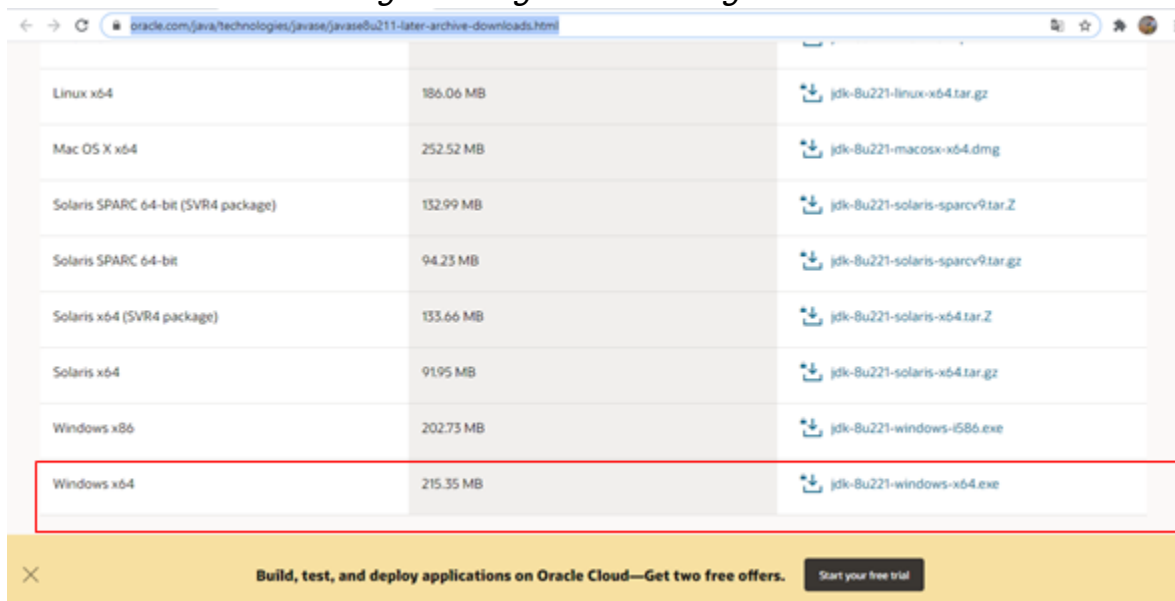
Necesidades de software

- PostgreSQL Versión 11 o superior.
- IDE: Eclipse
- JAVA JDK 8 o superior.
- WEKA 3.8.4 o superior.

2. INSTALACION de jdk y jre

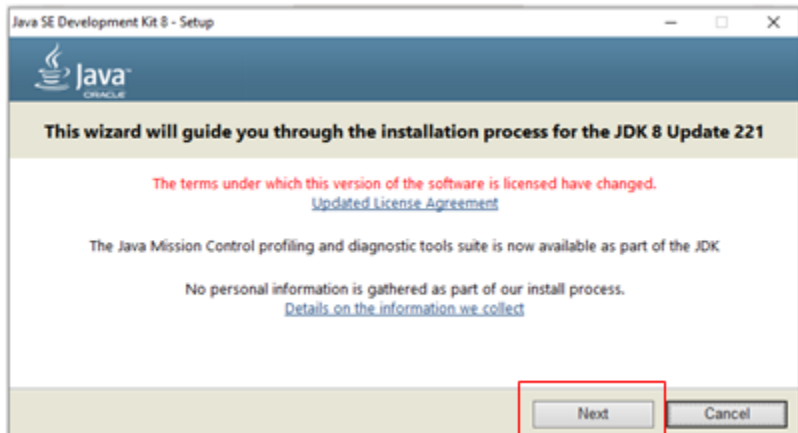
Para la instalación del JDK, es necesario descargar de la página de Oracle el siguiente componente:

Imagen 1. Página de descarga del JDK

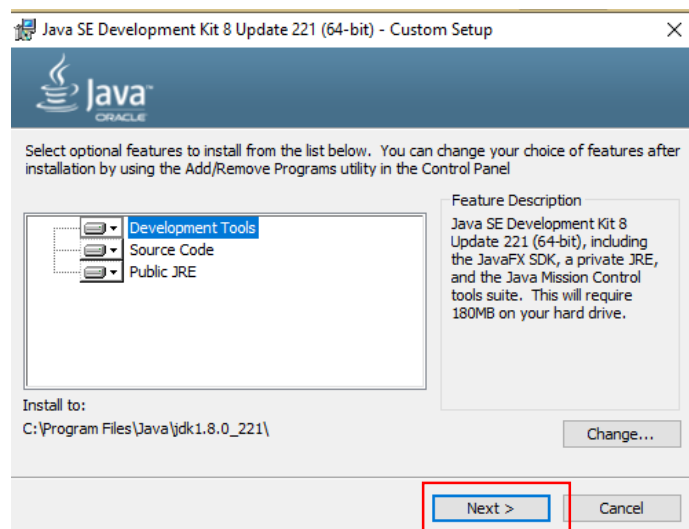


Una vez se ha descargado el archivo .EXE, se procede a la ejecución del mismo

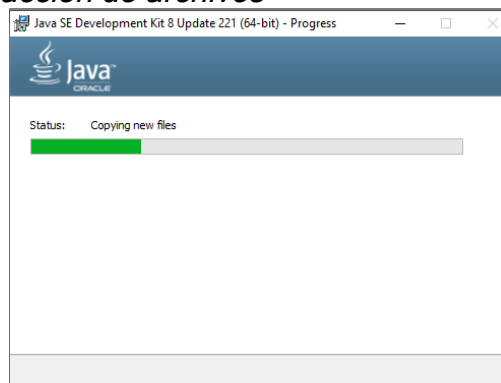
Imagen 2. Ejecución de archivo JDK



Se debe dar clic en el botón “Next”, una vez seleccionado se mostrara una nueva ventana con los componentes a instalar
Imagen 3. Ventana de componentes a instalar

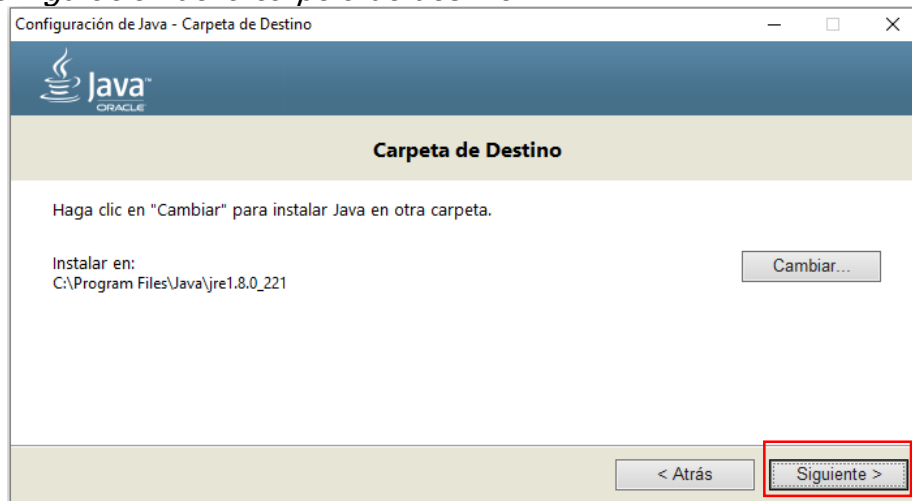


Nuevamente se debe seleccionar el botón “Next”, allí aparecerá una barra de carga indicando el estado del proceso de copiado de archivos.
Imagen 4. Proceso de extracción de archivos



A continuación se abrirá una nueva ventana de configuración, allí se debe especificar en qué lugar quedara guardado el JDK.

Imagen 5. Configuración de la carpeta de destino



Posteriormente seleccionaremos el botón siguiente, allí iniciara el proceso de instalación de los componentes.

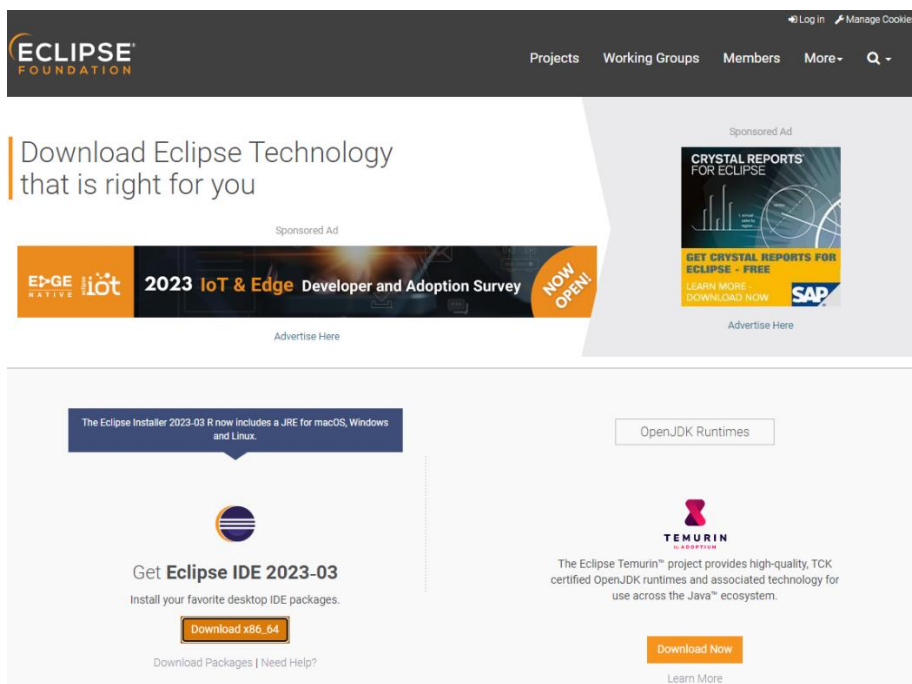
Imagen 6. Instalación de componentes



3. Instalación eclipse

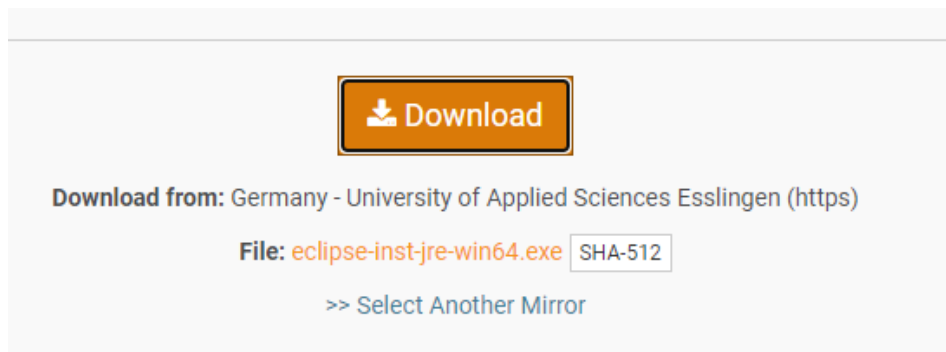
La descarga de la última versión de Eclipse es muy sencilla, se realiza desde: <https://www.eclipse.org/downloads/>

Imagen 7. Acceder al sitio oficial de eclipse



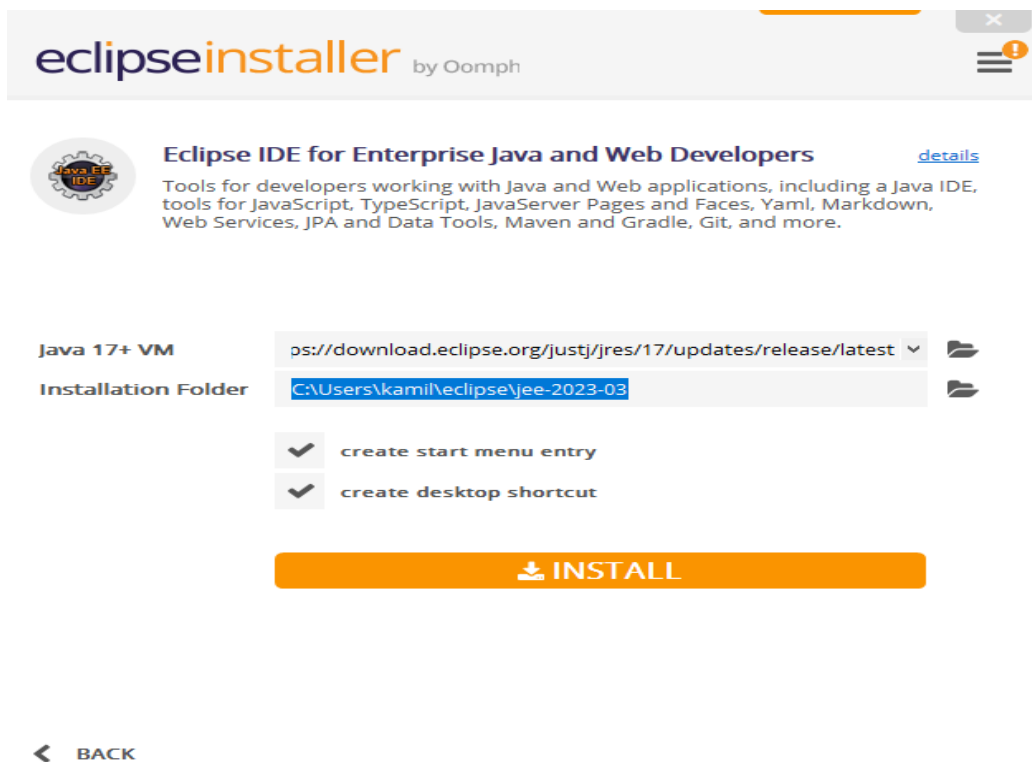
Hacemos click en “Download”

Imagen 8. Descargar el archivo



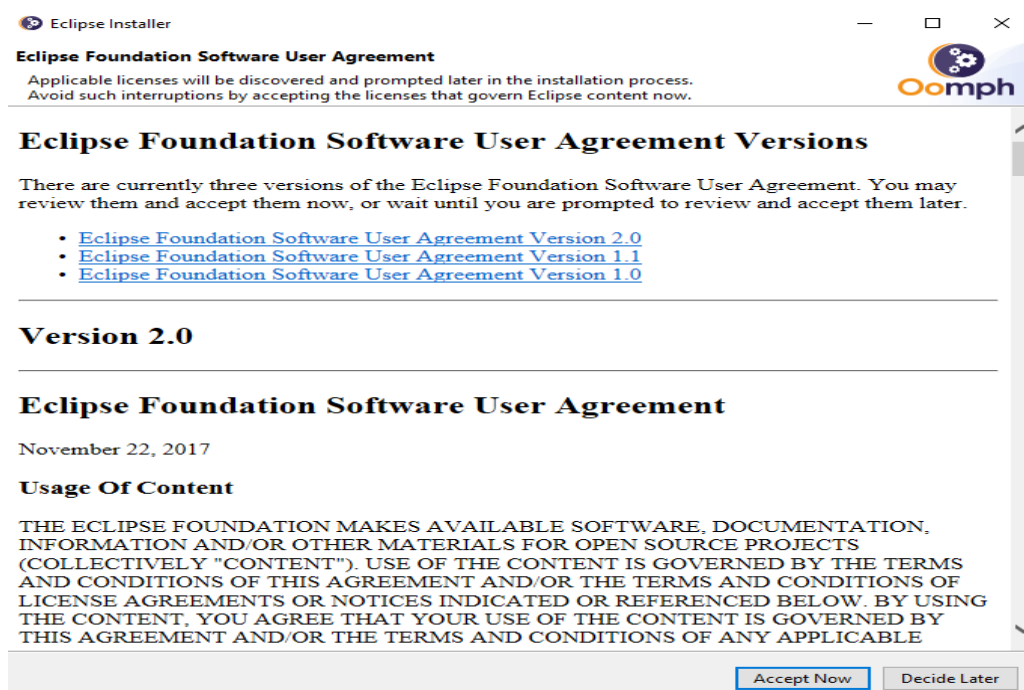
Elegimos la opción para instalar la versión que necesitamos. En este caso vamos a elegir la de Java y Web.

Imagen 9. Instalación de archivos



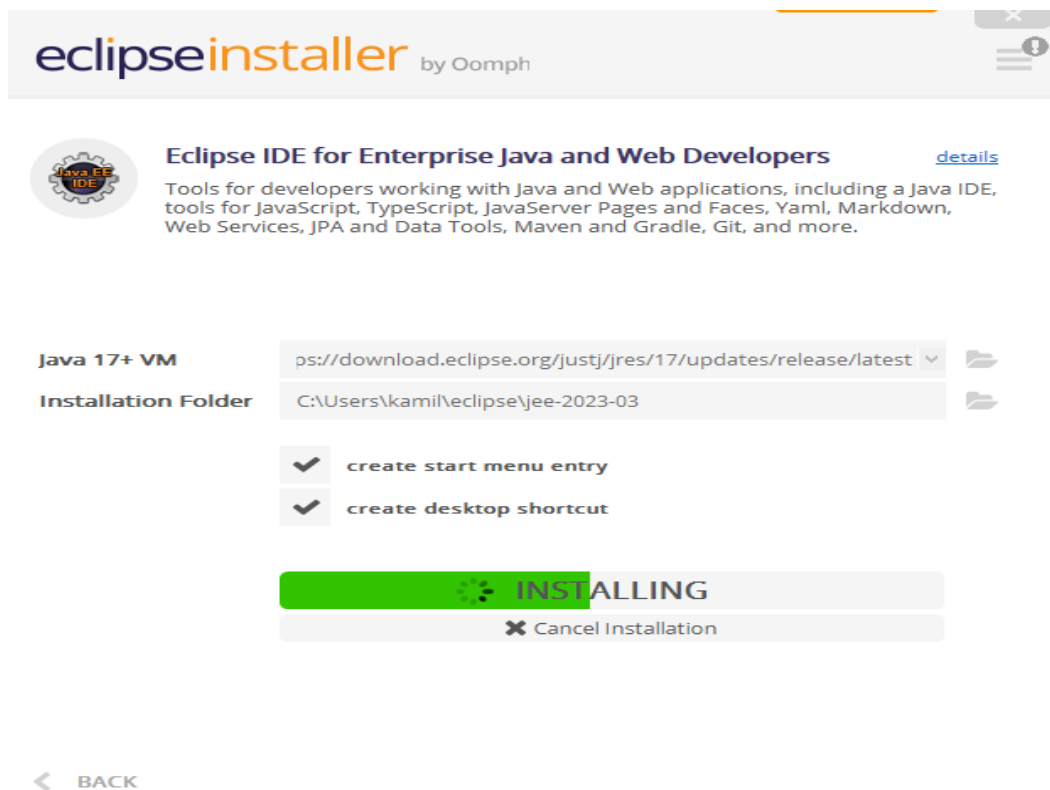
Debemos leer y aceptar los términos de licencia de eclipse haciendo click en “Accept Now”

Imagen 10. Aceptar términos y condiciones



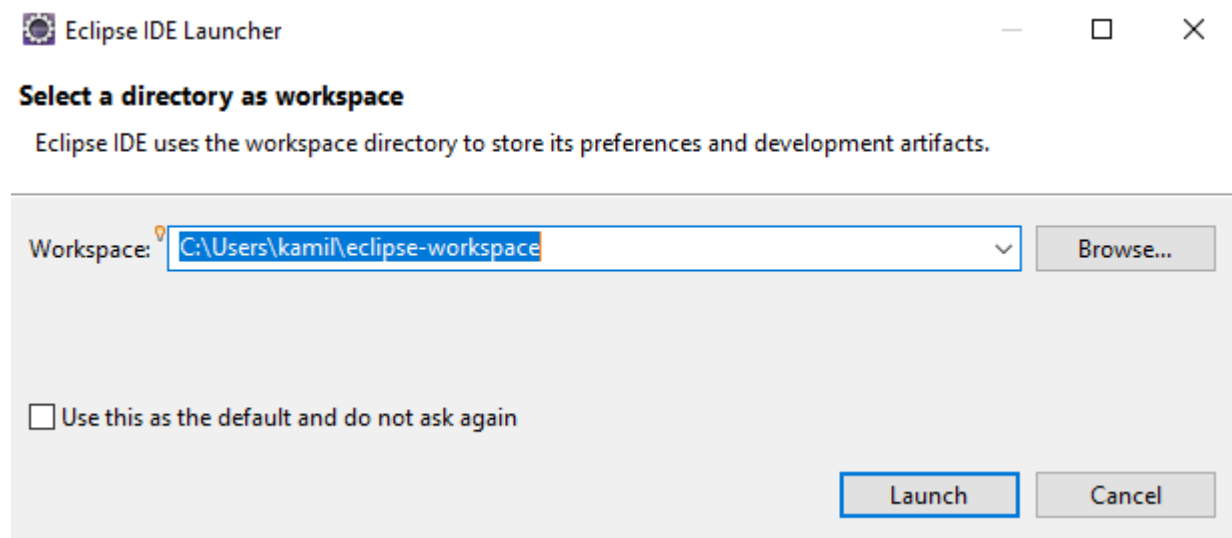
Comienza la instalación

Imagen 11. Instalación de componentes



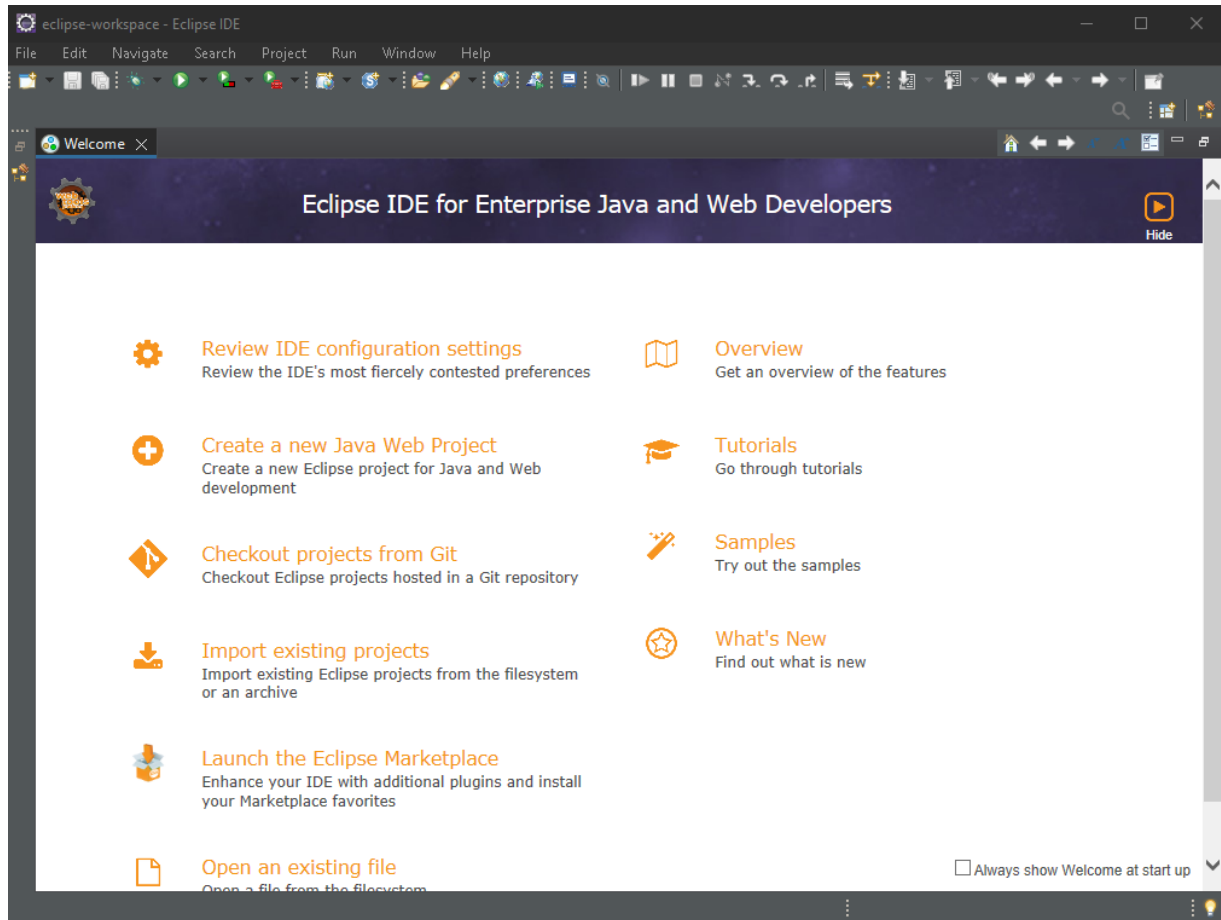
Cuando termine la instalación ya podemos ejecutar por primera vez eclipse haciendo click en "launch"

Imagen 12. Ehecucion de eclipse



Ya nos encontramos ante la interfaz de eclipse. Lo primero que se abre es la pantalla de bienvenida, en la que tenemos varias opciones, tutoriales y recomendaciones básicas.

Imagen 13. Interfaz eclipse



4. Instalación de PostgreSQL

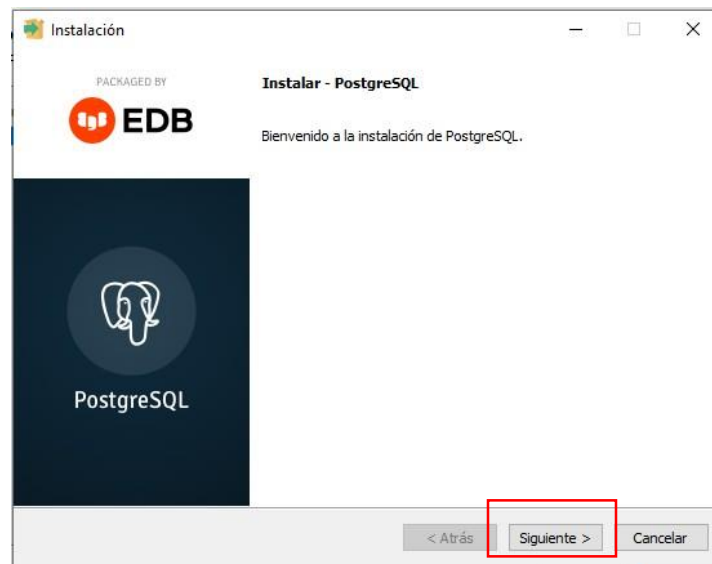
Para la instalación se debe ingresar a la página de EDB y allí seleccionar el siguiente componente:

Imagen 14. Descarga del PostgreSQL

| Version | Linux x86-64 | Linux x86-32 | Mac OS X | Windows x86-64 | Windows x86-32 |
|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 13 | N/A | N/A | Download | Download | N/A |
| 12.4 | N/A | N/A | Download | Download | N/A |
| 11.9 | N/A | N/A | Download | Download | N/A |
| 10.14 | Download | Download | Download | Download | Download |
| 9.6.19 | Download | Download | Download | Download | Download |
| 9.5.23 | Download | Download | Download | Download | Download |
| 9.4.26 (Not Supported) | Download | Download | Download | Download | Download |
| 9.3.25 (Not Supported) | Download | Download | Download | Download | Download |

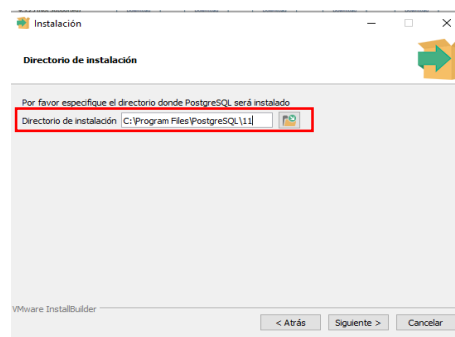
Una vez descargado el ejecutable procedemos a iniciar la instalación.

Imagen 15. Ventana principal de instalación del PostgreSQL



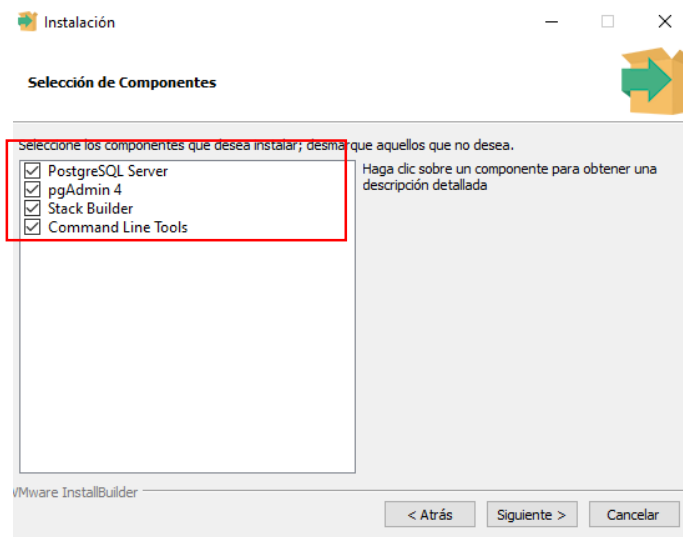
A continuación, se debe especificar el directorio donde quedara guardado el programa.

Imagen 16. Configuración de directorio de PostgreSQL



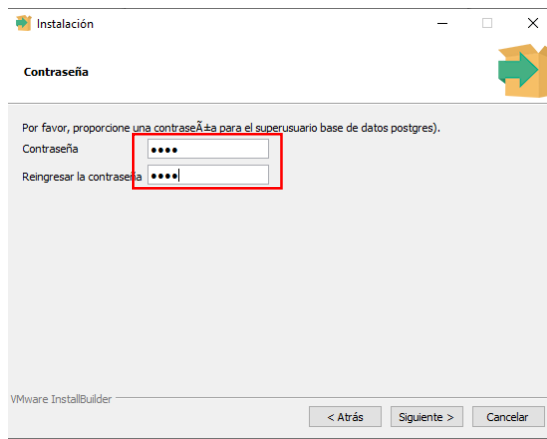
Luego seleccionaremos que componentes de PostgreSQL queremos instalar, para nuestro caso dejaremos seleccionado como esta por defecto. Damos clic en siguiente.

Imagen 17. Configuración de componentes a instalar



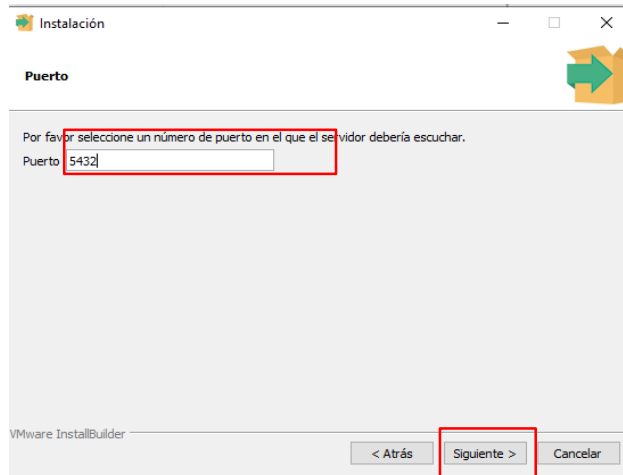
Ahora aparecerá un formulario en el cual debemos indicar la contraseña de acceso a PostgreSQL

Imagen 18. Configuración de contraseña



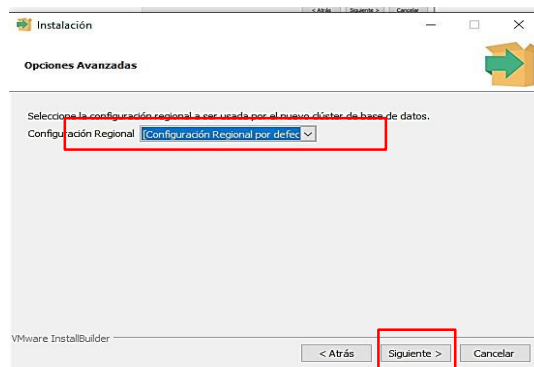
A continuación, solicitara ingresar el puerto de conexión con el servidor, por defecto siempre es el 5432

Imagen 19. Configuración del puerto de conexión



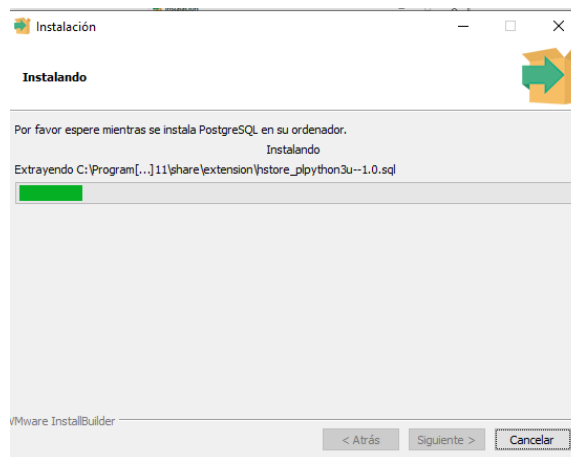
Luego de dar clic en siguiente, se debe seleccionar la configuración regional, por defecto dejamos la que está.

Imagen 20. Configuración regional



Por último esperamos a que termine la instalación.

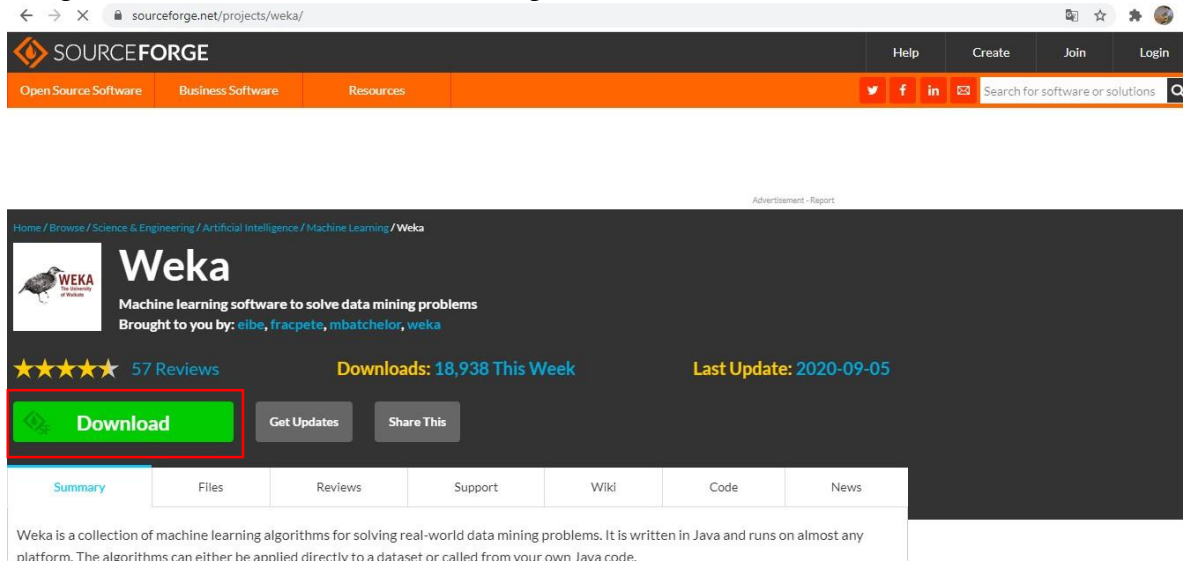
Imagen 21. Instalación de postgresQL



5. Instalación de WEKA

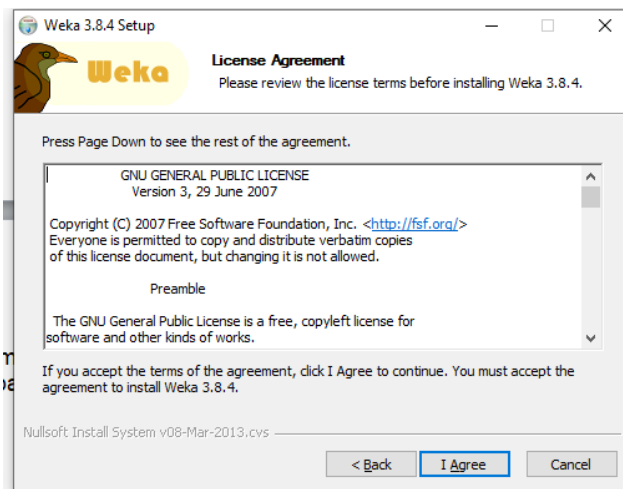
Para la descarga de WEKA debemos ingresar a la página de SourceForge y buscarla herramienta WEKA.

Imagen 21. Descarga de WEKA



Procedemos a instalar el programa, en la primera parte debemos aceptar términos de licencia para ello seleccionaremos el botón “I Agree”.

Imagen 22. Términos y condiciones de licencia



Una vez seleccionado el botón “Next”, debemos indicar en donde queremos guardarla instalación del programa.

Imagen 23. Configuración del directorio

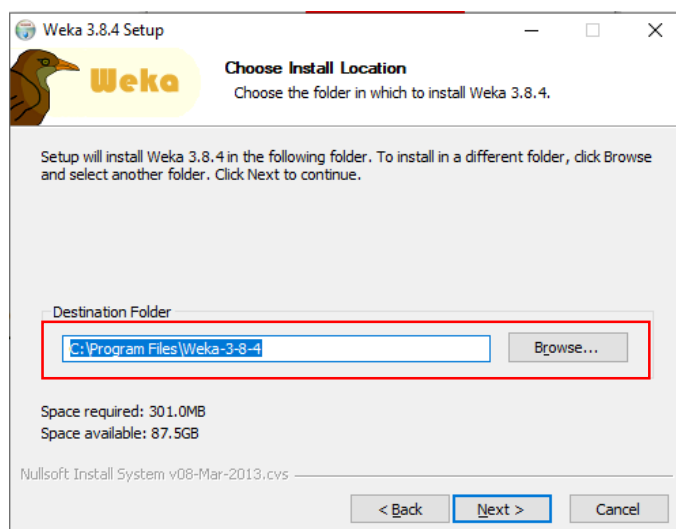
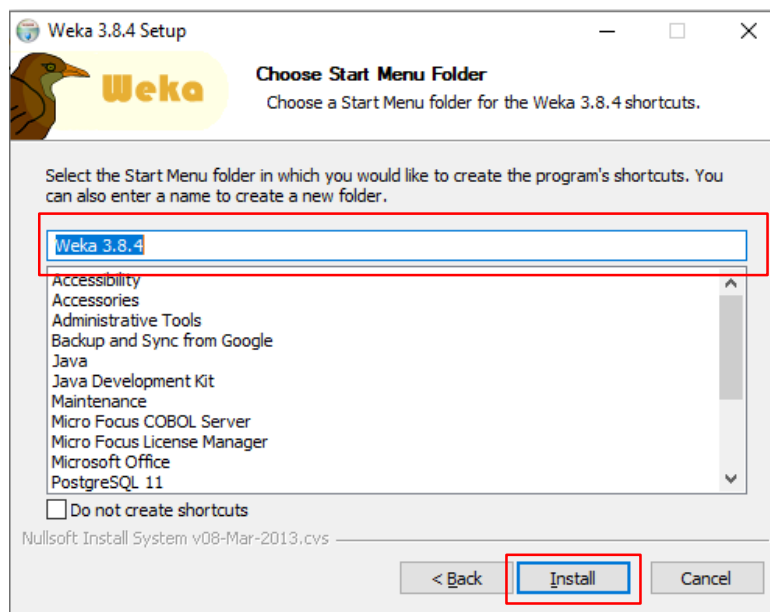


Imagen 24. Configuración del folder inicial



Por ultimo damos clic sobre el botón “install”, esperamos a que instale todos los componentes.

Revisión literaria de indicadores que influyen en la educación mediante técnicas de minería de datos.

Luis Felipe Vergara Serrato, David Santiago Mahecha Hernández, y cesar Yesid Barahona Rodríguez

Resumen - Esta revisión literaria integrativa o estado del arte ha sido organizada de manera clara para abordar los temas relacionados con la educación de los estudiantes y las técnicas de minería de datos. La educación es un tema vital en la vida de los seres humanos, en Colombia, esta se divide en educación inicial, la educación preescolar, la educación básica (primaria cinco grados y secundaria cuatro grados), la educación media (dos grados y culmina con el título de bachiller), y la educación superior. Lo que se busca es analizar diferentes indicadores de desempeño en los estudiantes colombianos por medio de técnicas de minería de datos. Lo primero que haremos es encontrar cuales son estos indicadores por medio de diferentes autores, después de esto se hará una búsqueda de los patrones más eficaces y eficientes del análisis de datos.

I. INTRODUCCION

Los indicadores de desempeño son instrumentos que proporcionan información cuantitativa sobre el desenvolvimiento y logros de una institución, programa, actividad o proyecto a favor de la población u objeto de su intervención, en el marco de sus objetivos estratégicos y su Misión. Los indicadores de desempeño establecen una relación entre dos o más variables, que, al ser comparados con periodos anteriores, productos similares o metas establecidas, permiten realizar inferencias sobre los avances y logros de las instituciones y/o programas.[1], para la búsqueda de estos indicadores se toman los conceptos claves como: academic performance, indicators, factors, school, análisis y family.

La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos. [2], Para el presente estudio se han examinado particularmente cuatro temáticas principales: data mining, decisión trees, y education.

Se usan como instrumentos los motores de búsqueda de las bases de datos de Scopus, ScienceDirect e IEEE, con el fin de determinar las publicaciones más recientes, así como el análisis de los temas que hacen parte del área. Por medio de los conceptos o temáticas principales se realizaron ecuaciones de búsqueda específicas.

La revisión bibliográfica fue un recurso para aplicar técnicas de mapeo VOS con el fin de realizar mediciones bibliométricas y descubrir tendencias en las investigaciones ligadas a la minería de datos e indicadores. Este análisis bibliométrico basado en un

Documento recibido el 9 de octubre de 2001. (Anote la fecha en que usted presentó su documento para su revisión.) Este trabajo fue apoyado en parte por los U.S. Departamento of Commerce under Grant S123456 (reconocimiento al patrocinador y apoyo financiero va aquí). los títulos del Documento deben ser escritos en letras mayúsculas y minúsculas, no todas las mayúsculas. Evite escribir fórmulas extensas con subíndices en el título; Utilice Fórmulas cortas que identifiquen los elementos (por ejemplo, "Nd-Fe-B"). No escriba "(invitados)" en el título. Escriba los Nombres completos de los autores en el campo autor, pero no es necesario. Ponga un espacio entre los autores.

F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (corresponding author to provide phone: 303-555-5555; fax: 303-555-5555; e-mail: author@ boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

mapa de coocurrencia de palabras claves, permite encontrar conexiones de temáticas propias de la minería de datos, como el uso de diferentes técnicas o algoritmos para la obtención de patrones y además encontrar los factores e incidencias del desempeño de los estudiantes.

II. MÉTODO

En la metodología se consideran las fuentes de los documentos estudiados, así como la manera de llegar a ellos mediante las ecuaciones de búsqueda, han sido elegidos Scopus, ScienceDirect e IEEE para la construcción del estado del arte debido a que son enfocados en la rama de la ingeniería además permite acceder a publicaciones procedentes de más de 5.000 editoriales internacionales revisadas por especialistas. Las publicaciones, revistas y bases de datos que fueron tenidas en cuenta en este trabajo son las expuestas a continuación:

- Elsevier: Es un editorial multimedia internacional de análisis de información general que asiste a instituciones en el avance de la ciencia, cuidados avanzados en temas médicos, así como mejorar la realización de los mismos para el beneficio de la sociedad. Los productos que se ofrecen incluyen revistas, colecciones de revistas electrónicas y el índice de citas bibliográficas de Scopus, entre otros. Es una importante editorial de libros académicos muy reconocida en la comunidad científica internacional por la calidad y su amplia lista de publicaciones, además sus divulgaciones incluyen también autores reconocidos, así como instituciones científicas [3].

- Scopus: Esta es la base de datos más grande de resúmenes y citas de literatura revisada la cual cuenta con revistas, libros y actas de congresos. Scopus proporciona una descripción general completa de los resultados de la investigación global en los campos de la ciencia, la tecnología, la medicina, las ciencias sociales, las artes y las humanidades, y presenta herramientas inteligentes para rastrear, analizar y visualizar la investigación [4].

- Institute of Electrical and Electronics Engineers (IEEE): Es una entidad global que se dedica a promover la creatividad, el desarrollo y la integración dando avance a la tecnología en beneficio de las personas y a la normalización. Es la mayor organización a nivel mundial sin ánimo de lucro de profesionales de las nuevas tecnologías. Esta base de datos ofrece conferencias, publicaciones y herramientas, como la Encuesta y guía de comunicación de IEEE, la Red IEEE, la Revista de comunicación de IEEE y muchas más [3].

El proceso llevado a cabo para la construcción de este documento tuvo en cuenta ecuaciones de búsqueda. En el momento de generar la ecuación se tuvieron en cuenta los siguientes pasos para cada uno de los temas claves:

- Temas principales relacionados con el operador AND.
- Años cubiertos: desde 2016 hasta el 2022.
- Campos de estudio: ciencias de la computación, ingeniería.

Los documentos tenidos en cuenta fueron, artículos, revistas y libros.

Las palabras clave correspondientes a los temas principales mencionados al inicio, son:

A. Minería de datos

A continuación, se realizara la definición de los conceptos usados en la búsqueda de las técnicas de minería de datos en las investigaciones realizadas en las bases de datos Scopus, ScienceDirect e IEEE XPLORE.

- Data mining: La minería de datos incluye un conjunto de técnicas que estudian automáticamente grandes bases de datos, con el objetivo de encontrar patrones, tendencias o reglas de repetición que expliquen el comportamiento de los datos recopilados. Estos patrones se pueden encontrar usando estadísticas o algoritmos de búsqueda cercana a la inteligencia artificial y redes neuronales. Por lo tanto, es de gran importancia ya que estos datos son el medio por el cual se pueden sacar conclusiones y, por lo tanto, se pueden transformar estos datos en información destacada, para que las entidades puedan realizar mejoras y resolver el problema. formas de ayudarlos a lograr sus objetivos. [5].

- Decision trees: Un árbol de decisión es un algoritmo que clasifica la información de una manera que produce un modelo en forma de árbol. Incluye un modelo de información esquemática que representa las diferentes alternativas, así como los posibles resultados de cada una. Este algoritmo se utiliza para clasificar, predecir y segmentar datos para obtener información analizable. [6].

- Education: Es la formación práctica y metódica que se le da a una persona en el camino del desarrollo y la madurez. Es un proceso por el cual la persona recibe las herramientas y conocimientos necesarios para ponerlos en práctica en la vida cotidiana. El aprendizaje de una persona comienza en la niñez, al ingresar a instituciones llamadas escuelas o colegios donde se ha recibido educación y capacitación previa se le inculca al niño identidades culturales, valores y valores para que en el futuro el niño se convierta en una buena persona. [7].

B. Indicadores de desempeño

A continuación, se realizará la definición de los conceptos usados en la búsqueda de los indicadores de

desempeño en las investigaciones realizadas en las bases de datos Scopus, ScienceDirect e IEEE XPLORE.

- **Performance indicators:** Permite identificar y valorar el estado en que se encuentra el estudiante con referencia a un conocimiento, valor, sentimiento, actitud, habilidad o destreza con lo que se convierte en un verdadero criterio de evaluación [8].

- **Academic performance:** Es la capacidad del alumno, que expresa lo que él ha aprendido a lo largo del proceso formativo [9].

- **Indicators:** Un indicador es una característica específica, observable y medible que puede ser usada para mostrar los cambios y progresos que está haciendo un programa hacia el logro de un resultado específico [10].

- **Factors:** Un factor es lo que contribuye a que se obtengan determinados resultados al caer sobre él la responsabilidad de la variación o de los cambios [11].

- **School:** Institución destinada a la enseñanza primaria o secundaria

- **Análisis:** identificar los componentes de un todo, separarlos y examinarlos para lograr acceder a sus principios más elementales [12].

- **Family:** grupo de personas que poseen un grado de parentesco y conviven como tal [13].

A continuación, las ecuaciones de búsqueda usadas en Scopus, ScienceDirect e IEEE:

C. Ecuación de búsqueda scopus

Ahora veremos las diferentes ecuaciones de búsqueda implementadas en la investigación acerca de los indicadores de desempeño y las técnicas de minería de datos para el análisis de datos.

- TITLE-ABS-KEY ("data mining" AND "decision trees" AND "academic performance" AND "education") AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016))
- (TITLE-ABS-KEY ("academic performance") AND TITLE-ABS-KEY (indicator) OR TITLE-ABS-KEY (factor) AND TITLE-

ABS-KEY (school)) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018))

D. Ecuación de búsqueda ScienceDirect

- Year(2016-2022) - Title, abstract, keywords("data mining" AND "decision trees" AND "academic performance" AND "education")
- Year(2018-2022)- Title, abstract, keywords (indicators, factors) - Title: Education; AND indicators AND "academic performance" AND factors AND school AND family AND analysis.

E. Ecuación de búsqueda IEEE xplora

- ("All Metadata":data mining) AND ("All Metadata":decision trees) AND ("All Metadata":academic performance) AND ("All Metadata":education) Filters Applied: 2016 – 2022
- (All Metadata":"academic performance") AND ("All Metadata":indicator) OR ("All Metadata":factor) AND ("All Metadata":school) AND ("All Metadata":analysis) AND ("Publication Title: performance) Filters Applied: 2018-2022

En las anteriores ecuaciones de búsqueda se aplicaron los temas principales como palabras claves, de este modo, se tendrán en cuenta documentos de las temáticas definidas, además, en la búsqueda realizada los trabajos considerados como revistas, artículos y libros están limitados a los que fueron publicados a partir del 2016.

III. DESARROLLO

A. Búsqueda de indicadores de desempeño

El rendimiento académico es una medida de las capacidades de un estudiante, que se puede expresar con lo aprendido en un periodo de tiempo. Rodríguez, Ordoñez e Hidalgo [14] afirman que, el acceso a herramientas tecnológicas de aprendizaje como computadoras y conexión a internet, el acceso a herramientas tecnológicas de aprendizaje como

computadoras y conexión a internet, el mayor nivel educativo de los padres de familia, la condición de ser varón y el estudiar en una institución educativa oficial urbana aumenta la probabilidad de obtener un mejor rendimiento académico.

El nivel socioeconómico es el conjunto de factores, como lo son los ingresos, el patrimonio y las condiciones generales del entorno en las que vive un ser humano. el componente social es un factor crucial para mejorar la calidad de la educación y que variables como las actividades extracurriculares apoyen la formación integral de los estudiantes para impactar positivamente en el rendimiento [15]. Devi, K., Ratnoo, S., & Bajaj, A. [16] afirman, que las variables socioeconómicas de los estudiantes, como la casta, la residencia y la ocupación del padre, impactan su rendimiento académico en el sexto grado. Por otro lado, Muelle [17] afirma, La condición social del alumno y la composición social de su escuela destacan como los factores que afectan mayormente el bajo rendimiento, asociados a factores contextuales como la repetición, la lengua materna, la matrícula oportuna, la dimensión de la escuela, el ausentismo y el género.

En Colombia la educación escolar se puede dividir en colegios públicos y privados, es decir que los públicos los sostiene económicamente el estado, mientras que en la privada se sostiene gracias a los mismos alumnos, por medio de sus padres o algún acudiente legal. Por otra parte, también se puede dividir en escuelas rurales y urbanas.

En Sudáfrica el sistema educativo se caracteriza por escuelas acomodadas y otras que no lo están. Las escuelas acomodadas se desempeñan a un ritmo mucho mejor, esto conlleva a que se vuelva un factor en el rendimiento académico de los estudiantes. [18] Por otra parte, Sarmiento Espinel, J. A., Silva Arias, A. C., & van Gameren, E, [19] afirman que, la desigualdad de oportunidades educativas ha aumentado con el tiempo. La resiliencia, es la capacidad que tiene una persona para superar circunstancias traumáticas, esta se relaciona positivamente con el rendimiento académico y otros factores como la relación profesor alumno, la participación de los padres y los métodos de estudio [20].

El rendimiento académico aumenta. Según JHU (Johns Hopkins University), aspectos como la comprensión y la fluidez lectora mejoran cuando hay participación de los padres, aún más si los papás dedican tiempo para leer con sus hijos, ya que los alumnos saben que sus papás están al pendiente, tratan de mejorar por ellos, se sienten más motivados a aprender y mejorar sus calificaciones. [21] Además, La relación que tienen las

variables sociofamiliares y no cognitivas sobre los estudiantes en cuanto a su rendimiento académico es un elemento muy importante para el éxito en la Educación Secundaria [22]. En Rusia un estudio realizado por Kotomina, O. v., & Sazhina, A. [23] afirma que, la revisión radica en la consideración de tres formas en que la familia impacta en desempeño del estudiante: el estatus socioeconómico de la familia, el capital social de la familia, y la participación de los padres en el proceso educativo. Las dos primeras formas han sido ampliamente estudiadas en la investigación, mientras que la participación de los padres a menudo se considera como un factor significativo en el rendimiento escolar.

Algo que afectó al mundo en general fue la pandemia, por temas de contagios, a la gran mayoría de los colegios el anuncio de cierre de clases los tomó por sorpresa y sin previa preparación para desarrollar su programa de educación a distancia. Fueron muy pocos los colegios los que ya tenían un programa de aprendizaje remoto listo para ser implementado [24]. Estudios realizados en China por Wang, Y., Xia, M., Guo, W., Xu, F., & Zhao, Y [25] afirman que, los padres dedicaron más del doble del tiempo normal a apoyar el aprendizaje y el desarrollo de sus hijos durante el período de COVID-19. Se encontró que los factores de apoyo y motivación de los padres son la contribución más efectiva en el desarrollo de las emociones positivas de los niños y el logro del aprendizaje. Por otra parte, otro estudio en China los resultados indicaron que el aprendizaje en línea no aumentó el rendimiento académico en el bachillerato rural y se observó una caída significativa del rendimiento en matemáticas e inglés [26].

A continuación, se muestra una tabla la cual resume cuales son los indicadores y cuales autores lo afirman:

| Indicadores | Autores | Título del artículo |
|----------------------|-------------------------------|--|
| Acceso a tecnologías | Rodríguez, Ordoñez e Hidalgo. | Determinantes del rendimiento académico de la educación media en el Departamento de Nariño, Colombia |
| Conexión a internet | Rodríguez, Ordoñez e Hidalgo. | Determinantes del rendimiento académico de la educación media en el Departamento de Nariño, Colombia |

| | | |
|-------------------------------|--|--|
| Nivel educativo de los padres | Rodríguez, Ordoñez e Hidalgo. | Determinantes del rendimiento académico de la educación media en el Departamento de Nariño, Colombia |
| Genero | Rodríguez, Ordoñez e Hidalgo. Muelle. | Determinantes del rendimiento académico de la educación media en el Departamento de Nariño, Colombia. Factores socioeconómicos y contextuales asociados al bajo rendimiento académico de alumnos peruanos en PISA 2015. |
| Institución (Urbana o rural) | Rodríguez, Ordoñez e Hidalgo. | Determinantes del rendimiento académico de la educación media en el Departamento de Nariño, Colombia |
| Ocupación del padre | Devi K, Ratnoo S, & Bajaj A. | Impact of Socio-Economic Factors on Students' Academic Performance: A Case Study of Jawahar Navodaya Vidyalaya |
| La residencia | Devi K, Ratnoo S, & Bajaj A. | Impact of Socio-Economic Factors on Students' Academic Performance: A Case Study of Jawahar Navodaya Vidyalaya |
| Condición | Velásquez, | Multidimensional |

| | | |
|----------------------------------|---|---|
| social | m y Crissien. Muelle | indicator to measure quality in education. Factores socioeconómicos y contextuales asociados al bajo rendimiento académico de alumnos peruanos en PISA 2015. |
| Composición social de la escuela | Muelle. Adebayo, K. A., Ntokozo, N., & Grace, N. Z. | Factores socioeconómicos y contextuales asociados al bajo rendimiento académico de alumnos peruanos en PISA 2015. Availability of Educational Resources and Student Academic Performances in South Africa. |
| Ausentismo | Muelle | Factores socioeconómicos y contextuales asociados al bajo rendimiento académico de alumnos peruanos en PISA 2015. |
| Participación de los padres | Sarmiento Espinel, J. A., Silva Arias, A. C., & van Gameren, E. Kotomina, O. v., & Sazhina | Evolution of the inequality of educational opportunities from secondary education to university. Influencia de los factores familiares sobre el desempeño de escolares y estudiantes: |

| | | |
|---|---|---|
| | | revisión de estudios extranjeros. |
| Desigualdad de oportunidades educativas | Sarmiento Espinel, J. A., Silva Arias, A. C., & van Gameren, E | Evolution of the inequality of educational opportunities from secondary education to university. |
| Estatus socioeconómico de la familia | Kotomina, O. v., & Sazhina | Influencia de los factores familiares sobre el desempeño de escolares y estudiantes: revisión de estudios extranjeros. |
| Capital social de la familia | Kotomina, O. v., & Sazhina | Influencia de los factores familiares sobre el desempeño de escolares y estudiantes: revisión de estudios extranjeros. |
| Pandemia | Wang, Y., Xia, M., Guo, W., Xu, F., & Zhao Zeng, L., & Luo, H. | Academic performance under COVID-19: The role of online learning readiness and emotional competence. Online Academic Performance during the COVID-19: Evidence from a Rural High School in Western China |
| Condiciones sociofamiliares | Rodríguez-Rodríguez, D., & | Academic performance of secondary education students |

| | | |
|--|-----------|---------------------------------|
| | Guzmán, R | in socio-familial risk contexts |
|--|-----------|---------------------------------|

Figura 1. Elaboración propia

B. Técnicas de minería de datos para la búsqueda de patrones.

Ahora bien, frente a la medición del desempeño de las diferentes instituciones educativas, se encuentran varios estudios en el cual se relacionan los diferentes indicadores y la manera de analizar dichos datos, según la investigación de Nuankaew y Sararat [27] realizada a 1859 estudiantes de la escuela Manchasuksa en el distrito de La Mancha Khiri, provincia de Khon Kaen, Tailandia, durante el año académico 2015-2020, en donde las herramientas de investigación están separadas en 2 secciones. La primera sección es un paso básico de análisis estadístico, este se compone de análisis de frecuencia, análisis de porcentaje, análisis de media y análisis de desviación estándar. Otra sección es la fase de análisis de minería de datos, que consiste en la técnica de discretización, la técnica de clasificación XGBoost (árbol de decisión, árboles potenciados por gradiente y random forest), análisis de rendimiento de matriz de confusión y análisis de rendimiento de validación cruzada.

Además, uno de los temas de enfoque en la investigación de Vasiliki Matzavela y Efthimios Alepis [28] trata de la Minería de Datos Educativos (EDM) la cual es una aplicación de Técnicas de Minería de Conocimiento a partir de datos educativos, y su objeto es analizar datos, con el fin de resolver problemas de investigación en el campo de la Educación. Sus datos provienen de diferentes fuentes, como bases de datos de sistemas educativos, sistemas de Internet.

Por otro lado, se evidencia un tema de gran importancia, los árboles de decisión pues son una de las técnicas usadas como modelo predictivo del Rendimiento Académico de los Estudiantes en entornos Inteligentes de M-Learning, en el área del aprendizaje automático y la ciencia de datos, es uno de los métodos más populares dentro de las técnicas de clasificación por ser fácil de entender e interpretar por medio de gráficas, así como también el manejo de datos numéricos y categóricos. Los sistemas M-Learning se consolidan recientemente como uno de los métodos de mayor interés para una educación más efectiva y un aprendizaje adaptativo proporcional a las habilidades de aprendizaje de cada estudiante.

Un clasificador de árbol de decisión es uno de los métodos de aprendizaje supervisado más utilizados para la exploración de datos, aproximando una función por regiones constantes a trozos, y no necesita información

previa de la distribución de los datos esto según Mitra S y Acharya T. [29]. En el estudio de Witten et al. [30] los modelos de árboles de decisión son comúnmente utilizados en la minería de datos para examinar los datos e inducir el árbol y sus reglas que serán utilizadas para hacer las predicciones. Según Rud [31] el verdadero propósito de los árboles de decisión es clasificar los datos en grupos distintos o ramas que generen la separación más fuerte en los valores de la variable dependiente, siendo superiores en identificar segmentos con un comportamiento deseado como la respuesta o la activación, proporcionando así una solución fácilmente interpretable.

Teniendo en cuenta lo anterior, la minería de datos cuenta con diferentes sistemas que contribuyen al descubrimiento de los factores principales para un mejor rendimiento académico, en las últimas décadas se han generado avances significativos de las nuevas tecnologías en el ámbito educativo entre las cuales la Minería de Datos Educativo- EDM, juega un papel indispensable para la búsqueda del mejoramiento pedagógico, permitiendo a los investigadores por medio de bases de datos agrupar variables que ayuden a identificar los factores que influyen en el rendimiento académico de los estudiantes, algunos de estos como la información demográfica de los estudiantes, la disposición o voluntad de aprendizaje y la interacción familiar, usando EDM como regresión lineal, regresión árbol, random forest y red neuronal.(Yucheng-jin y Xiaomeng Yang) [32].

Con los conceptos tratados por Han, J. and Kamber, M. [33] en los últimos años, el campo de la minería de datos se vuelve muy importante para diferentes industrias, corporaciones y empresas, etc. debido a su capacidad para utilizar una gran cantidad de datos que antes no tenían uso y respecto de los cuales se pueden realizar análisis, predicción de tendencias y patrones.

Para determinar los indicadores de desempeño se tendrán presentes las técnicas de minería de datos implementado la herramienta WEKA como lo hicieron los investigadores Sadiq Hussain y Neema Abdulaziz [34] para la selección de los atributos o factores y de este modo permitiendo clasificar la información donde los autores determinaron mediante los resultados que el algoritmo de clasificación random forest destaca en precisión. Para confirmar que tecnología es más óptima en el estudio del desempeño académico de los estudiantes se pondrán a prueba varios métodos de selección de características para así extraer los indicadores fundamentales tomando como ejemplo el artículo de Talha Mahboob, Mubbashar Mushtaq y Kamran Shaukat [35] donde en este estudio proponen un método novedoso para la medición del desempeño de las instituciones educativas haciendo uso de varios modelos

de aprendizaje automático como los árboles de decisión, bosque de rotación, bosque aleatorio, entre otros, ya que esto permite adaptar los diferentes factores a tratar según las necesidades del estudio.

Según en el artículo de Umair Shafique y Haseeb Qaiser [36] se puede centrar la investigación haciendo uso de tres modelos de procesos de minería de datos que son muy populares y que principalmente son empleados por expertos e investigadores en minería de datos los cuales son Knowledge Discovery Databases (KDD), CRISP-DM y SEMMA. De acuerdo con ello, en la investigación de Samsudin [37] propone hacer uso de una máquina de vectores de soporte el cual es un algoritmo de aprendizaje que sirve para determinar patrones de desempeño académico durante la pandemia de COVID-19.

En el siguiente artículo se explica el proceso de aplicación de la metodología CRISP-DM para detectar factores relacionados con el rendimiento académico de estudiantes colombianos quienes presentaron las pruebas saber 11. Principalmente se construye un repositorio o base de datos con la información socioeconómica y académica disponible por el ICFES, para luego realizar una limpieza de este, se realizó un modelo de clasificación basado en arboles de decisión para predecir los patrones asociados con el buen o bajo rendimiento académico Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. [38].

En la investigación que realizaron Pandey y Pal [39] de minería de datos utilizando clasificación de Naïve Bayes para analizar, clasificar y predecir estudiantes de alto y bajo rendimiento donde La clasificación de Naïve Bayes se usa como técnica probabilística simple, que asume que todos los atributos dados en un conjunto de datos son independientes entre sí, de ahí el nombre "Naïve". Realizaron esta investigación con una muestra de datos de estudiantes matriculados en un Diploma de Postgrado en Aplicaciones Informáticas Aplicaciones informáticas (PGDCA) en la Universidad Dr. R. M. L. Awadh, Faizabad, India.

De acuerdo con el artículo de Ajibade y Bahiah Binti Ahmad [40] para construir un modelo predictivo, se utilizan varias técnicas de la minería de datos, que son la clasificación, la regresión y la agrupación. Para la investigación de los dos autores mencionados eligieron parámetros como las notas internas, las notas de las sesiones y la puntuación de admisión además hicieron uso del algoritmo de aprendizaje SVM (máquina de vectores de soporte) desarrollado por Subaira [41] que sirve para manejar los desafíos del reconocimiento de patrones y la predicción como también para el análisis y mapeo de funciones, asimismo se usó el algoritmo random forest igual que en el trabajo investigativo de Sabzevari M. [42]. Ahora bien, este método trata de una

colección de algoritmos de árboles de decisión que no están correlacionados. Random Forest genera una gran cantidad de árboles de decisión a partir de subconjuntos del conjunto de datos a estudiar donde cada subconjunto proporciona un árbol de decisión. Ahora, cada modelo de árbol de decisión clasifica una instancia en una clase y así la clase más votada se toma como instancia, es decir, de reiterar o ser insistente en una orden dada, todo esto es conforme a lo dicho por Amrieh E. A. [43]. En la investigación de Moisa V. [44] utilizaron algunas técnicas de conjuntos como Bagging, Adaboosting y random forest para predecir el rendimiento académico de los estudiantes con mayor precisión.

La finalidad del trabajo de Contreras Leonardo y Fuentes Héctor [45] es predecir el rendimiento académico de estudiantes mediante técnicas de aprendizaje automático donde se analizan 324 variables con métodos de selección de características, con el objetivo de determinar las variables más destacadas. El modelo de predicción del rendimiento académico universitario es estudiado por medio de algoritmos supervisados como (KNN, SVC, Naive Bayes y árbol de decisión), los cuales son optimizados mediante lenguaje Python. Además, son implementados algoritmos de ensamble que permiten mejorar la exactitud de los clasificadores previos, también se implementan métodos Bagging (CART, Random Forest), métodos Boosting (AdaBoost, GBM, XGBoost).

A continuación, veremos una tabla resumiendo las técnicas de minerías de datos con sus autores:

| | | |
|------------------|-------------------|---|
| Minería de datos | Arbol de decisión | 27.Mitra S, Acharya T. Data Mining 2013. |
| | | 28.Witten, I. H., Frank, E., & Hall, M. A. (2011). |
| | | 36.Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. (2020). |
| | | 29. Rud, O. P. (2012). Data Mining Cookbook |
| | | 33. Alam, T. M., Mushtaq, M., & Shaukat, K. (2021). |
| | | 41.Amrieh E. A. 2017. Database Theory and Application 9 119-136. |
| | | 25.Nuankaew, P., & Sararat, W. (2022). Student Performance Prediction Model for Predicting Academic |
| | | 43. Contreras Bravo, L. E., Fuentes López, H. J., |

| | | |
|--|---------------------|--|
| | | & Rivas Trujillo, E. (2022). Análisis del rendimiento académico |
| | Random forest | 42.Moisa V. 2018 Journal of Mobile Embedded and Distributed Systems 5 70-77 |
| | | 41.Amrieh E. A. 2017. Database Theory and Application 9 119-136. |
| | | 39.Subaira A. 2016 IEEE 8th Int. Conf. on Intelligent System and Control (ISCO) 978 274-280 |
| | | [33]. Alam, T. M., Mushtaq, M., & Shaukat, K. (2021). |
| | | [32]. Hussain, S. (2018, 1 febrero). Educational Data Mining and Analysis of Students |
| | | [40.] Sabzevari M. 2018 Cornell Uni. arXiv preprint arXiv:1802.07877 |
| | | [25]. Nuankaew, P., & Sararat, W. (2022). Student Performance Prediction Model for Predicting Academic |
| | Naive Bayes | [43]. Contreras Bravo, L. E., Fuentes López, H. J., & Rivas Trujillo, E. (2022). Análisis del rendimiento académico |
| | | [37]. Pandey, U.K. and Pal, S., 2011. Data Mining: A prediction of performer or underperformer using classification. |
| | vectores de soporte | [35].Samsudin, N. A. M. (2021). Modeling Student's Academic Performance. |
| | | [39].Subaira A. 2016 IEEE 8th Int. Conf. on Intelligent System and Control (ISCO) 978 274-280 |
| | | [38].Ajibade, S. S. M., & Bahiah Binti Ahmad, N. |

que son varios los factores determinantes del alto o bajo desempeño educativo, dentro de estos se enmarcan el Familiar, Social y Económico, siendo el Familiar un indicador ampliamente determinante pues es el común denominador de la población estudiantil ya que se encuentra en edad joven, es decir, aún dependen de la ayuda económica de su núcleo familiar, partiendo de esta suposición y en concordancia con la información presente en desarrollo de este artículo se puede asumir que la Familia es el indicador principal del cual se desglosa el factor social y económico, de modo que una familia que se encuentra situada en una región poblacional en la cual cuentan con todos los recursos sociales y económicos para desarrollarse íntegramente reúne las condiciones para que el estudiante tenga un mejor rendimiento académico.

- Gracias a toda la investigación realizada en este artículo, podemos decir con certeza que las técnicas más usadas por los autores y que más ayudan al análisis de datos son los árboles de decisión y random forest.

REFERENCIAS

- [1] Lima, “MINISTERIO DE ECONOMÍA Y FINANZAS DIRECCIÓN GENERAL DEL PRESUPUESTO PÚBLICO Instructivo para la Formulación de Indicadores de Desempeño,” 2010.
- [2] “Conceptos de minería de datos | Microsoft Docs.” <https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions> (accessed May 26, 2022).
- [3] E. Serna, Desarrollo e Innovación en Ingeniería. IAI, 2019. Accedido el 10 de agosto de 2022. [En línea]. Disponible: <https://doi.org/10.5281/zenodo.3387679>
- [4]. Investigación [en línea]. (sin fecha). Publiciencia. [Consultado el 8 de julio de 2022]. Disponible en: <https://publiciencia.com/investigación>
- [5]. Bello. "¿Qué es el minado de Datos o Data Mininig? Técnicas y pasos a seguir". Thinking for Innovation. <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/> (accedido el 10 de agosto de 2022).
- [6] “Árboles de decisiones en la minería de datos - Conecta Software.” <https://conectasoftware.com/analytics/arboles-de-decisiones-en-la-mineria-de-datos/> (accessed May 09, 2022).
- [7] A. Sánchez. "¿Qué es la Educación?» Su Definición y Significado 2021". Concepto de - Definición de. <https://conceptodefinicion.de/educacion/> (accedido el 10 de agosto de 2022).
- [8] “Algunos conceptos importantes sobre educación en Colombia - ESE.” <https://eservicioseducativos.com/editorial/conceptos-importantes-sobre-la-educacion-en-colombia/> (accessed May 26, 2022).
- [9] “Definición de rendimiento académico - Qué es, Significado y Concepto.” <https://definicion.de/rendimiento-academico/> (accessed May 26, 2022).
- [10] “Indicadores.” <https://www.endvawnow.org/es/articulos/336-indicadores.html> (accessed May 26, 2022).
- [11] “Definición de Factores» Concepto en Definición ABC.” <https://www.definicionabc.com/general/factores.php> (accessed May 26, 2022).
- [12] “Definición de análisis - Qué es, Significado y Concepto.” <https://definicion.de/analisis/> (accessed May 26, 2022).
- [13] “Significado de Familia (Qué es, Concepto y Definición) - Significados.” <https://www.significados.com/familia/> (accessed May 26, 2022).
- [14] D. D. R. Rosero, R. E. O. Ortega, and M. E. H. Villota, “Academic performance determinants of high school students in the Department of Nariño, Colombia,” *Lecturas de Economía*, no. 94, pp. 87–126, Jan. 2021, doi: 10.17533/UDEA.LE.N94A341834.
- [15] J. V. Rodríguez, D. N. Rodado, T. Crissien Borrero, and A. Parody, “Multidimensional indicator to measure quality in education,” *International Journal of Educational Development*, vol. 89, Mar. 2022, doi: 10.1016/j.ijedudev.2021.102541.
- [16] K. Devi, S. Ratnoo, and A. Bajaj, “Impact of Socio-Economic Factors on Students’ Academic Performance: A Case Study of Jawahar Navodaya Vidyalaya,” *Lecture Notes in Networks and Systems*, vol. 419 LNNS, pp. 774–785, 2022, doi: 10.1007/978-3-030-96299-9_73.
- [17] L. Muelle, “Socioeconomic and contextual factors associated with low academic performance of peruvian students in PISA 2015,” *Apuntes*, vol. 47, no. 86, pp. 111–146, 2020, doi: 10.21678/APUNTES.86.943.
- [18] K. A. Adebayo, N. Ntokozi, and N. Z. Grace, “Availability of Educational Resources and Student Academic Performances in South Africa,” *Universal Journal of Educational Research*, vol. 8, no. 8, pp. 3768–3781, Aug. 2020, doi: 10.13189/ujer.2020.080858.
- [19] J. A. Sarmiento Espinel, A. C. Silva Arias, and E. van Gameren, “Evolution of the inequality of educational opportunities from secondary education to university,” *International Journal of Educational*

- Development, vol. 66, pp. 193–202, Apr. 2019, doi: 10.1016/j.ijedudev.2018.09.006.
- [20] G. Bester and N. Kuyper, “The Influence of Additional Educational Support on Poverty-Stricken Adolescents’ Resilience and Academic Performance,” <https://doi-org.ucundinamarca.basesdedatosezproxy.com/10.1080/18146627.2019.1689149>, vol. 17, no. 3, pp. 158–174, May 2020, doi: 10.1080/18146627.2019.1689149.
- [21] “La participación de los padres en la enseñanza — Observatorio | Instituto para el Futuro de la Educación.” <https://observatorio.tec.mx/edu-news/la-importancia-de-la-participacion-de-los-padres-en-la-educacion> (accessed Aug. 09, 2022).
- [22] D. Rodríguez-Rodríguez and R. Guzmán, “Academic performance of secondary education students in socio-familial risk contexts,” *Suma Psicológica*, vol. 28, no. 2, pp. 104–111, 2021, doi: 10.14349/sumapsi.2021.v28.n2.5.
- [23] O. v. Kotomina and A. I. Sazhina, “The Influence of Family Factors on the Academic Performance of Schoolchildren and University Students: Review of Foreign Studies,” *Education and Self Development*, vol. 16, no. 4, pp. 74–92, 2021, doi: 10.26907/esd.16.4.07.
- [24] “Así ha afectado el Covid-19 la educación en Colombia - Forbes Colombia.” <https://forbes.co/2020/04/30/actualidad/asi-ha-afectado-el-covid-19-la-educacion-en-colombia/> (accessed Aug. 09, 2022).
- [25] Y. Wang, M. Xia, W. Guo, F. Xu, and Y. Zhao, “Academic performance under COVID-19: The role of online learning readiness and emotional competence,” *Current Psychology*, 2022, doi: 10.1007/s12144-022-02699-7.
- [26] L. Zeng and H. Luo, “Online Academic Performance during the COVID-19: Evidence from a Rural High School in Western China,” *Proceedings - 2021 10th International Conference of Educational Innovation through Technology, EITT 2021*, pp. 112–116, 2021, doi: 10.1109/EITT53287.2021.00030.
- [27] Nuankaew, P., & Sararat, W. (2022). Student Performance Prediction Model for Predicting Academic Achievement of High School Students. *European Journal of Educational Research*, 11(2), 949–963. <https://doi.org/10.12973/eu-jer.11.2.949>
- [28] Matzavela, V., & Alepis, E. (2021, 5 octubre). E-Biblioteca Ucundinamarca. ScienceDirect. Recuperado 17 de abril de 2022, de <https://login.ucundinamarca.basesdedatosezproxy.com/login?url=https://www.sciencedirect.com/farticle%2fpii%2fS2666920X21000291%3fvia%253Dihub>
- [29] Mitra S, Acharya T. Data Mining. Multimedia, Soft Computing, and Bioinformatics. John Wiley & Sons, Inc., Hoboken, New Jersey; 2013.
- [30] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann Publishers.
- [31] Rud, O. P. (2012). *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management* (1.a ed.). John Wiley & Sons Inc.
- [32.] Jin, Y., & Yang, X. (2021). Educational Data Mining: Discovering Principal Factors for Better Academic Performance. 2021 the 3rd International Conference on Big Data Engineering and Technology (BDET). <https://doi.org/10.1145/3474944.3474945>
- [33]. Han, J. and Kamber, M. “Data Mining: Concepts and Techniques. Second Edition”, Morgan Kaufmann Publishers, San Francisco, 2010.
- [34]. Hussain, S. (2018, 1 febrero). Educational Data Mining and Analysis of Students’ Academic Performance Using WEKA | Hussain | Indonesian Journal of Electrical Engineering and Computer Science. Indonesian Journal. Recuperado 17 de abril de 2022, de <http://ijeecs.iaescore.com/index.php/IJECS/article/view/9746>
- [35]. Alam, T. M., Mushtaq, M., & Shaukat, K. (2021). A Novel Method for Performance Measurement of Public Educational Institutions Using Machine Learning Models. *Applied Sciences*, 11(19), 9296. <https://doi.org/10.3390/app11199296>
- [36]. Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.
- [37]. Samsudin, N. A. M. (2021). Modeling Student’s Academic Performance During Covid- 19 Based on Classification in Support Vector Machine. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(5), 1798-1804.
- [38]. Timarán-Pereira, R., Hidalgo-Troya, A., & Caicedo-Zambrano, J. (2020). Academic performance patterns of middle school students in the knowledge natural science test 11 with decision trees. *RISTI - Revista De Sistemas e Tecnologias De Informacao*, 2020(E32), 190-201. Retrieved from www.scopus.com
- [39]. Pandey, U.K. and Pal, S., 2011. Data Mining: A prediction of performer or underperformer using classification. (IJCSIT) *International Journal of Computer Science and Information Technologies*, Vol. 2 (2), 2011, 686- 690.
- [40]. Ajibade, S. S. M., & Bahiah Binti Ahmad, N. (2019). Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods.

- IOP Conference Series: Materials Science and Engineering, 551(1), 012061.
<https://doi.org/10.1088/1757-899x/551/1/012061>
- [41]. Subaira A. 2016 IEEE 8th Int. Conf. on Intelligent System and Control (ISCO) 978 274-280
- [42]. Sabzevari M. 2018 Cornell Uni. arXiv preprint arXiv:1802.07877
- [43]. Amrieh E. A. 2017 Int'l. Journal of Database Theory and Application 9 119-136.
- [44]. Moisa V. 2018 Journal of Mobile Embedded and Distributed Systems 5 70-77
- [45]. Contreras Bravo, L. E., Fuentes López, H. J., & Rivas Trujillo, E. (2022). Análisis del rendimiento académico mediante técnicas de aprendizaje automático con métodos de ensamble. Revista Boletín Redipe, 10(13), 171–190.
<https://doi.org/10.36260/rbr.v10i13.1737>

Análisis predictivo de indicadores de desempeño por medio de la técnica de árboles de decisión j48 para las pruebas saber 11 Colombia 2020

Luis Felipe Vergara Serrato, David Santiago Mahecha Hernández, y cesar Yesid Barahona Rodríguez

Resumen - En Colombia el ICFES (Instituto Colombiano para la Evaluación de la Educación) realiza las pruebas saber 11 para medir las aptitudes de los estudiantes en diferentes áreas como: matemáticas, inglés, lectura crítica, sociales y ciencias naturales.

Se realizó un análisis predictivo de las pruebas saber por medio de las técnicas de minería de datos, para este caso de estudio se usará la herramienta WEKA ya que permite realizar una investigación a detalle de las pruebas saber 11 en Colombia del año 2020. Por otro lado, debido al estudio ya realizado donde, por medio de diferentes autores, se tiene como base para este estudio los indicadores de desempeño que pueden influir en los estudiantes, por medio de árboles de decisión (donde estos son una estructura representativa y analítica de todos los eventos que puede surgir de una decisión tomada), obtendremos resultados que serán analizados detalladamente, con esto poder demostrar si en la presente investigación estos indicadores influyen en el rendimiento académico de los estudiantes de secundaria.

I. INTRODUCCION

Colombia tiene las pruebas saber 11 el cual es considerado el examen de estado de la educación media, siendo una herramienta de medición estándar que evalúa la calidad de la educación formal.

Los indicadores de desempeño son instrumentos que proporcionan información cuantitativa y cualitativa de un objeto de estudio los cuales establecen una relación con diferentes variables, que, al ser comparados, se logra detectar fallas o avances del objeto de estudio [1].

Para esta investigación se realizó análisis predictivo consiste en el uso de técnicas estadísticas, con el fin de predecir tendencias futuras, además, se tiene un enfoque descriptivo para cuantificar las relaciones entre los datos usados, estudiando los beneficios pasados con base en los desempeños.

Los árboles de decisión son una técnica de clasificación que permiten examinar los resultados y determinar visualmente el flujo del modelo. Los resultados visuales ayudan a encontrar subgrupos y relaciones específicos que quizás no encontremos con las estadísticas tradicionales. Además, es una técnica estadística para la predicción; Para la elaboración de estos, se usa el algoritmo J48 el cual es una implementación libre en java del algoritmo C4.5, que utiliza el concepto de entropía de la información para la selección de variables que mejor clasifiquen al objeto estudiado [2].

II. MÉTODO

En la revisión bibliográfica donde se recolectaron los indicadores de desempeño académico, se utilizó la metodología implementada en el artículo de investigación de [3] donde se hace uso de herramientas y ecuaciones de búsqueda para las bases de datos de Scopus, ScienceDirect e IEEE.

Luego de realizar lo anterior y analizar las publicaciones que hacen parte del tema a desarrollar, se tomaron en cuenta los siguientes indicadores: el primer indicador es el estatus socioeconómico de los padres el cual gira entorno a las condiciones económicas como los ingresos familiares, como también la educación de los padres [4]; el capital social consta de la participación de los padres, es decir, si se encuentran los dos padres juntos [4]; la pandemia se tiene presente ya que se puede tener en cuenta las falencias con el acceso a internet o a computadores,[5]; también hace parte la desigualdad de oportunidades que se compone de el estrato socioeconómico como la profesión de los padres [6]; la composición social de la escuela se toma en cuenta si son colegios públicos o privados, entre otros indicadores [7].

Ahora bien, continuando con la metodología CRISP-DM (Tabla 1) ya que es el modelo de referencia más utilizada en el

desarrollo de proyectos enfocados en la minería de datos [8] se accedió a la base de datos del Icfes donde se consiguieron los resultados de las pruebas saber 11, en cuanto a su contenido nos ofrece información familiar, social, socioeconómica, etc., para poder visualizar este contenido de manera organizada se debe realizar el proceso de transformación limpieza de los datos la cual consiste en organizar, limpiar y eliminar algunas inconsistencias que pudiera tener una base de datos. En ocasiones, cuando se extraen datos de un sitio web estos contienen errores, los cuales deben ser corregidos para utilizar la información. La manera en la que realizamos la limpieza consta del uso de funciones para corregir el uso de caracteres especiales y eliminar datos erróneos o vacíos.

| Fases CRISP-DM | Concepto |
|--------------------------|--|
| Comprensión del negocio | Comprensión de los objetivos de proyecto |
| Comprensión de los datos | Consta de la recopilación de datos inicial |
| Preparación de los datos | Transformación y la limpieza de datos |
| Modelado | Se selecciona y aplica las técnicas de modelado |
| Evaluación | Obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos. |
| Implementación | Generación de un informe |

Tabla 1. Pasos metodología CRISP-DM.

1. Obtención de datos.



Fig.1 Obtención de datos

2. Transformación y limpieza

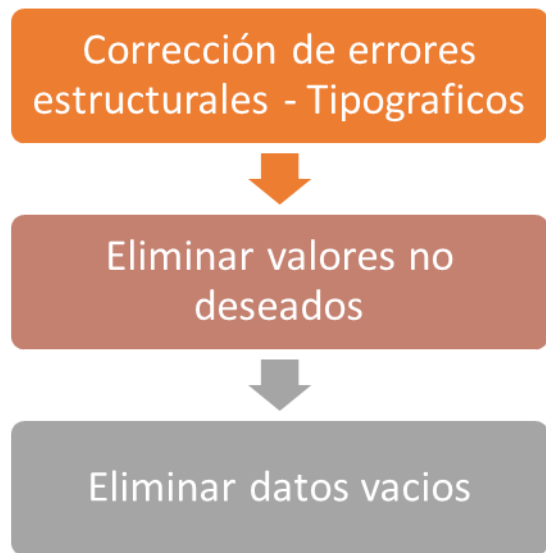


Fig.2 Transformación y limpieza de datos

3. Cargue de datos e implementación de los árboles de decisión

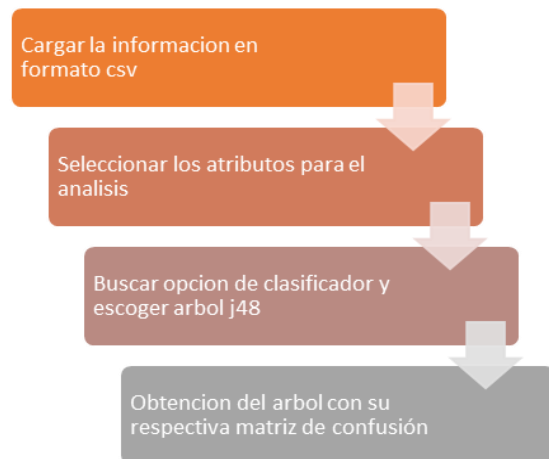


Fig.3 Cargue e implementación de los datos en los árboles de decisión

III. RESULTADOS Y DISCUSIÓN

1.

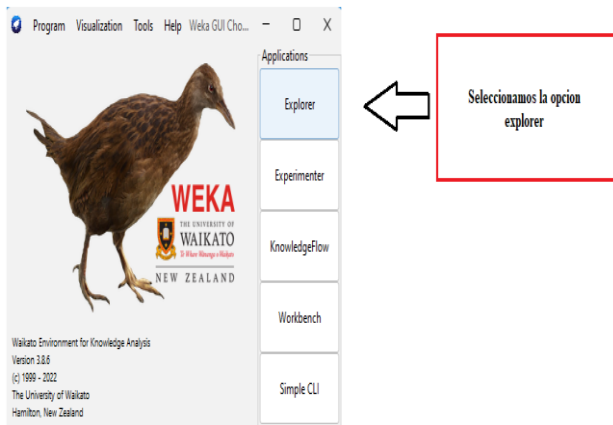


Fig.4 Inicio Weka.

2.



Fig.5 Selección del archivo.

3.

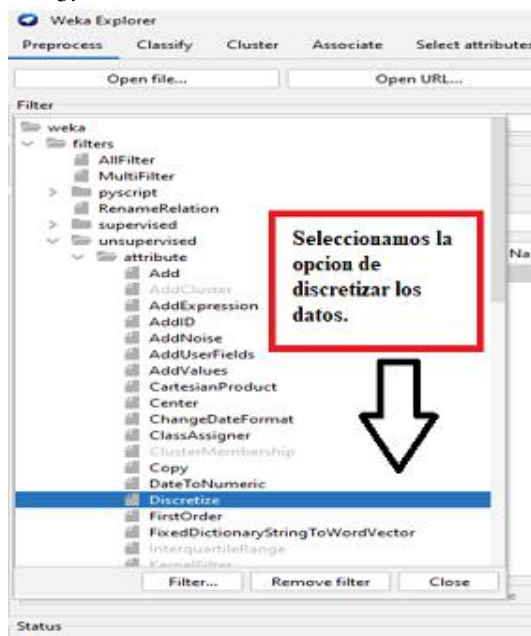


Fig.6 Selección del filtro.

El filtro de discretización es necesario cuando se quiere hacer una clasificación sobre un atributo numérico, en nuestro caso se usó para que el atributo de puntaje global se dividiera en rangos.

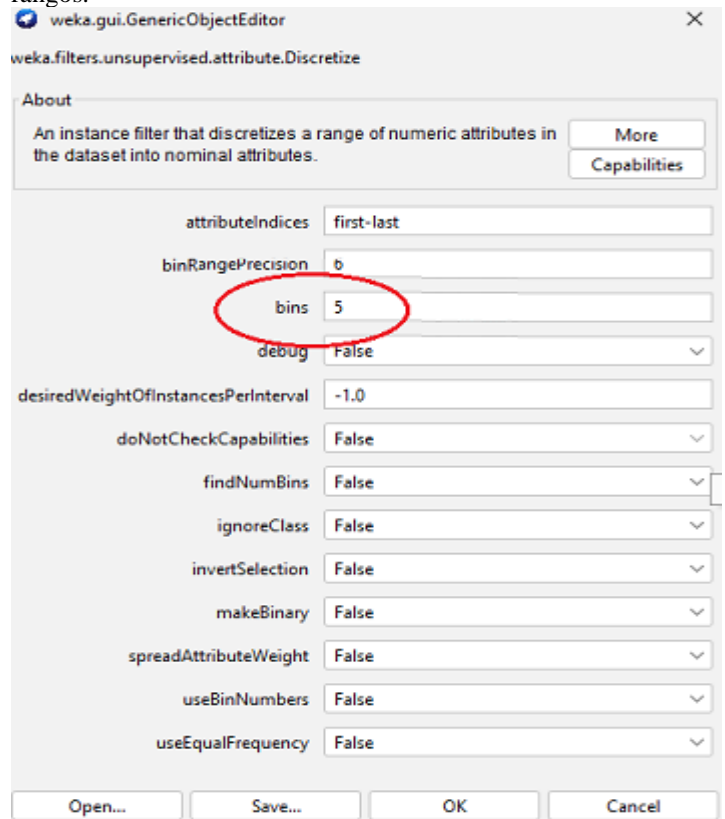


Fig.7 Selección de bis.

En la opción de bins podemos escoger en cuantos rangos vamos a dividir los atributos numéricos.

4.

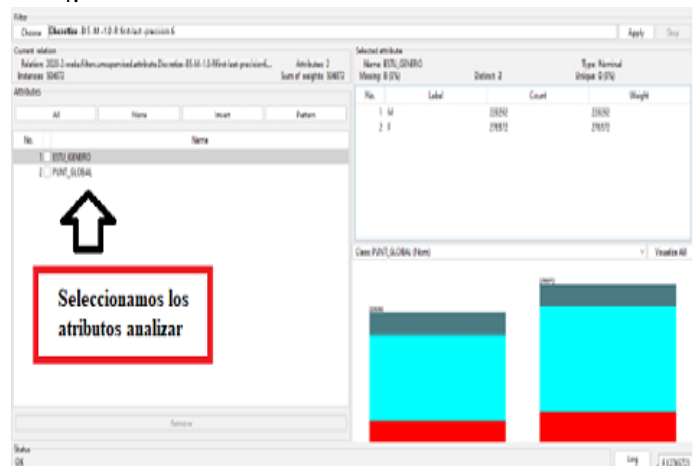


Fig.8 Selección de atributos.

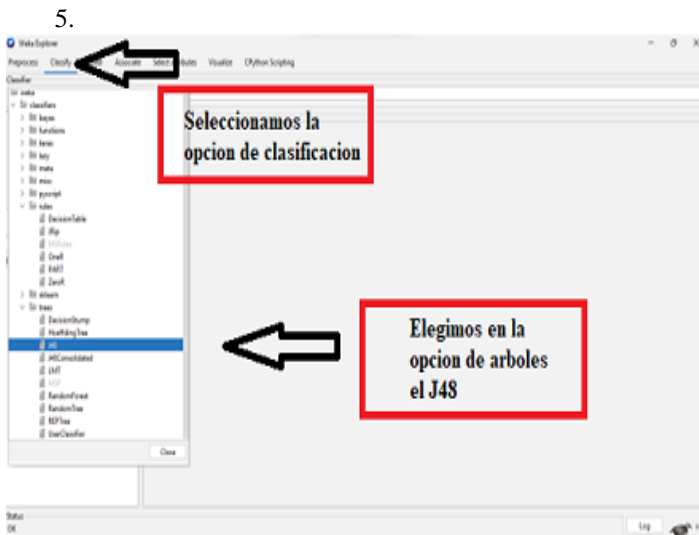


Fig.9 Selección del clasificador.

Se hizo uso de los árboles de decisión j48 que además de su gran reputación, permite la selección del atributo principal el cual da lugar a la construcción de dicho árbol buscando y dividiendo en subcategorías o subconjuntos donde haya relación con los demás atributos seleccionados, además, tienen como ventaja que sus resultados son fáciles de interpretar, se calculan rápido y se puede aplicar a variables continuas o categóricas [9].

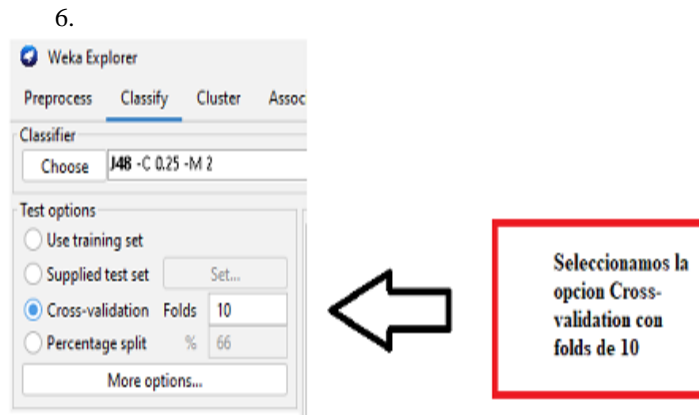


Fig.10 Selección de la prueba.

“Esta opción permite hacer una validación cruzada de K iteraciones. Los datos de la muestra se dividen en K subconjuntos. En cada una de las K iteraciones, el subconjunto K se utiliza como datos de prueba y el resto (K-1), se utiliza como datos de entrenamiento. Una vez se termina el proceso de validación cruzada, se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado. Este método resulta preciso puesto que evalúa K combinaciones de prueba y entrenamiento, aunque desde el punto de vista computacional resulta más lento mientras más iteraciones se añadan. La elección del número de iteraciones debe depender de la medida del conjunto de datos. El programa utiliza por defecto 10 iteraciones.” [10]

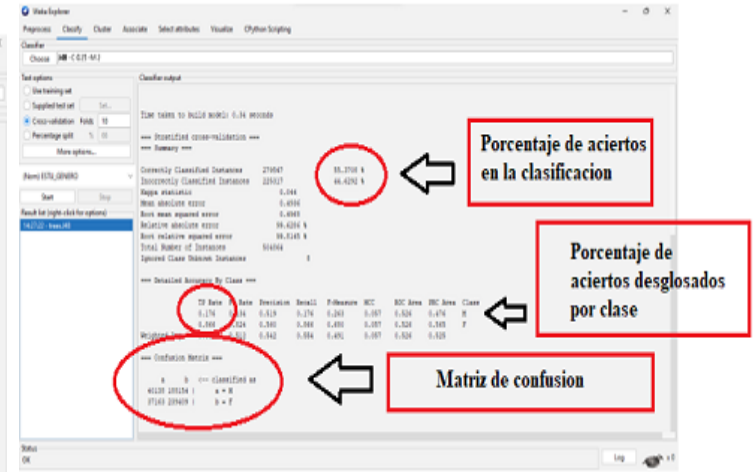


Fig.11 Información del clasificador.

Análisis de los indicadores:

- **Genero:**
✓ **2018:**

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 303544 | 55.1964 % |
| Incorrectly Classified Instances | 246390 | 44.8036 % |
| Kappa statistic | 0.0654 | |
| Mean absolute error | 0.4939 | |
| Root mean squared error | 0.4969 | |
| Relative absolute error | 99.4415 % | |
| Root relative squared error | 99.7207 % | |
| Total Number of Instances | 549934 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| M | 0.283 | 0.220 | 0.522 | 0.283 | 0.367 | 0.073 | 0.533 | 0.489 | M |
| F | 0.780 | 0.717 | 0.562 | 0.780 | 0.653 | 0.073 | 0.533 | 0.561 | F |
| Weighted Avg. | 0.552 | 0.489 | 0.544 | 0.552 | 0.522 | 0.073 | 0.533 | 0.528 | |

=== Confusion Matrix ===

| a | b | ←-- classified as | |
|-------|--------|-------------------|--|
| 71349 | 181079 | a = M | |
| 65311 | 232195 | b = F | |

Fig.12 Informe suministrado del árbol de decisión J48 indicador género 2018.

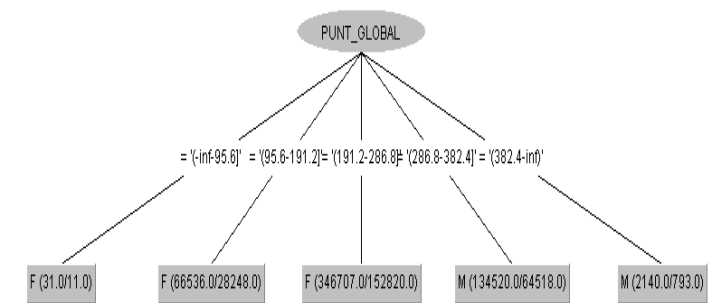


Fig.13 Árbol de decisión J48 indicador genero 2018.

7.

✓ 2020:

```
Classifier output

Time taken to build model: 0.34 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      279547      55.3708 %
Incorrectly Classified Instances    225317      44.6292 %
Kappa statistic                    0.044
Mean absolute error                0.4936
Root mean squared error            0.4968
Relative absolute error            99.6286 %
Root relative squared error        99.8145 %
Total Number of Instances         504864
Ignored Class Unknown Instances    8

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
-----
0.176  0.134  0.519     0.176  0.263     0.057  0.526   0.476   M
0.866  0.824  0.560     0.866  0.680     0.057  0.526   0.565   F
Weighted Avg.  0.554  0.512  0.542     0.554  0.491     0.057  0.526   0.525

=== Confusion Matrix ===

      a  b  <-- classified as
40138 188154 | a = M
37163 239409 | b = F
```

Fig.14 Informe suministrado del árbol de decisión J48 indicador género 2020.

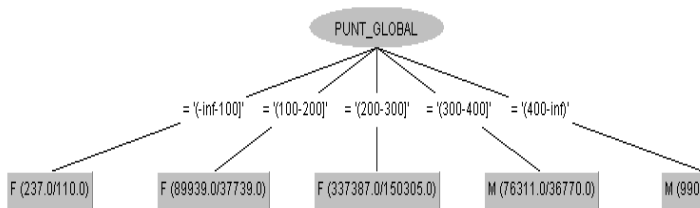


Fig.15 Árbol de decisión J48 indicador genero 2020.

Análisis: Podemos observar que en ambos años el género femenino obtiene puntajes de 300 hacia abajo, mientras que el género masculino obtiene resultados de 300 hacia arriba. En este caso vemos que el porcentaje de acierto que nos genera esta clasificación está en aproximadamente el 55%.

• Educación del padre:

✓ 2018:

```
Correctly Classified Instances      351252      63.8717 %
Incorrectly Classified Instances    199602      36.1283 %
Kappa statistic                    0.0387
Mean absolute error                0.2013
Root mean squared error            0.3167
Relative absolute error            95.3036 %
Root relative squared error        97.469 %
Total Number of Instances         549934

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
-----
0.000  0.000  ?          0.000  ?          ?       0.599  0.000   '(-inf-95.6)'
0.000  0.000  ?          0.000  ?          ?       0.620  0.177   '(95.6-191.2)'
0.995  0.965  0.637     0.995  0.777     0.113  0.596  0.691   '(191.2-286.8)'
0.047  0.006  0.713     0.047  0.089     0.140  0.676  0.414   '(286.8-382.4)'
0.000  0.000  ?          0.000  ?          ?       0.819  0.026   '(382.4-inf)'
Weighted Avg.  0.639  0.610  ?          0.639  ?          ?       0.619  0.558

=== Confusion Matrix ===

      a  b  c  d  e  <-- classified as
0  0  31  0  0  | a = '(-inf-95.6)'
0  0  66318  218  0  | b = '(95.6-191.2)'
0  0  344871  1836  0  | c = '(191.2-286.8)'
0  0  128139  6381  0  | d = '(286.8-382.4)'
0  0  1627  513  0  | e = '(382.4-inf)'
```

Fig.16 Informe suministrado del árbol de decisión J48 indicador educación del padre 2018.

```
FAMI_EDUCACIONPADRE = Ninguno: '(191.2-286.8)' (20094.78/7536.91)
FAMI_EDUCACIONPADRE = Primaria incompleta: '(191.2-286.8)' (110193.57/31863.46)
FAMI_EDUCACIONPADRE = Tecnica o tecnologica completa: '(191.2-286.8)' (37205.54/16548.77)
FAMI_EDUCACIONPADRE = Primaria completa: '(191.2-286.8)' (61616.5/19058.33)
FAMI_EDUCACIONPADRE = Secundaria (Bachillerato) completa: '(191.2-286.8)' (122404.4/44165.37)
FAMI_EDUCACIONPADRE = Tecnica o tecnologica incompleta: '(191.2-286.8)' (11050.79/4473.52)
FAMI_EDUCACIONPADRE = No sabe: '(191.2-286.8)' (35203.01/13322.75)
FAMI_EDUCACIONPADRE = Secundaria (Bachillerato) incompleta: '(191.2-286.8)' (77909.36/25208.17)
FAMI_EDUCACIONPADRE = Educacion profesional completa: '(191.2-286.8)' (48599.53/25816.68)
FAMI_EDUCACIONPADRE = Educacion profesional incompleta: '(191.2-286.8)' (10386.85/5560.26)
FAMI_EDUCACIONPADRE = Postgrado: '(286.8-382.4)' (9566.8/3113.03)
FAMI_EDUCACIONPADRE = No Aplica: '(191.2-286.8)' (5702.88/2326.6)
```

Fig.17 Árbol de decisión J48 indicador educación del padre 2018.

✓ 2020:

```
Correctly Classified Instances      339435      67.2319 %
Incorrectly Classified Instances    165437      32.7681 %
Kappa statistic                    0.0405
Mean absolute error                0.1918
Root mean squared error            0.3096
Relative absolute error            96.1439 %
Root relative squared error        98.004 %
Total Number of Instances         504872

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
-----
0.000  0.000  ?          0.000  ?          ?       0.769  0.001   '(-inf-100)'
0.000  0.000  ?          0.000  ?          ?       0.634  0.264   '(100-200)'
0.990  0.964  0.674     0.990  0.802     0.090  0.557  0.697   '(200-300)'
0.072  0.009  0.577     0.072  0.129     0.165  0.698  0.306   '(300-400)'
0.000  0.000  ?          0.000  ?          ?       0.782  0.009   '(400-inf)'
Weighted Avg.  0.672  0.645  ?          0.672  ?          ?       0.592  0.559

=== Confusion Matrix ===

      a  b  c  d  e  <-- classified as
0  0  237  0  0  | a = '(-inf-100)'
0  0  89583  357  0  | b = '(100-200)'
0  0  333913  3480  0  | c = '(200-300)'
0  0  70790  5522  0  | d = '(300-400)'
0  0  781  209  0  | e = '(400-inf)'
```

Fig.18 Informe suministrado del árbol de decisión J48 indicador educación del padre 2020.

```
FAMI_EDUCACIONPADRE = Secundaria Bachillerato completa: '(200-300)' (123499.56/37026.87)
FAMI_EDUCACIONPADRE = Postgrado: '(300-400)' (9820.68/4266.76)
FAMI_EDUCACIONPADRE = Tecnica o tecnologica completa: '(200-300)' (33946.42/11286.21)
FAMI_EDUCACIONPADRE = Primaria completa: '(200-300)' (47143.98/14403.87)
FAMI_EDUCACIONPADRE = Primaria incompleta: '(200-300)' (94520.96/29191.96)
FAMI_EDUCACIONPADRE = Ninguno: '(200-300)' (21347.25/9385.25)
FAMI_EDUCACIONPADRE = Educacion profesional incompleta: '(200-300)' (10501.19/4152.98)
FAMI_EDUCACIONPADRE = No sabe: '(200-300)' (30660.88/9747.93)
FAMI_EDUCACIONPADRE = Educacion profesional completa: '(200-300)' (47971.27/20257.75)
FAMI_EDUCACIONPADRE = No Aplica: '(200-300)' (7939.27/2617.59)
FAMI_EDUCACIONPADRE = Secundaria Bachillerato incompleta: '(200-300)' (67029.62/19940.2)
FAMI_EDUCACIONPADRE = Tecnica o tecnologica incompleta: '(200-300)' (10490.92/3286.89)
```

Fig.19 Árbol de decisión J48 indicador educación del padre 2020.

Análisis: En ambos años podemos observar que el puntaje en casi todas las opciones es el mismo, menos cuando el padre realizo un postgrado, notamos que esto influye positivamente ya que en el 2018 los estudiantes están en el rango de (286.8-

382.4) en comparación al resto que todos están en el rango de (191.2-286.8) y en le año 2020, observamos que los estudiantes están en el rango de (300-400), mientras que el resto está en (200-300). En esta clasificación se obtuvo un porcentaje de acierto en el 2018 del 63%, mientras que en el 2020 un porcentaje de 67%.

• **Educación de la madre:**
 ✓ **2018:**

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 351211 | 63.8642 % |
| Incorrectly Classified Instances | 198723 | 36.1358 % |
| Kappa statistic | 0.0408 | |
| Mean absolute error | 0.2003 | |
| Root mean squared error | 0.3159 | |
| Relative absolute error | 94.8501 % | |
| Root relative squared error | 97.2222 % | |
| Total Number of Instances | 549934 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-----------------|
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.632 | 0.000 | '(-inf-95.6]' |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.641 | 0.186 | '(95.6-191.2]' |
| | 0.993 | 0.962 | 0.638 | 0.993 | 0.777 | 0.111 | 0.600 | 0.691 | '(191.2-286.8]' |
| | 0.052 | 0.008 | 0.683 | 0.052 | 0.097 | 0.140 | 0.679 | 0.414 | '(286.8-382.4]' |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.829 | 0.024 | '(382.4-inf)' |
| Weighted Avg. | 0.639 | 0.608 | ? | 0.639 | ? | ? | 0.625 | 0.560 | |

=== Confusion Matrix ===

| a | b | c | d | e | <-- classified as |
|---|---|--------|------|---|---------------------|
| 0 | 0 | 31 | 0 | 0 | a = '(-inf-95.6]' |
| 0 | 0 | 66286 | 250 | 0 | b = '(95.6-191.2]' |
| 0 | 0 | 344221 | 2486 | 0 | c = '(191.2-286.8]' |
| 0 | 0 | 127530 | 6990 | 0 | d = '(286.8-382.4]' |
| 0 | 0 | 1635 | 505 | 0 | e = '(382.4-inf)' |

Fig.20 Informe suministrado del árbol de decisión J48 indicador educación de la madre 2018.

FAMI_EDUCACIONMADRE = Primaria incompleta: '(191.2-286.8]' (86678.46/25384.87)
 FAMI_EDUCACIONMADRE = Secundaria (Bachillerato) incompleta: '(191.2-286.8]' (84163.67/25707.07)
 FAMI_EDUCACIONMADRE = Tecnica o tecnologica incompleta: '(191.2-286.8]' (16432.16/6406.77)
 FAMI_EDUCACIONMADRE = Secundaria (Bachillerato) completa: '(191.2-286.8]' (142672.86/49359.76)
 FAMI_EDUCACIONMADRE = No sabe: '(191.2-286.8]' (9780.21/4282.32)
 FAMI_EDUCACIONMADRE = Tecnica o tecnologica completa: '(191.2-286.8]' (55092.37/23639.73)
 FAMI_EDUCACIONMADRE = Primaria completa: '(191.2-286.8]' (61710.12/18740.34)
 FAMI_EDUCACIONMADRE = Educacion profesional completa: '(191.2-286.8]' (55816.99/29649.54)
 FAMI_EDUCACIONMADRE = Ninguno: '(191.2-286.8]' (12894.57/5062.85)
 FAMI_EDUCACIONMADRE = Educacion profesional incompleta: '(191.2-286.8]' (12836.86/6516.43)
 FAMI_EDUCACIONMADRE = Postgrado: '(286.8-382.4]' (10934.47/3861.66)
 FAMI_EDUCACIONMADRE = No Aplica: '(191.2-286.8]' (921.27/463.64)

Fig.21 Árbol de decisión J48 indicador educación de la madre 2018

✓ **2020:**

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 339236 | 67.1925 % |
| Incorrectly Classified Instances | 165636 | 32.8075 % |
| Kappa statistic | 0.0442 | |
| Mean absolute error | 0.1909 | |
| Root mean squared error | 0.3088 | |
| Relative absolute error | 95.6948 % | |
| Root relative squared error | 97.7668 % | |
| Total Number of Instances | 504872 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|--------------|
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.683 | 0.001 | '(-inf-100]' |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.651 | 0.275 | '(100-200]' |
| | 0.987 | 0.959 | 0.675 | 0.987 | 0.801 | 0.090 | 0.566 | 0.704 | '(200-300]' |
| | 0.081 | 0.012 | 0.552 | 0.081 | 0.142 | 0.169 | 0.711 | 0.311 | '(300-400]' |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.791 | 0.009 | '(400-inf)' |
| Weighted Avg. | 0.672 | 0.643 | ? | 0.672 | ? | ? | 0.604 | 0.566 | |

=== Confusion Matrix ===

| a | b | c | d | e | <-- classified as |
|---|---|--------|------|---|-------------------|
| 0 | 0 | 237 | 0 | 0 | a = '(-inf-100]' |
| 0 | 0 | 89488 | 452 | 0 | b = '(100-200]' |
| 0 | 0 | 333029 | 4364 | 0 | c = '(200-300]' |
| 0 | 0 | 70105 | 6207 | 0 | d = '(300-400]' |
| 0 | 0 | 774 | 216 | 0 | e = '(400-inf)' |

Fig.22 Informe suministrado del árbol de decisión J48 indicador educación de la madre 2020.

FAMI_EDUCACIONMADRE = Secundaria Bachillerato completa: '(200-300]' (141424.7/41352.66)
 FAMI_EDUCACIONMADRE = Educacion profesional completa: '(200-300]' (57890.35/24006.71)
 FAMI_EDUCACIONMADRE = Tecnica o tecnologica incompleta: '(200-300]' (15615.11/4471.07)
 FAMI_EDUCACIONMADRE = Secundaria Bachillerato incompleta: '(200-300]' (70052.22/20229.6)
 FAMI_EDUCACIONMADRE = Primaria completa: '(200-300]' (47249.1/15049.31)
 FAMI_EDUCACIONMADRE = Primaria incompleta: '(200-300]' (73245.51/24172.13)
 FAMI_EDUCACIONMADRE = Postgrado: '(300-400]' (11547.45/5301.52)
 FAMI_EDUCACIONMADRE = Ninguno: '(200-300]' (12702.3/5915.4)
 FAMI_EDUCACIONMADRE = Educacion profesional incompleta: '(200-300]' (12967.38/5003.1)
 FAMI_EDUCACIONMADRE = Tecnica o tecnologica completa: '(200-300]' (51873.63/16182.42)
 FAMI_EDUCACIONMADRE = No Aplica: '(200-300]' (1202.11/562.25)
 FAMI_EDUCACIONMADRE = No sabe: '(200-300]' (9102.13/3541.72)

Fig.23 Árbol de decisión J48 indicador educación de la madre 2020.

Análisis: Al igual que el indicador anterior, notamos el mismo caso, el cual es que influye positivamente el hecho de que la madre tenga una educación de postgrado con los siguientes resultados, en el 2018 se obtiene un rango de (286.8-382.4) en comparación al resto que todos están en el rango de (191.2-286.8) y en el año 2020, observamos que los estudiantes están en el rango de (300-400), mientras que el resto está en (200-300). En esta clasificación se obtuvo un porcentaje de acierto en el 2018 del 63%, mientras que en el 2020 un porcentaje de 67%.

• **Tiene internet:**
 ✓ **2018:**

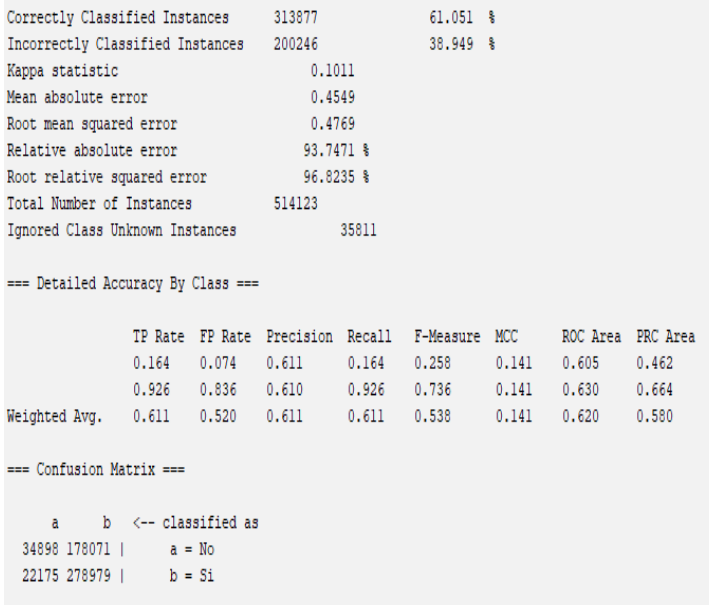


Fig.24 Informe suministrado del árbol de decisión J48 indicador el estudiante tiene internet 2018.

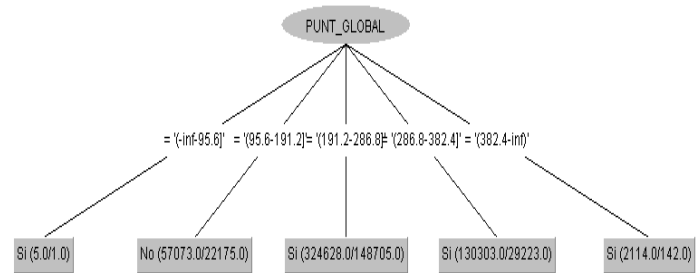


Fig.25 Árbol de decisión J48 indicador el estudiante tiene internet 2018.

✓ **2020:**

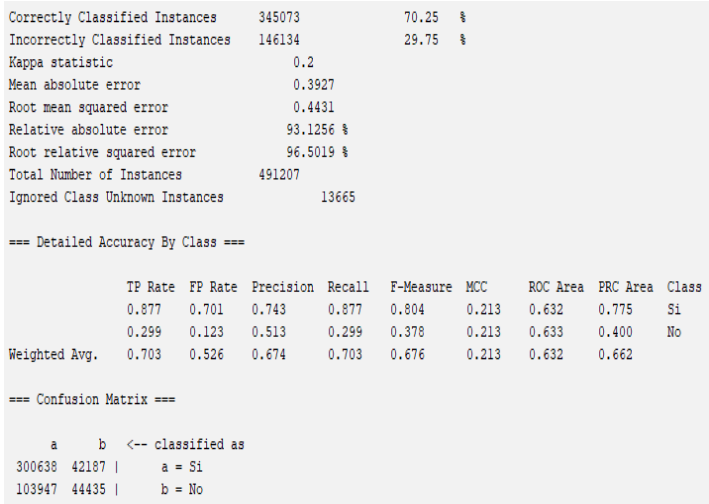


Fig.26 Informe suministrado del árbol de decisión J48 indicador el estudiante tiene internet 2020.

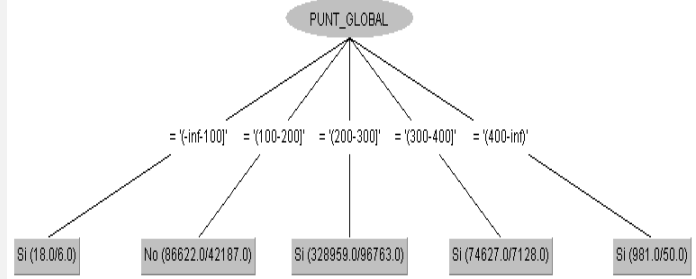


Fig.27 Árbol de decisión J48 indicador el estudiante tiene Internet 2020.

Análisis: En este indicador vemos claramente en ambos años que el hecho de tener internet influye positivamente en los resultados. En el 2018 obtuvimos un 63% de acierto en la clasificación, mientras que en el 2020 un 70%.

• **Tiene computador:**
 ✓ **2018:**

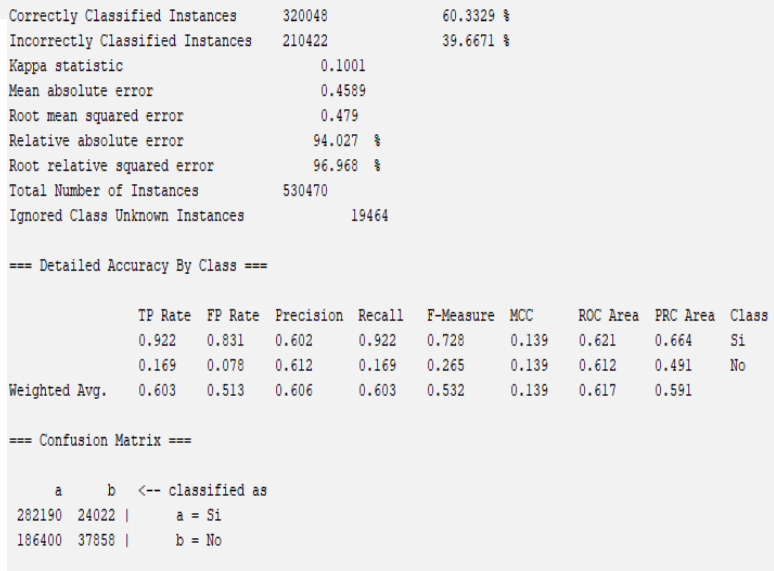


Fig.28 Informe suministrado del árbol de decisión J48 indicador estudiante tiene computador 2018.

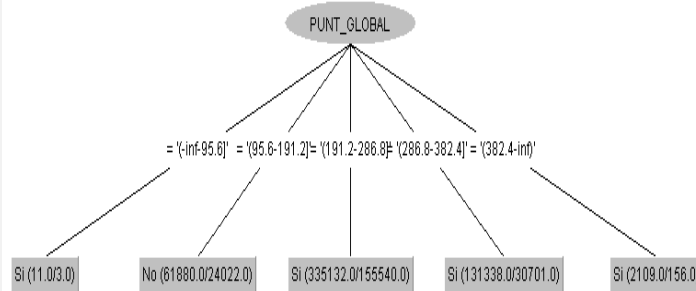


Fig.29 Árbol de decisión J48 indicador el estudiante tiene computador 2018.

✓ 2020:

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 310428 | 64.0483 % |
| Incorrectly Classified Instances | 174250 | 35.9517 % |
| Kappa statistic | 0.1669 | |
| Mean absolute error | 0.4439 | |
| Root mean squared error | 0.4711 | |
| Relative absolute error | 92.9315 % | |
| Root relative squared error | 96.4014 % | |
| Total Number of Instances | 484678 | |
| Ignored Class Unknown Instances | 20194 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| | 0.884 | 0.733 | 0.649 | 0.884 | 0.749 | 0.194 | 0.625 | 0.692 | Si |
| | 0.267 | 0.116 | 0.599 | 0.267 | 0.369 | 0.194 | 0.623 | 0.477 | No |
| Weighted Avg. | 0.640 | 0.490 | 0.630 | 0.640 | 0.599 | 0.194 | 0.624 | 0.607 | |

=== Confusion Matrix ===

| a | b | <-- classified as |
|--------|-------|-------------------|
| 259387 | 34123 | a = Si |
| 140127 | 51041 | b = No |

Fig.30 Informe suministrado del árbol de decisión J48 indicador estudiante tiene computador 2020.

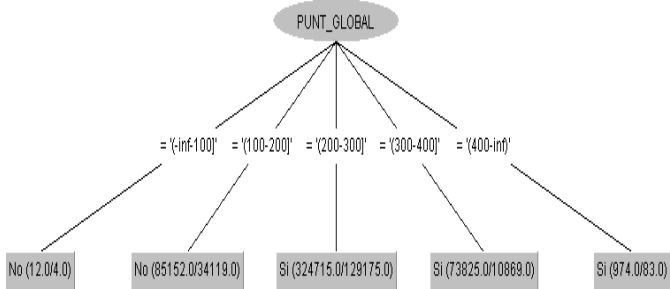


Fig.31 Árbol de decisión J48 indicador el estudiante tiene computador 2020.

Análisis: En este indicador vemos claramente en ambos años que el hecho de tener computador influye positivamente en los resultados. En el 2018 obtuvimos un 60% de acierto en la clasificación, mientras que en el 2020 un 64%.

IV. CONCLUSIONES

Nuestro análisis predictivo de los resultados conseguidos con el árbol de decisión J48 para determinar si los indicadores analizados influyen en el desempeño académico de los estudiantes colombianos en las pruebas de estado para la educación media saber 11° en los años 2018 y 2020, señalando que estos si tienen influencia no solo con el análisis realizado con los datos del ICFES sino también con el respaldo teórico. En nuestro primer indicador notamos una brecha de genero bastante marcada donde el género masculino tiende a obtener mejores resultados en estas pruebas que el género femenino, En el segundo y tercer indicador el cual es la educación alcanzada por los padres notamos que en ambos casos donde alguno de los 2 tuviese un postgrado tiende a que los estudiantes obtengan un mejor rendimiento en las pruebas y los 2 últimos indicadores los cuales son si el estudiante cuenta con internet y si cuenta con computador, vemos que en ambos casos influye positivamente el hecho de obtener alguno de estos 2 estos.

Por otro lado, en el análisis del COVID el cual se conforma por tener acceso a internet y tener computador en una comparación con el año donde empieza todo siendo 2020 con uno anterior como 2018 se evidencia que no hay gran influencia en cuanto a los resultados obtenidos.

REFERENCIAS

[1] Lima, “MINISTERIO DE ECONOMÍA Y FINANZAS DIRECCIÓN GENERAL DEL PRESUPUESTO PÚBLICO Instructivo para la Formulación de Indicadores de Desempeño,” 2010.

[2] Vanesa Berlanga Silvente and María José Rubio Hurtado, “Cómo aplicar árboles de decisión en SPSS.,” 2013, Accessed: Aug. 20, 2022. [Online]. Available: <http://diposit.ub.edu/dspace/bitstream/2445/43762/1/618361.pdf>

[3] Mahecha Hernandez David Santiago and Vergara Serrato Luis Felipe, “Revisión literaria de indicadores que influyen en la educación mediante técnicas de minería de datos”.

[4] O. v. Kotomina and A. I. Sazhina, “The Influence of Family Factors on the Academic Performance of Schoolchildren and University Students: Review of Foreign Studies,” Education and Self Development, vol. 16, no. 4, pp. 74–92, 2021, doi: 10.26907/esd.16.4.07.

[5] Y. Wang, M. Xia, W. Guo, F. Xu, and Y. Zhao, “Academic performance under COVID-19: The role of online learning readiness and emotional competence,” Current Psychology, 2022, doi: 10.1007/s12144-022-02699-7.

[6] J. A. Sarmiento Espinel, A. C. Silva Arias, and E. van Gameren, “Evolution of the inequality of educational opportunities from secondary education to university,” International Journal of Educational Development, vol. 66, pp. 193–202, Apr. 2019, doi: 10.1016/j.ijedudev.2018.09.006.

[7] K. A. Adebayo, N. Ntokozo, and N. Z. Grace, “Availability of Educational Resources and Student Academic Performances in South Africa,” Universal Journal of Educational Research, vol. 8, no. 8, pp. 3768–3781, Aug. 2020, doi: 10.13189/ujer.2020.080858.

[8] J. H. Orallo, *Introducción a la minería*. PRENTICE HALL, 2008.

[9] Cardona Carlos and Garcia Nancy, “Análisis de datos mediante el algoritmo de clasificación J48, sobre un cluster en la nube de AWS,” Nov. 2016. <https://revistas.udistrital.edu.co/index.php/REDES/article/view/11985/12615> (accessed Nov. 13, 2022).

[10] Rodríguez Manzanares, C. (2018). Análisis de los métodos de predicción aplicados a los desvíos en el sistema eléctrico ibérico. (Trabajo Fin de Máster Inédito). Universidad de Sevilla, Sevilla.

Agradecimientos

Este trabajo se desarrolló gracias al apoyo de la Universidad de Cundinamarca y el grupo de Investigación GISTFA (Grupo de Investigación de Sistemas y Tecnología de Facatativá), Colombia

Conflicto de Intereses

Los autores consideran que no existen conflictos de intereses.

Contribución de los autores

Luis Felipe Vergara Serrato. ORCID: <https://orcid.org/0000-0003-2679-8375>

Diseño y desarrollo de los experimentos. Redacción del manuscrito.

David Santiago Mahecha Hernández: <https://orcid.org/0000-0002-2801-6333>

Diseño y desarrollo de los experimentos. Redacción del manuscrito.

Cesar Yesid Barahona Rodríguez. ORCID: <https://orcid.org/0000-0001-7673-7381>

Diseño y desarrollo de los experimentos. Redacción del manuscrito.