

**Técnicas de Machine Learning para la predicción de desempeño académico en el
Desarrollo del espacio proyectivo del Pensamiento Espacial**

ÓSCAR ALEXANDER MÉNDEZ AGUIRRE

JAHIR GUILLERMO LÓPEZ MARTÍNEZ

**Proyecto para optar al Título de Maestría en Tecnologías de la Información Aplicadas a la
Educación**

UNIVERSIDAD PEDAGÓGICA NACIONAL

FACULTAD DE CIENCIA Y TECNOLOGÍA

MAESTRÍA EN TECNOLOGÍAS DE LA INFORMACIÓN APLICADAS A LA

EDUCACIÓN

BOGOTÁ, D.C.

2019

**Técnicas de Machine Learning para la predicción de desempeño académico en el
Desarrollo del espacio proyectivo del Pensamiento Espacial**

**ÓSCAR ALEXANDER MÉNDEZ AGUIRRE
JAHIR GUILLERMO LÓPEZ MARTÍNEZ**

**DIRECTOR
VICTOR QUINTERO SUAREZ**


**UNIVERSIDAD PEDAGÓGICA NACIONAL
FACULTAD DE CIENCIA Y TECNOLOGÍA
MAESTRÍA EN TECNOLOGÍAS DE LA INFORMACIÓN APLICADAS A LA
EDUCACIÓN
BOGOTA, D.C.
2019**

Derechos de autor

—Para todos los efectos, declaro que el presente trabajo es original y de mi total autoría; en aquellos casos en los cuales he requerido del trabajo de otros autores o investigadores, he dado los respectivos créditosl. (Artículo 42, párrafo 2, del Acuerdo 031 del 4 de diciembre de 2007 del Consejo Superior de la Universidad Pedagógica Nacional)



Este trabajo de grado se encuentra bajo una Licencia Creative Commons de **Reconocimiento – No comercial – Compartir igual**, por lo que puede ser distribuido, copiado y exhibido por terceros si se muestra en los créditos. No se puede obtener ningún beneficio comercial y las obras derivadas tienen que estar bajo los mismos términos de licencia que el trabajo original.

	FORMATO	
	RESUMEN ANALÍTICO EN EDUCACIÓN - RAE	
Código: FOR020GIB	Versión: 01	
Fecha de Aprobación: 10-10-2012	Página v de 122	

1. Información General	
Tipo de documento	Tesis de grado de maestría de investigación
Acceso al documento	Universidad Pedagógica Nacional. Biblioteca Central
Título del documento	Técnicas de Machine Learning para la predicción de desempeño académico en el Desarrollo del espacio proyectivo del Pensamiento Espacial
Autor(es)	Méndez Aguirre, Óscar Alexander; López Martínez, Jahir Guillermo
Director	Quintero Suárez, Víctor
Publicación	Bogotá. Universidad Pedagógica Nacional, 2019. 129 p.
Unidad Patrocinante	Universidad Pedagógica Nacional.
Palabras Claves	PENSAMIENTO ESPACIAL, ESPACIO PROYECTIVO, MACHINE LEARNING, PREDICCIÓN, DESEMPEÑO ACADEMICO.

2. Descripción

El presente trabajo de investigación busca aplicar técnicas de machine learning para predecir el desempeño académico de los estudiantes en el espacio proyectivo del pensamiento espacial e identificar los grupos de estudiantes que puedan presentar un desempeño bajo en el desarrollo de esta proyectiva con el fin de tomar acciones tempranas para mejorar su desempeño en las pruebas realizadas en la institución al finalizar su periodo académico.

Antes de aplicar la técnica de Machine Learning, los estudiantes resolvieron dos pruebas: una prueba simulacro donde los estudiantes objeto de estudio respondieron una prueba simulacro tomada del modelo de pruebas que aplicó la Universidad Nacional de Colombia en el proceso de admisión para el segundo semestre del año 2010. La prueba estaba conformada por 15 preguntas y se aplicó a los 153 estudiantes de la Institución Educativa Departamental General Carlos Albán del municipio de Albán – Cundinamarca.

La otra prueba que resolvieron los estudiantes de grado octavo, noveno, décimo y once del año 2019 fue una encuesta cuyo propósito fue el de conocer las condiciones sociodemográficas de los estudiantes. La prueba estuvo conformada por 33 grupos de preguntas relacionadas con datos personales, estudios realizados; datos de estudio, datos laborales, salario de los padres; datos de vivienda y comodidades de vivienda.

En la investigación se usó como enfoque metodológico el modelo CRISP-DM que significa Cross Industry Standard Process for Data Mining. El modelo CRISP-DM actúa como marco de trabajo propio del desarrollo de proyectos que abordan investigaciones basadas en minería de datos y técnicas de machine learning. (Fuentes, 2018).

La investigación se desarrolló en tres etapas: 1-) adaptación de la Prueba de Admisión de la Universidad Nacional de Colombia, 2-) diseño de encuesta simulacro conformada por variables sociodemográficas y 3-) diseño del modelo de: procesamiento, sanidad de datos y desarrollo del modelo predictivo con técnicas de Machine Learning.

Se concluye que la aplicación de técnicas de Machine Learning y simulación para el desarrollo de un modelo que permita la predicción del desempeño de los estudiantes de Educación Básica y Media se constituye como una herramienta eficaz para el docente, ya que pueden clasificar a los estudiantes y conocer con alto grado de precisión las categorías de DESEMPEÑO_BAJO y DESEMPEÑO_ALTO de los aprendices; esta es una ventaja que permite a los docentes crear estrategias en cada una de sus asignaturas para orientar al desarrollo del pensamiento espacial a partir de la Teoría del Desarrollo del Conocimiento Espacial de Piaget.

Por último, las actividades que se desarrollan en el aula se pueden transformar con la aplicación sistemática de una metodología como la CRISP donde el Machine Learning se constituye en una aplicación que da respuesta a algunas preguntas de investigación, hace productivos este tipo de modelos y convierte las instituciones educativas en centros inteligentes de formación.

3. Fuentes

Amadiou, F., Tricot, A., Gog, Paas, & Marine. (2009). Apprendre avec le numérique. Mythes et réalités. *RETZ*.

Baker, D. (2015). Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Wiley periodicals inc.*

Bazzan, A., Heinen, M. R., & Constantine, E. (2015). An Agent-Based Simulator for Intelligent Transportation Systems. *Research gate*.

Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2016). Early Detection of Students at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods. *CESifo Working Papers*.

Boulesteix, A., Bender, A., Lorenzo, J., & Strobl, C. (05 de 2013). Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations. *BMC bioinformatics*, 13.

Bravo, J., Ortega, M., & Prieto, M. (Enero - Abril de 2009). Entornos de Simulación en la Educación a Distancia. (E. U. Calatrava, Ed.) *Revista de Enseñanza Tecnológica*, 9.

Castellanos, J. L. (2017). Using machine learning based on eye gaze to predict targets: An exploratory study. *Symposium Series on Computational Intelligence (SSCI)*.

Castro, J. (2004). *El desarrollo de la noción del espacio en el niño de Educación inicial*. (U. d. Andes, Ed.) Táchira, Venezuela.

Cataldi, Z., Dominighini, C., & Lage, F. J. (2013). Fundamentos para el Uso de Simulaciones en la Enseñanza. *Revista de Informática Educativa y Medios Audiovisuales*, 10, 12.

Chang, Y. (2014). 3D-CAD effects on creative design performance of different spatial.

CleverData. (2014). <https://cleverdata.io/que-es-machine-learning-big-data/>.

Cohen, C., & Hegarty, M. (2014). Visualizing cross sections: Training spatial thinking using interactive animations and virtual objects. *Learning and Individual Differences*.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *Springer*.

Conati, C., & Mavrikis, M. (2018). AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling. *Cornell University*.

Consejo Provincial de Educación. (1999). *Acerca de la Enseñanza del Espacio*. Provincia Río Negro, Argentina.

Dicovski, L. M., & Pedroza, M. E. (2018). Mining Data, an Innovation of Quantitative Research Methods, in the Measurement of University Academic Performance. *Revista Científica de FAREM-Esteli*, 10.

Duque, C. (2009). *Geometría intuitiva desde el cuarto de baño*. Obtenido de www.sinewton.org: http://www.sinewton.org/numeros/numeros/70/Experaula_01.pdf

Durán, E., Costaguta, R., Maldonado, M., Únzaga, S., Chequer, G., Menini, M., . . . Fernandez, N. (2010). Técnicas de aprendizaje de máquina y personalización en educación. *Workshop de Investigadores en Ciencias de la Computación*.

Fuentes, A. (2018). *Hands-On Predictive Analytics with Python*. Packt Publishing.

Galvez, G. (1994). La geometría, la psicogénesis de las nociones espaciales y la enseñanza de la geometría en la escuela elemental. En G. Galvez, *Didáctica de las matemáticas. Aportes y reflexiones*. Buenos Aires: Paidós. Educador.

Garbanzo Vargas, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista Educación*.

García, J. (05 de 2016). Líneas de investigación en minería de datos en aplicaciones en ciencia e ingeniería: Estado del arte y perspectivas. *Arxiv, Artificial Intelligence (cs.AI)*.

Gardner, H. (2011). *Inteligencias Múltiples, la Teoría en la Práctica*. España: Editorial Paidós.

Gardner, H. (2011). *Inteligencias Múltiples. La Teoría en la Práctica*. España: Ediciones Paidós.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc.

Giraldo, M. (2017). *Descubrimiento de patrones*. Bogotá: Universidad Nacional de Colombia.

González, J. (2017). *Pensamiento Espacial: Una experiencia de aula apoyada por realidad aumentada y Learning Catalytics, para el desarrollo de habilidades de razonamiento geométrico*. Bogotá: Universidad de los Andes.

Gutierrez, D. D. (2015). *Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R*. Technics Publications.

Hale, J. (12 de 10 de 2018). *Towards Data Science* . (Towards Data Science) Recuperado el 12 de 04 de 2019, de Towards Data Science : <https://towardsdatascience.com/smarter-ways-to-encode-categorical-data-for-machine-learning-part-1-of-3-6dca2f71b159>

Hart, K., & Moore, G. (1971). *Desarrollo de la cognición espacial*. (P. P. Research, Ed.) Worcester, 7.

Hederich, C. (2017). Estilística educativa un campo de investigación en educación y pedagogía. *Revista latinoamericana de estudios educativos*.

Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*.

ICFES. (2016). *Marco de Factores Asociados. Saber 31, 5º y 9º*. Bogotá.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to statistical learning with Applications in R* (Vol. 8). (Springer, Ed.) New York , Heidelberg Dordrecht London: Springer Texts in Statistics.

Jensen, K. (2016). <file:///Users/oscardenendezaguirre/Downloads/ModelerCRISPDM.pdf>.
Obtenido de <file:///Users/oscardenendezaguirre/Downloads/ModelerCRISPDM.pdf>:
<file:///Users/oscardenendezaguirre/Downloads/ModelerCRISPDM.pdf>

Kakade, S., & Ozdemir, S. (2018). *Principles of Data Science - Second Edition*. Packt Publishing.

Lawson, J. (2015). *Data Science in Higher Education: A Step-by-Step Introduction to Machine Learning for Institutional Researchers*. CreateSpace Independent Publishing Platform.

Martín, N. (2008). Development of a fast remedial course to improve the spatial abilities of engineering students. *Journal of Engineering Education*, 505-513.

Martin, O. (2018). *Bayesian Analysis with Python - Second Edition*. Packt Publishing.

Matemático, O. c. (2007). Orientaciones curriculares para el campo de Pensamiento Matemático. En S. d. Distrital, *Orientaciones curriculares para el campo*. Bogota. Distrito capital: Secretaria de educación.

Mavrikis. (2010). *La Inteligencia Artificial en la Educación Necesita Aprendizaje Automático Interpretable*.

Mc Cormick, K., & Salcedo, J. (2017). *IBM SPSS Modeler Essentials*. USA: Packt Publishing.

MEN. (1998). *Lineamientos curriculares para el área de matemáticas*. Bogotá. Distrito capital.: Creamos alternativas.

MEN. (2006). *Estandares básicos de competencias en matemáticas*. Bogotá: Ministerio de educación nacional.

Mitchell, T., Chen, S., & Macredie, R. (2004). Adapting Hypermedia to Cognitive Styles: Is it necessary? *Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK*.

Moncayo, C. M. (2016). Determinantes que influyen en el Rendimiento Académico: Un Estudio Aplicado para Colombia a Partir de las Pruebas Saber 11 - 2014. Bogotá, Colombia.

Morales, C. A. (2012). El desarrollo del pensamiento espacial y la competencia matemática. Una aproximación desde el estudio de los cuadriláteros. *Revista Amazonia Investiga / Florencia, Colombia*.

Navarro, R. (2009). La capacidad espacial y su relación con la ingeniería. *DYNA*.

Olaya, C., & Diaz, C. C. (2017). *Social Systems Engineering: The Design of Complexity*. Wiley.

Ott, M., & Pozzi, F. (2012). Digital games as creativity enablers for children. *Behaviour & Information Technology*.

Pérez, D., & González, C. (2007). *Minería de Datos. Técnicas y Herramientas*. B & N.

Phillips, L., Norris, S., & Macnab, J. (2010). Visualization in mathematics, reading and science education. *Springer*.

Piaget, J. (1947). *La representación del Espacio en el Niño*. Paris: PUF.

Piaget, J. (1948). *La Geometría Espontánea en el Niño*. Paris: PUF.

Piaget, J., & Inhelder, B. (1997). *Psicología del niño*. Madrid: Morata.

Prabha, L., & Shanavas, M. (07 de 2014). EDUCATIONAL DATA MINING APPLICATIONS. *Operations Research and Applications: An International Journal (ORAJ)*, 1.

Rainey, R., & Talk. (2012). FTC Final Privacy Report Draws a Map to Meaningful Privacy Protection in the Online World. *ELECTRONIC FRONTIER FOUND*.

Ramírez. (19 de 06 de 2018). *medium.com*. (medium.com) Recuperado el 12 de 04 de 2019, de medium.com: <https://medium.com/bluekiri/curvas-pr-y-roc-1489fbd9a527>

Ramírez, F., Hernández, J., & Ramírez, M. J. (2004). *Introducción a la minería de datos*. Pearson Educación.

Rice, M. &. (10 de 2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, 29.

- Salas. (2005). Hallazgos de la investigación sobre la inserción de las Tecnología de la información y la comunicación (TIC) en la enseñanza: la experiencia de los últimos diez años en los Estados Unidos. *Revista Educación.*, 29(2), 53-66.
- Salas, R. P., & Zuleta, P. A. (1995). La Simulación como Método de Enseñanza y Aprendizaje. Educación Médica Superior. *Educ Med Super v.9 n.1 Ciudad de la Habana*, 72.
- Samsudin, K., Rafi, A., & Hanif, A. (2011). *Training in Mental Rotation and Spatial*.
- Sharma, T., Bali, R., & Sarkar, D. (2017). *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*. Apress.
- Sison, R., & Shimura, M. (1998). Student Modeling and Machine Learning. *International Journal of Artificial Intelligence in Education (IJAIED)*.
- Sposito, O., Etcheverry, M., Ryckeboer, H., & Bossero, J. (2008). *Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil*. San Justo – Provincia de Buenos Aires – Argentina: Universidad Nacional de La Matanza.
- Tamoghna, Raghav, B., & Dipanjan, S. (2018). *Hands-On Transfer Learning with Python*. Packt Publishing.
- Thanaki, J. (2018). *Machine Learning Solutions*. Packt Publishing.
- Timarán, R., & Jiménez, J. (2014). Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM. *Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación*.
- UNMC Nebraska Medicine. (7 de 3 de 2001). <https://www.unmc.edu/>. Obtenido de [www.unmc.edu: http://gim.unmc.edu/dxtests/roc3.htm](http://gim.unmc.edu/dxtests/roc3.htm)
- Van der laan, M. J. (2011). Targeted learning. *Springer*.
- Vasco, C. E. (2006). *Sistemas geométricos. Un nuevo enfoque para la didáctica de las matemáticas*. Bogota. Distrito capital.: Ministerio de Educacion Nacional.
- Vázquez, S., & Noriega, M. (2011). Razonamiento espacial y rendimiento académico. *Interdisciplinaria. Revista de psicología y ciencias afines*.
- Vera, Y. (2003). *El uso de la Internet como herramienta educativa y su relación en el rendimiento cualitativo de los alumnos del sexto grado de Educación Básica*. Maracaibo, Venezuela.: Trabajo de Grado, Universidad Rafael Beloso Chacín.
- Wirth, R. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.

4. Contenidos

El documento se estructura en cuatro partes que corresponden al paso a paso en el desarrollo de la investigación.

Una primera parte presenta el planteamiento del problema donde se encuentra la estructura general del documento y traza el camino que sigue la presente investigación donde se incluye pregunta de investigación y objetivos de investigación.

En la segunda parte se encuentran los antecedentes que direccionan las necesidades de investigar, dónde se encuentran falencias y dónde es posible realizar una intervención más asertiva. También se encuentra el marco teórico que incluye las teorías que fundamentan la investigación y que se centra en las teorías del pensamiento espacial y las fases del proceso del machine learning.

En la tercera parte se plantea la metodología que incluye población, muestra, variables planteadas como lo son el desempeño académico en el desarrollo del espacio proyectivo del pensamiento espacial, y la aplicación de técnicas de machine learning con múltiples valores, las etapas de la investigación y el modelo del machine learning.

En la cuarta parte se encuentran los análisis de resultados de las variables analizadas, para finalmente realizar las conclusiones, mediante las cuales se da respuesta a la pregunta de investigación, presentando unas recomendaciones.

5. Metodología

El marco metodológico y procedimental empleado en el desarrollo de la investigación, está integrado por tipo de investigación, población y muestra, técnicas empleadas y etapas e instrumentos utilizados y descripción del ambiente computacional.

En la investigación se usó como enfoque metodológico el modelo CRISP-DM que significa Cross Industry Standard Process for Data Mining. El modelo CRISP-DM actúa como marco de trabajo propio del desarrollo de proyectos que abordan investigaciones basadas en minería de datos y técnicas de machine learning. (Fuentes, 2018).

Los reportes que resultan del trabajo desarrollado con esta metodología se enmarcan en los procesos conceptuales, algorítmicos y estadísticos descritos por (James, Witten, Hastie, & Tibshirani, 2013) en su libro *An Introduction to Statistical Learning with Applications in R*. y en

aplicaciones desarrolladas en contextos educativos cuyos resultados señalan que el uso de técnicas de Machine Learning favorece la optimización de procesos propios del campo educativo (Lawson, 2015) (Castellanos, 2017).

Antes de aplicar la técnica de Machine Learning, los estudiantes resolvieron dos pruebas: una prueba simulacro y una encuesta de factores sociodemográficos.

6. Conclusiones

Una vez realizado el estudio y aplicado el algoritmo para dar respuesta a la pregunta planteada se puede evidenciar que en los procesos de aprendizaje existen factores que inciden directamente en el proceso de aprendizaje del estudiante y que no le permiten un desarrollo adecuado de dicho proceso.

La aplicación de técnicas de Machine Learning y simulación para el desarrollo de un modelo que permita la predicción del desempeño de los estudiantes de Educación Básica y Media se constituyen en una herramienta eficaz para el docente, ya que pueden clasificar a los estudiantes y conocer con alto grado de precisión las categorías de DESEMPEÑO_BAJO y DESEMPEÑO_ALTO de los aprendices; esta es una ventaja que permite a los docentes crear estrategias en cada una de sus asignaturas para orientar al desarrollo del pensamiento espacial a partir de la Teoría del Desarrollo del Conocimiento Espacial de Piaget.

Actualmente existen muchas aplicaciones de minería de datos en educación y trabajos realizados en diferentes partes del mundo que señalan que, la intervención de dichas tecnologías en el manejo de los múltiples datos de los estudiantes brinda soluciones que favorecen los procesos de enseñanza – aprendizaje y previenen situaciones que conllevan a la deserción del aprendiz.

Los algoritmos, el Machine learning, las simulaciones, la extracción, procesamiento y almacenamiento de datos se configura en una metodología necesaria e imprescindible en el desarrollo de estrategias y conocimientos que promuevan y potencien las actividades en el aula de clase y todos los procesos pedagógicos que esto conlleva.

El desarrollo tecnológico no busca reemplazar los docentes, por el contrario, se suma a la ardua tarea de concebir nuevas formas, procesos, procedimientos, métodos, metodologías y

estrategias que eleven el nivel de eficacia en el desarrollo de actividades de aprendizaje. La apropiación de la información se convierten en un insumo de mucho valor que sugiere sumar esfuerzos para su análisis y exploración afín de aplicarlo en ámbitos en diferentes ámbitos educativos.

Elaborado por:	Méndez Aguirre, Óscar Alexander; López Martínez, Jahir Guillermo		
Revisado por:	Quintero Suárez, Víctor		
Fecha de elaboración del Resumen:	15	11	2019

Índice de Contenido

1. Planteamiento del problema.....	20
2. Objetivos.....	27
2.1. Objetivo general.....	27
2.2. Objetivos específicos.....	27
3. Antecedentes.....	28
3.1. Investigaciones sobre el espacio proyectivo del pensamiento espacial.	28
3.2. Investigaciones sobre el desarrollo del pensamiento espacial por medio de las TIC. ...	34
3.3. Investigaciones sobre el uso de técnicas de Machine Learning en la educación.	37
4. MARCO TEÓRICO.....	45
4.1. Relaciones espaciales según Piaget.....	48
4.1.1. Relaciones espaciales:	48
4.2. Características del Pensamiento espacial según el MEN.....	50
4.2.1. Características cognitivas.....	50
4.2.2. Características individuales.....	51
4.2.3. Características del entorno físico, cultural, social e histórico	51
4.2.4. Estrategias pedagógicas.....	52
4.3. Teorías sobre el pensamiento espacial	57
4.3.1. Teoría de Piaget y el espacio Proyectivo	59
4.3.2. Teorías de Howard Gardner y Grecia Gálvez	64
4.4. Minería de Datos y Machine Learning.....	66
4.4.1. Fases del Proceso del Machine Learning.....	69
4.4.2. Ámbitos de Aplicación del Machine Learning.	75
4.4.3. Simulación educación	76
4.5. CRISP – DM	78
5. Metodología	82
5.1. Prueba Simulacro	83
5.2. Encuesta sobre Factores Sociodemográficos	83
5.2.1. Población.....	83

5.2.2.	Muestra.....	84
5.2.3.	Variable Dependiente.....	84
5.2.4.	Variable Independiente.....	84
5.3.	Etapas de la investigación.....	85
5.3.1.	Primera etapa.....	85
5.3.2.	Segunda etapa.....	85
5.3.3.	Tercera etapa.....	86
5.4.	Diseño de ambiente computacional.....	88
5.5.	Modelo de Machine Learning.....	89
5.5.1.	Simulador Web.....	92
6.	Análisis de Resultados.....	96
6.1.	Prueba simulacro grado octavo.....	96
6.2.	Prueba simulacro grado noveno.....	97
6.3.	Prueba simulacro grado décimo.....	98
6.4.	Prueba simulacro grado once.....	99
6.5.	Desarrollo del modelo.....	101
6.5.1.	Comprensión del contexto.....	101
6.5.2.	Información de la base de Datos.....	103
6.5.3.	Interpretación de los datos.....	104
7.	Discusión y Análisis de Resultados del Uso del Clasificador Final con un Grupo Real de Estudiantes.....	120
8.	Conclusiones y Recomendaciones.....	124
8.1.	Recomendaciones.....	127
	Referencias Bibliográficas.....	130
	Anexos.....	137

Índice de Tablas

Tabla 1. Etapas del Proceso de Conocimiento del Espacio según Jean Piaget.....	49
Tabla 2. El espacio proyectivo según Jean Piaget	59
Tabla 3. Tipos de Algoritmos según Categorías	72
Tabla 4. Herramientas y Librerías del Algoritmo	73
Tabla 5. Fases de la Metodología CRISP – DM.....	79
Tabla 6. Diseño de la investigación	82

Índice de Figuras

Figura 1. Teorías sobre Pensamiento Espacial	57
Figura 2. Diagrama de Ven Data Sciencie, (Kakade & Ozdemir, 2018).....	67
Figura 3. CRISP -DM WorkFlow (Jensen Own, 2016).....	78
Figura 4. Diagrama de Componentes del Simulador Web	90
Figura 5. Esquema Conceptual de la Solución.	91
Figura 6. Formulario de Ingreso de Datos	92
Figura 7. Resumen de Estudiantes Clasificados.	92
Figura 8. Detalle del estudiante.	93
Figura 9. Listado de Estudiantes.....	94
Figura 10. Resultados del Simulacro Grado Octavo - 2019.	97
Figura 11. Resultados Simulacro Grado Noveno - 2019.....	98
Figura 12. Resultados Simulacro Grado Decimo - 2019	99
Figura 13. Resultados Simulacro Grado Once - 2019.....	100
Figura 14. Información Inicial Data Set.	103
Figura 15. Tipo de Variables.	104
Figura 16. Información de Datos Faltantes o Perdidos.....	105
Figura 17. Variable DANE_RESIDE_DPTO.....	105
Figura 18. Variable DEPART_NACE.....	105
Figura 19. Distribución DANE_RESIDE_DEPTO.....	106

Figura 20. Distribución EDAD.	107
Figura 21. Tipos de Datos de Dataset.	108
Figura 22. Implementación Simple de Montecarlo.....	109
Figura 23. Variable HABITACION_DUERME_HOGAR.	110
Figura 24. Importancia Acumulada de Variables.	112
Figura 25. Variables por Importancia.	113
Figura 26. ROC - AUC.	114
Figura 27. Matriz de Confusión.....	115
Figura 28. Reporte del clasificador.....	116
Figura 29. Ejecutor Simulación en Batch.	118
Figura 30. Salida de la Simulación para el grupo comparador - 2019.....	119
Figura 31. Porcentaje de coincidencia 2019.	122

Índice de Anexos

Anexo 1. Facsímil Prueba de Admisión Universidad Nacional 2010	137
Anexo 2. Diccionario de Variables MEN - ICFES	143
Anexo 3. Encuesta de factores sociodemográficos	146
Anexo 4. Salida de la Simulación para grupo comparador - 2019	147
Anexo 5. Documentación del algoritmo	149

1. Planteamiento del problema

El desempeño académico de los estudiantes está asociado con múltiples variables que lo afectan; en este sentido, el efecto de algunas de tales variables se ha estudiado desde diversos campos, pero existe un sinnúmero de variables que se desconoce o se ha estudiado muy poco y que no permiten predecir con un mínimo nivel de acierto futuros resultados de desempeño para tomar acciones que eviten la mortalidad académica derivada de los bajos desempeños. A renglón seguido, se describen algunos estudios que han abordado las variables que afectan el desempeño académico desde el dominio de las TIC's y luego se muestran otras variables desconocidas o poco estudiadas que afectan dicho desempeño y que motivan el presente trabajo de investigación.

Respecto a los estudios que han abordado variables que afectan el desempeño académico se observa que en el área específica de la geometría dinámica concretamente en lo que tiene que ver con la capacidad de representación y el razonamiento espacial, se han desarrollado trabajos desde el dominio de las TIC's, dirigidos a aportar en la solución de esta problemática, es así como, la incursión de las tecnologías de la información y la comunicación TIC al campo educativo ha generado grandes posibilidades debido a las múltiples ventajas que supuestamente éstas ofrecen, tanto en la resolución de problemas de cualquier dominio del conocimiento como en mejorar los procesos de aprendizaje.

En este sentido, algunos investigadores sostienen que los ambientes de aprendizaje computacionales permiten puntualizar en factores críticos que facilitan el aprendizaje entre otras razones porque: dan margen a que los estudiantes organicen su proceso de aprendizaje de forma flexible, debido a que ofrecen disponibilidad inmediata de contenidos de aprendizaje durante las 24 horas del día, los 7 días de la semana y accesibilidad a gran cantidad de información desde cualquier lugar del mundo, lo mismo que, permiten verificar y contrastar en tiempo real ideas y

modelos numéricos con el resultado de operaciones matemáticas obtenidas previamente mediante el apoyo de lápiz y papel o a través del uso del cálculo mental (Hederich, 2017).

Los escenarios computacionales también permiten la representación y la manipulación de objetos geométricos, con lo cual se constituyen en una herramienta privilegiada para explorar, representar y refinar el pensamiento espacial; en fin, el uso de escenarios computacionales de aprendizaje hacen más accesible e importante para los estudiantes temas de la geometría, probabilidad, estadística y álgebra, evidenciando entonces que, las tecnologías de información y comunicación aligeran y superan la capacidad de cálculo de la mente humana.

Al respecto, algunos investigadores sostienen que el pensamiento espacial es maleable y responde positivamente a la mejora de habilidades de visualización, manipulación y representación de objetos a partir del entrenamiento basado en tecnologías de información (Ministerio de Educación Nacional de Colombia, 2006; Martín, 2008; Navarro, 2009; Samsudin, Rafi, & Hanif, 2011; Chang, 2014; Cohen & Hegarty, 2014). Los autores concluyen que el uso de herramientas digitales para el entrenamiento de las habilidades espaciales es significativo (Ott & Pozzi, 2012; Villa, 2016). En la actualidad la visualización del aprendizaje de la geometría y las matemáticas no sólo es contemplada como una propuesta ilustrativa, sino que está siendo reconocida como una componente clave del razonamiento, la resolución de problemas y la demostración (Phillips, Norris, & Macnab, 2010).

No obstante, las múltiples ventajas que en teoría ofrecen los ambientes computacionales también se encuentran desventajas; algunos investigadores encontraron evidencia científica que muestra que no todos los estudiantes se benefician de estos entornos, ni alcanzan los logros mínimos esperados cuando navegan en estos ambientes; mientras algunos aprendices navegan y

aprenden con facilidad, otros navegan sin control y sin un propósito definido, en consecuencia, obtienen muy bajos desempeños (Amadiou, Tricot, Gog, Paas, y Marine, 2009; Mitchell, Chen, y Macredie, 2004) . Esta problemática se presenta debido a lo poco significativo que resulta la incorporación de las TIC's en algunos niveles de primaria, secundaria y aún en la universidad por fallas en: procesos de planeación, soporte técnico, apoyo a la investigación, participación de los usuarios en el diseño, falta de estructuración de los ambientes de aprendizaje computacionales entre otros como lo expresa (Vera, 2003;Salas, 2005)

A pesar de los esfuerzos que se han realizado los investigadores en el dominio de las TIC's, aún se desconocen muchas variables que causan bajos desempeños de aprendizaje; en esta línea, algunos investigadores sostienen, que no existen evidencias suficientes que demuestren una relación lineal entre aprendizaje y TIC, en este orden de ideas es posible encontrar estudiantes que no han tenido exposición directa a las TIC con mejores niveles de aprendizaje que aquellos que sí han tenido oportunidad de participar en procesos educativos mediados por TIC, sin que este hecho sea causa-efecto. Esta problemática puede explicar en parte, los bajos logros de aprendizaje reportados por los estudiantes cuando aprenden en escenarios computacionales.

De lo anteriormente expuesto se deduce que pueden existir muchas variables que afectan el proceso de aprendizaje, gran parte de las cuales no se conocen o no se tienen en cuenta al momento de diseñar entornos computacionales de aprendizaje. Los resultados también indican que aún, hay vacíos e interrogantes que es necesario resolver, por esta razón, la investigación debe avanzar en formular estudios que aporten conocimiento en torno a conocer factores que inciden en que los desempeños de los estudiantes no resultan significativos tales como: académicas, socio-culturales, recursos económicos, estrategias de implementación de programas, la capacitación de

docentes, falta de motivación hacia el uso de las tecnologías de información y comunicación (TIC) entre otras; de ser posible conocer el efectos de estas variables en el desempeño académico, probablemente se pueda hacer un mejor diagnóstico de la problemática que incide en el proceso de aprendizaje, en este sentido, la incorporación del análisis de datos a los procesos de aprendizaje contribuye con la solución de esta problemática.

Es así como, la incorporación del análisis automático de los datos a los procesos educativos permite comprender gran cantidad de variables externas al estudiantes que afectan su desempeño, tales como: edad, nivel educativo de sus padres, estrato socioeconómico, entorno, integrantes de su familia, alimentación, tipo de trabajo o tarea que realizan los padres, entorno familiar, adultos responsables del estudiante, el contexto, diferencias sociales, variables demográficas; y variables que hacen parte del estudiante como: contactos que se hace a través de entornos virtuales de aprendizaje, asistencia a clases, sus calificaciones etc. Todas esas variables generan información.

Garbanzo (2007) afirma que:

El desempeño académico está determinado por la suma de diferentes y complejos factores que actúan en la persona que aprende y ha sido definido con un valor atribuido al logro del estudiante en las tareas académicas el cual se mide mediante las calificaciones obtenidas, con una valoración cuantitativa, cuyos resultados muestran las materias ganadas o perdidas, la deserción y el grado de éxito académico. (Garbanzo, 2007)

En el caso concreto de construcción del espacio, el Ministerio de Educación Nacional de Colombia (1998) sostiene que esté condicionado e influenciado tanto por las características individuales como por la influencia del entorno físico, cultural, social e histórico (MEN, 1998).

En la misma línea del análisis automático CleverData (2014) hace énfasis en si una institución educativa maneja un alto volumen de información de sus estudiantes y dicha información se desea utilizar para realizar predicciones que le permitan dinamizar sus procesos de aprendizaje, la institución puede tomar ese volumen de datos debidamente organizados, tratarlos en bloque, generar una base de datos que puede explotar y así predecir comportamientos benéficos y comportamientos perjudiciales en pro del mejoramiento de la institución; ellos añaden que estas predicciones se pueden realizar utilizando Machine Learning.

El Machine Learning es una alternativa que nos permite contemplar muchas variables distintas de las que se han trabajado con las tecnologías de la información y comunicación y estas variables son las que permiten aproximarse e identificar el desempeño académico del estudiante. (Lawson, 2015) y otros promueven el uso de la ciencia de datos y sostienen que esta puede ser una parte integral de las operaciones de las instituciones educativas para empoderar mejor los procesos de seguimiento y orientación de los estudiantes, si se trabaja hombro a hombro con facultades, áreas y educadores para recopilar, filtrar y extraer todos los datos posibles de los comportamientos actuales y pasados de los estudiantes. El autor materializa lo anterior con una serie de estudios que caracterizan la aplicación de técnicas de Machine Learning y de la ciencia de datos para mejorar y posicionar instituciones educativas de California en procesos de certificación y acreditación.

Los autores destacan la información como el activo máspreciado que puede tener cualquier institución junto al uso de técnicas de simulación para obtener respuestas a problemas que requieren de dicha información y del tratamiento de altos volúmenes de datos (Rainey & Talk, 2012; Bazzan, Heinen, & Constantine, 2015) llaman la atención sobre los altos costos que implica la aplicación de tales técnicas en instituciones del común y aclaran que son un medio necesario para lograr objetivos y optimizar el desarrollo real de los procesos de aprendizaje. En esta línea,

las acciones relacionadas con el procesamiento de datos requieren de un Sistema Inteligente que permita modelar actividades, conductas, habilidades y estados de los estudiantes, que en la mayoría de los casos no están bien definidos, entendidos o capturados.

En este sentido, Conati y Mavrikis (2018) en el artículo *“AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling”* plantean que las técnicas de Machine Learning, son instrumentos que permiten abordar estos desafíos, ya que pueden ayudar a aprender de los datos, los conocimientos y modelos que son difíciles de obtener de expertos humanos y calcular las predicciones de los estados cognitivos y mentales de los estudiantes. (Conati y Mavrikis, 2018). Sumado a la anterior (Olaya & Díaz, 2017) mencionan como los sistemas socio técnicos complejos se caracterizan por fenómenos sociales y de comportamiento que afectan significativamente el funcionamiento y el rendimiento de sistemas de ingeniería o diseños complicados y muestra ejemplos de aplicaciones del Machine Learning relacionadas con estados de estudiantes y habilidades como la autoeficacia (Mavrikis, 2010) reacciones emocionales (Conati & Maclaren, 2009), predicción de estudiantes con capacidad para realizar consultas científicas con éxito en entornos virtuales (Baker, 2015).

La interpretación de estas técnicas de Machine learning es fundamental para que la aplicación explique a los usuarios sus inferencias y acciones. A través de ellas se puede mejorar la efectividad pedagógica ya que pueden ayudar a los estudiantes a comprender por qué el sistema considera que sus respuestas son incorrectas frente a un tema en particular que debe o no aprender.

Teniendo en cuenta los planteamientos anteriores y considerando que las dificultades en el desarrollo del pensamiento espacial están asociadas a acciones tan finas y factores externos transparentes para la escuela y que se ha encomendado a los educadores la tarea de identificarlas

y mediar para superarlas a través de sus prácticas pedagógicas; el presente trabajo de investigación busca aplicar técnicas de machine learning para predecir el desempeño académico de los estudiantes en el espacio proyectivo del pensamiento espacial e identificar los grupos de estudiantes que puedan presentar un desempeño bajo en el desarrollo de esta proyectiva con el fin de tomar acciones tempranas para mejorar su desempeño en las pruebas realizadas en la institución al finalizar su periodo académico.

Esta investigación está encaminada a determinar si el aprovechamiento de la información que brinda el uso de los factores sociodemográficas asociados al desempeño académico y la aplicación de técnicas como el Machine Learning y la simulación podrían llegar a predecir el desempeño que pueda presentar el estudiante frente al desarrollo del espacio proyectivo del pensamiento espacial.

La investigación está encaminada a resolver la siguiente pregunta:

¿La aplicación de técnicas de Machine Learning en procesos educativos permiten predecir el desempeño de estudiantes de educación básica y media en el desarrollo del espacio proyectivo del pensamiento espacial?

2. Objetivos

2.1. Objetivo general

- Evaluar la incidencia del uso de machine learning en la predicción del desempeño en el espacio proyectivo del pensamiento espacial.

2.2. Objetivos específicos

- Determinar el efecto que los factores sociodemográficos provocan en la predicción del desempeño del desarrollo del espacio proyectivo del pensamiento espacial.
- Desarrollar un modelo de machine learning que permita procesar el efecto de las variables académicas y sociodemográficas en la predicción del desempeño de los estudiantes en el espacio proyectivo del pensamiento espacial.
- Validar el modelo de machine learning para la predicción del desempeño de los estudiantes en el espacio proyectivo del pensamiento espacial.
- Comparar las predicciones del modelo de machine learning con los resultados del test de pensamiento espacial aplicado al grado octavo.
- Realizar recomendaciones sobre las posibilidades del uso pedagógico del Machine Learning en la predicción de problemáticas escolares y el análisis masivo de datos.

3. Antecedentes

A continuación se presenta una revisión de investigaciones realizadas entre los años 1998 al 2017 afines con los estudios sobre el espacio proyectivo del pensamiento espacial, el desarrollo del pensamiento espacial por medio de las TIC e investigaciones sobre el uso de técnicas de Machine Learning en la educación y su influencia en la predicción del desempeño académico de los estudiantes; se espera que tales investigaciones aporten significativamente en fundamentar el uso de técnicas de machine learning para categorizar el comportamiento académico de los aprendices y predecir su desempeño.

3.1. Investigaciones sobre el espacio proyectivo del pensamiento espacial.

Castro (2004) realizó investigación de tipo documental relacionada con la noción de espacio que se constituye en uno de los marcos lógico-matemáticos fundamentales. Los resultados del estudio se utilizan para estructurar el futuro pensamiento abstracto- formal, lo mismo que para reestructurar los aspectos organizativos, curriculares y pedagógicos de la educación de los niños entre 0 y 6 años. El autor aporta algunas referencias que se constituyen en fundamentos esclarecedores de muchas de las estrategias de enseñanza y de actividades de aprendizaje que los docentes realizan o pudieran realizar con sus alumnos como, por ejemplo, las actividades cognitivo - procedimentales que favorecen el desarrollo de la noción de espacio en el niño; también realza la importancia que tienen los primeros años de vida de los niños (Castro, 2004)

El autor señala que, resulta imperioso el conocimiento del desarrollo de las nociones espaciales por parte de los docentes que atienden en sus primeros años de vida académica a niños

de nivel de pre-escolar; de dicho conocimiento depende la adecuada selección de estrategias de enseñanza y de actividades de aprendizaje que fomenten el desarrollo de las nociones de carácter topológico, proyectivo y euclidiano, así como, la comprensión de los principios fundamentales de la Geometría.

El autor concluye: el desarrollo del espacio proyectivo permite a los niños visualizar los cambios que sufren ángulos y longitudes en la representación del objeto observado; así por ejemplo, cuando dibujan un paisaje con árboles cada vez más pequeños reflejan la profundidad y el alejamiento que resultan de los cambios en las longitudes y los ángulos que contienen, mientras que las líneas, puntos y proporciones permanecen invariables; paralelamente a los conceptos proyectivos, los conceptos topológicos se transforman también en conceptos Euclidianos, lo que equivale a decir que el niño comienza a percibir los objetos de su espacio exterior no como algo estático, sino como objetos móviles; del mismo modo, puede describir y dibujar la trayectoria del recorrido de un automóvil (no sólo su punto de partida y llegada como ocurría antes), comprender la congruencia de un cuerpo al sufrir un cambio rígido (movimiento, rotación, traslado) y conservar las propiedades de longitud, ángulos, áreas y volúmenes.

En la investigación se encontró que el desarrollo del pensamiento espacial en los estudiantes facilita la exploración activa del espacio tridimensional del mundo externo, incrementa la imaginación y permite el desarrollo de representaciones, estáticas y dinámicas de los objetos en el espacio; para que este proceso se dé es importante promover la representación plana de las formas y las relaciones tridimensionales; de igual forma, los autores confirman la existencia de tres tipos de representaciones espaciales: topológicas, proyectivas y euclidianas que son complementarias entre sí por lo que, normalmente cada una de ellas necesita de las otras, en consecuencia, es necesario utilizar más de una para percibir un objeto en el espacio. El uso y la

relación de más de una representación permite una percepción completa, así, por ejemplo: en el dibujo técnico se emplean las proyecciones ortogonales y la representación del sólido para su completa comprensión.

Esta investigación aporta significativamente en la comprensión del trabajo que realizan los docentes para desarrollar nociones espaciales en la primera etapa académica de los niños; se ha encontrado que, de este aprendizaje depende el desarrollo de nociones topológicas, proyectivas y euclidianas; nociones que son fundamentales para que los niños puedan comprender con mayor facilidad los principios fundamentales de la geometría y amplíen sus conocimientos en el espacio proyectivo.

Vázquez y Noriega (2011) presentan los resultados de una investigación cuyos objetivos fueron: 1)- evaluar el nivel de pensamiento espacial en alumnos del ciclo básico común (CBC) de la facultad de arquitectura, diseño y urbanismo de la universidad de Buenos Aires (UBA), 2)- relacionar el pensamiento espacial y el desempeño académico, 3)- relacionar el pensamiento espacial y la educación media de origen, 4)- evaluar el progreso del pensamiento espacial al finalizar el CBC y 5)- determinar las posibles diferencias entre pensamiento espacial y desempeño académico según el sexo (Vázquez y Noriega, 2011)

En el estudio se tomó una muestra de 596 alumnos que hacen parte de la asignatura de dibujo y como metodología se utilizó un diseño cuasi-experimental de clases intactas con pre y post-test. En el tratamiento de los datos se realizaron varias pruebas estadísticas: un análisis de correlación (coeficiente de correlación de Pearson) con la finalidad de hacer una exploración previa de los mismos y un análisis de varianza en el que se relacionó el sexo como variable independiente y el pensamiento espacial y el desempeño académico como variables dependientes.

También se realizó un análisis de regresión múltiple para determinar en qué medida los factores de pensamiento espacial y el desempeño del primer semestre permitían predecir el desempeño final, y una prueba t de muestras relacionadas para evaluar el efecto del proceso de enseñanza - aprendizaje de los contenidos de la materia dibujo sobre el pensamiento espacial, lo mismo que para establecer la diferencia entre el desempeño del primer semestre y el desempeño final. En todos los casos se verificaron las condiciones que permitían realizar pruebas paramétricas.

Los resultados tanto del post-test como en pruebas finales de desempeño académico señalan que los varones superaron a las mujeres en capacidad de pensamiento espacial, en cambio, las mujeres registraron una menor tasa de deserción y un mayor incremento en los puntajes de pensamiento espacial que se interpretan en relación con el rol del esfuerzo.

Los autores concluyen que encontraron diferencias significativas en el nivel de pensamiento espacial entre el comienzo y el final del CBC que pueden ser consideradas como un efecto del aprendizaje. Este resultado avala los hallazgos de otras investigaciones referidas a la modificabilidad de esta capacidad como consecuencia del entrenamiento. Si se considera el total de la muestra, hay diferencias significativas a favor de los varones, tanto en el pre-test como en el post-test. Sin embargo, el incremento de los resultados en el último pos-test es mayor para las mujeres, aunque no llega a ser estadísticamente significativo. Por otra parte, tomando en cuenta la variable escuela de procedencia se verificó que en el caso de las mujeres que proceden de escuela técnica, no hay diferencias significativas en competencia espacial con respecto a los varones de la misma procedencia, lo que reafirma la interpretación hecha a propósito del rol diferencial del entrenamiento específico por sexos.

Se encontró que el pensamiento espacial es un buen predictor del desempeño de la primera parte del año en combinación con el nivel de pensamiento espacial en el momento del ingreso permite predecir el 60% del resultado final. De la investigación se destacan aspectos que aportan al modelo de predicción que se plantea en el trabajo de investigación como los factores espaciales y de desempeño que se manipularon para predecir el resultado final.

El Consejo Provincial de Educación (1999), realizó investigación cuyo objetivo se centró en analizar la importancia que tiene para todos los sujetos, pero especialmente para los niños, el conocimiento sobre el espacio cuyo aprendizaje se inicia mucho antes del inicio de la escolaridad. El estudio reporta que la interacción del infante con la madre y con el entorno es excepcionalmente útil en el conocimiento y representación del espacio de bebés niños de preescolar y básica primaria; debido a que, la representación espacial es una acción interiorizada que resulta de tal interacción. El niño una vez aprende la actividad sensorio motor y la liga con la percepción del objeto evoca la percepción con la imaginación y determina las primeras relaciones que él puede reconocer y representar gráficamente como las de vecindad, separación, orden, entorno y continuidad. (Consejo Provincial de Educación, 1999)

Según el consejo Provincial, las propiedades topológicas del espacio son las primeras nociones geométricas que desarrolla el sujeto. El niño comienza por distinguir figuras cerradas de abiertas, diferenciar el espacio interior del exterior a un límite definido y determinar posiciones relativas al interior de un orden lineal; desarrollar estos procesos le permite al sujeto la constitución de una geometría del objeto en singular; en otros términos, cuando el niño es capaz de predecir el aspecto de un objeto a partir de su observación desde diversos ángulos ha alcanzado el nivel de desarrollo de relaciones proyectivas.

A su vez, el desarrollo del espacio proyectivo permite al sujeto la construcción de una geometría del espacio exterior que comienza a generarse en el momento en que los ejes adelante-atrás, derecha -izquierda dejan de ser absolutos; es a partir del desarrollo de relaciones proyectivas como el sujeto se descentra y coordina distintos puntos de vista posibles para construir una representación del espacio con el que está interactuando; en últimas, la representación del espacio exterior le permite alcanzar el estadio de las relaciones euclidianas.

Como conclusión, el Consejo Provincial pone en duda la jerarquía paralela de las geometrías y los desempeños espaciales y señala que mientras muchas representaciones de zonas u objetos familiares están organizadas según reglas proyectivas o métricas, otros espacios confusos o imprecisos sólo son aprendidos a nivel topológico como la utilización del espacio y la creación de conocimientos prácticos que les permiten dominar sus desplazamientos.

También señalan evidencias de que la aparición de modos elaborados de aprehensión del espacio no suprime en absoluto los modos más simples de ese conocimiento, lo que les permite afirmar que, los estadios de desarrollo no determinan los estados de esos conocimientos sino más bien las eventuales potencialidades; cuestionan la relatividad de los factores de desarrollo en relación al pensamiento espacial y muestran que lo mínimo de desarrollo del pensamiento en forma espontánea que la mayoría de los sujetos adquiere es particularmente bajo, lo que, generaría duda sobre la suficiencia de estas adquisiciones para asegurar a la mayoría de los individuos puedan adaptarse autónomamente al medio en que viven.

El Consejo Provincial sugiere que, el trabajo de reconstrucción a nivel representativo sobre los conocimientos y esquemas de acción y perceptivos se vuelve completamente necesario para asegurar un aprendizaje eficaz de los saberes espaciales y espacio-geométricos; además es

necesario que el alumno se ubique en situaciones que tomen en cuenta las condiciones con las que se constituyen los conocimientos espaciales. (Consejo Provincial de Educación, 1999).

Esta investigación hace énfasis en la importancia que tiene para el ser humano el conocimiento del espacio y de su entorno para construir un espacio geométrico, con base en, el desarrollo de relaciones proyectivas que fundamentan el posterior desarrollo de relaciones euclidianas.

3.2. Investigaciones sobre el desarrollo del pensamiento espacial por medio de las TIC.

González (2017) investigó cómo el componente geométrico se constituye en el lenguaje a través del cual se comprende y caracteriza la realidad; la importancia de este componente en las Matemáticas se ha destacado por los beneficios cognitivos que conlleva su estudio. El trabajo se enmarca bajo el modelo de razonamiento geométrico de Van Hiele y pretende articular con las fases de desarrollo (Información, Orientación Dirigida, Explicitación, orientación Libre e Integración) permitirán desarrollar habilidades de razonamiento geométrico que el estudiante podrá adquirir alrededor del objeto matemático poliedros regulares, por medio de la técnica didáctica de la teoría de situaciones didácticas de Brousseau y un diseño instruccional ADDIE (González, 2017).

Con el desarrollo del estudio, los autores pretenden reconocer atributos y propiedades de los poliedros regulares con estudiantes de grado 6 del Gimnasio los Andes; el trabajo se apoya con tecnología bajo el marco de la geometría intuitiva, la cual entienden los autores como “aquella que relaciona la geometría tridimensional con la bidimensional a partir de la manipulación y la visualización” (Duque, 2009) . El desarrollo del trabajo se enfocó en la

visualización de poliedros regulares y sólidos platónicos a través del uso de la realidad aumentada.

El autor concluye que la propuesta de unidad didáctica desarrollada con el uso de realidad aumentada como estrategia permite estudiar y observar cada cuerpo geométrico desde todas sus perspectivas, así como, desplegar, descubrir y descomponer sus caras en figuras planas; la realidad aumentada también permite manipular los poliedros de forma natural con lo cual se mejora el desarrollo del pensamiento espacial. La propuesta combina los contenidos conceptuales con factores procedimentales (aplicación de fórmulas y cálculo) mediante distintos ejercicios: adivinar el poliedro a partir de unas pistas, comparar características de distintos poliedros, ejercicios de verdadero-falso o cálculo de áreas y volúmenes.

La investigación presenta como estrategia pedagógica novedosa y aporte interesante, el uso de la realidad aumentada como recurso didáctico que hace posible el establecimiento de relaciones entre la geometría tridimensional y la bidimensional a través de la manipulación de elementos de fácil visualización; con esta estrategia los estudiantes reconocen con mayor facilidad las propiedades de los poliedros regulares.

Morales (2012) en su investigación “El desarrollo del pensamiento espacial y la competencia matemática. Una aproximación desde el estudio de los cuadriláteros” contribuyen en el desarrollo del pensamiento espacial y los niveles de la competencia matemática formulando y resolviendo problemas, mediante el estudio del objeto matemático cuadriláteros con mediación de un programa de geometría dinámica, en estudiantes de grado séptimo de la Institución Educativa José Eustasio Rivera del municipio de Pitalito-Huila (Morales, 2012).

Para lograr el desarrollo del pensamiento espacial, se consideraron dos momentos o fases dentro del proyecto: la primera fase denominada diagnóstico permitió determinar el alcance de la

política nacional en relación con la propuesta institucional en torno al pensamiento espacial y las competencias matemáticas. Así mismo, se logró analizar cómo se estaba llevando a cabo el proceso de enseñanza y aprendizaje de la geometría y determinar cuál era el nivel de razonamiento geométrico de los estudiantes en el estudio de los cuadriláteros.

La fase dos consistió en la propuesta didáctica en torno a los cuadriláteros, donde se elaboraron actividades en las que se tuvo en cuenta el contexto social y económico de la región que teóricamente se consideran como una alternativa que contribuye al desarrollo del pensamiento espacial y la superación de los fenómenos didácticos encontrados en los estudiantes que participaron en esta investigación.

Los autores encontraron que existen diferencias entre el currículo propuesto tanto a nivel nacional como institucional y el currículo desarrollado en las clases de geometría, lo anterior se evidencia a partir del análisis de las concepciones expuestas por los profesores sobre elementos que se explicitan como directrices generales de la asignatura a orientar y de las cuales se tiene una aproximación inicial – empírica.

Los autores concluyen que existen diferencias entre el currículo propuesto a nivel nacional e institucional y el currículo desarrollado en las clases de geometría de la Institución. De acuerdo con el diagnóstico, el modelo predominante en las clases de geometría es el heteroestructurante centrado en la transmisión y recepción de contenidos y en las que prevalece el dominio casi exclusivo del sistema representación gráfico y en lenguaje natural por parte del profesor.

Con el diseño de la propuesta didáctica, la forma de entender la clase y los roles de profesores y estudiantes se propone una salida a dicha situación, a partir de las actividades propuestas, se elimina la jerarquía en el uso exclusivo de los sistemas mencionados

anteriormente y se incluye además el sistema de representación en lenguaje formal en actividades de conversión, tratamiento y comunicación.

Los autores recomiendan que los profesores de matemáticas que orientan geometría deben reconocer el tipo de dificultades a las que se puedan enfrentar sus estudiantes a la hora de realizar el estudio de cada uno de los objetos matemáticos, y de esta manera reconocer diversas estrategias que permitan un apoyo eficaz y aportes significativos en la superación de dichas falencias, el anterior es un aporte interesante de este trabajo de investigación.

3.3. Investigaciones sobre el uso de técnicas de Machine Learning en la educación.

Giraldo (2017) desarrolla investigación que propone un acercamiento a un modelo de descubrimiento de patrones basado en la observación y el análisis de datos educativos, lo mismo que, en las interacciones existentes entre estudiantes y plataformas virtuales de aprendizaje. El estudio tuvo el propósito de hacer algunas recomendaciones que permitieran adaptar y posiblemente personalizar el proceso de aprendizaje de acuerdo con las características e interacciones de los estudiantes (Giraldo, 2017).

El modelo se validó mediante un caso de estudio que contó con datos académicos de estudiantes de la Universidad Nacional de Colombia Sede Manizales durante los periodos comprendidos entre el primer semestre de 2009 - 2015 y sus interacciones en plataformas virtuales de aprendizaje entre el segundo semestre de 2013 - 2014.

Los resultados muestran que es posible implementar un modelo basado en técnicas de minería de datos que permita extraer información derivada del cruce de datos (hechos) guardados en diferentes matrices para almacenarla de forma coherente y organizada. Tal información se analiza con el fin de descubrir tendencias o patrones de interacciones entre estudiantes y

plataformas virtuales de educación mediante el uso de analíticas de aprendizaje; el modelo requiere que se cuente con los datos necesarios y suficientes.

El autor concluye que las instituciones educativas recolectan y generan gran cantidad de información de sus estudiantes sobre desempeño académico, condición socioeconómica, situación sociocultural, hábitos de estudio, horas de entrada, salida, tiempos de permanencia en las plataformas virtuales de educación entre otras, que se puede aprovechar para descubrir tendencias, predecir futuros desempeños y sugerir estrategias de aprendizaje acordes con las características de cada estudiante. Esta conclusión concuerda totalmente con los propósitos de del proyecto de investigación que se adelanta.

De esta investigación se puede recalcar la importancia de los datos recolectados, ya que, gracias a su valiosa utilidad en las plataformas virtuales de educación, se pueden llegar a predecir desempeños académicos y de igual forma sugerir estrategias de aprendizaje que den respuestas prácticas a las necesidades de los estudiantes.

Sposito, Etcheverry, Ryckeboer, y Bossero (2008) realizaron una investigación cuyo objetivo principal fue maximizar la calidad que los modelos ofrecen para detectar los patrones determinantes de algún factor a fin de clasificar y agrupar a los estudiantes de acuerdo con características académicas, factores sociales y demográficos, en particular se tuvo en cuenta la evaluación del desempeño académico y la deserción de los estudiantes del Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT) de la Universidad Nacional de La Matanza, Argentina (UNLaM) (Sposito et al.,2008).

La investigación aplicó el proceso de descubrimiento de conocimiento sobre los datos de alumnos del período 2003-2008 y usó el software MS SQL Server para generar el almacén de datos, también hizo un pre procesamiento de los datos con el software SPSS y con el software

Weka (Waikato Environment for Knowledge Analysis) se detectaron los patrones determinantes de la deserción estudiantil.

Los resultados de la investigación reportan la generación de un almacén de datos con base en las variables: planes de estudio, carrera, asignaturas, exámenes y otros datos de los colegios de educación secundaria donde habían estudiado; la información se tomó del almacén de datos de la UNLaM y de las bases de datos tanto del DIIT como de colegios de educación secundaria del Ministerio de Educación.

Como resultado del estudio se consolidó un grupo de investigación en las técnicas de minería de datos y se implementó un almacén de datos que permite tomar decisiones sobre los estudiantes con menor tiempo e incertidumbre, sin embargo, no se logró encontrar un clasificador del desempeño académico y de deserción estudiantil con un alto grado de precisión y compresibilidad; como aporte del estudio se resalta la gran cantidad de información que se genera por las distintas áreas de cualquier institución que conlleva a utilizar imprescindiblemente las Tecnologías de la Información y la Comunicación (TIC) para que la información pueda ser almacenada, transformada, analizada y visualizada.

De esta investigación se resalta la importancia de la recolección de datos y el uso de técnicas de minería de datos para monitorear el desempeño de los estudiantes con mayor posibilidad de deserción escolar.

Timarán y Jiménez (2014) realizaron trabajo de investigación con el uso de metodología de minería de datos CRISP-DM cuyo objetivo fue detectar patrones de deserción estudiantil a partir de los datos socioeconómicos y académicos de los estudiantes de los programas de pregrado de la Universidad de Nariño e Institución Universitaria CESMAG; como resultado se descubrieron perfiles socioeconómicos y académicos de los estudiantes y se construyó un

repositorio de datos con la información de los estudiantes que ingresaron a estas universidades entre el primer semestre de 2004 y segundo semestre de 2006 con proyecciones de observación hasta el 2011 (Timarán y Jiménez, 2014)

Del estudio se generaron reportes y un patrón general de deserción estudiantil común para las dos universidades que muestra promedio de notas bajo, pérdida de asignaturas en los primeros semestres de la carrera y un bajo puntaje promedio en pruebas de ICFES. Tales patrones sirvieron para soportar la toma de decisiones eficaces por parte de las directivas universitarias, quienes se enfocaron en formular políticas y estrategias relacionadas con los programas de retención estudiantil.

No obstante, los buenos resultados, se presentaron dificultades debido a la mala calidad de los datos guardados en la base de datos de las 2 universidades, las cuales tuvieron que hacer el filtrado de datos para descartar ciertas variables que no presentaban sus valores y que de alguna manera influyen en los resultados obtenidos sobre deserción estudiantil. Este antecedente es un referente esencial del estudio que actualmente se desarrolla.

De esta investigación cabe resaltar que gracias a la metodología CRISP-DM se pueden detectar patrones en común basados en datos socioeconómicos y académicos de los estudiantes que pueden tener un bajo desempeño frente a la deserción escolar, permitiéndole a los directivos de las universidades anticiparse a esta situación, creando estrategias para prevenir la deserción.

Duran, Costaguta, Maldonado, Unzaga, Chequer, Menini & Fernández (2010), en su investigación “*Técnicas de machine learning y personalización en educación*” estudian los fundamentos conceptuales, metodológicos y técnicos del machine learning, lo mismo que, la personalización del aprendizaje a través del diseño, desarrollo y evaluación de aplicaciones

computacionales específicas capaces de adaptar la enseñanza a las características y necesidades individuales de cada estudiante.

Según los autores, estas tecnologías computarizadas aportan a la educación aspectos innovadores que permiten mejorar las formas de enseñar y aprender.

Los autores concluyen que el proyecto contribuirá a la formación de los investigadores involucrados y también a la formación de estudiantes avanzados de informática. También se espera contribuir a la capacitación de dichos estudiantes en teorías, métodos y técnicas vinculadas con la personalización de sistemas en el ámbito de la educación. La investigación se destaca la importancia del machine learning ya que permite modificar y adaptar la enseñanza a las necesidades individuales de cada estudiante, permitiendo mejorar su formación por parte de las instituciones educativas.

Sison y Shimura (1998) en su investigación “modelado estudiantil y machine learning” evalúa cómo se han utilizado las técnicas de machine learning para automatizar la construcción de modelos de estudiantes, así como los conocimientos básicos necesarios para el estudiante modelado. Según los autores, el modelado de estudiantes implica la construcción de una representación cualitativa que tenga en cuenta el comportamiento de los estudiantes en términos de conocimientos previos existentes sobre un dominio y sobre los estudiantes que aprenden el dominio. El estudio reporta resultados que dan cuenta de la dificultad, la idoneidad y el potencial de usar el machine learning para encontrar los procesos de modelado de estudiantes y en menor medida, el potencial de usar técnicas de modelado de estudiantes en el machine learning (Sison y Shimura, 1998).

Los autores concluyen que los elementos esenciales para definir un modelo de estudiante son: su conocimiento, su comportamiento y su conocimiento básico, también identifican varios

problemas importantes del modelo de estudiante, lo mismo que, los principales enfoques y paradigmas del machine learning y un breve esbozo de cada uno; además, examinan las diversas formas en que se han utilizado las técnicas de machine learning para modelar el estudiante y para construir los conocimientos básicos necesarios para definir un modelo de estudiante, los autores señalan que faltan muchas cosas por hacer.

Berens, Schneider, Görtz, Oster, y Burghoff (2016) realizan una investigación titulada “Detección temprana de estudiantes en Riesgo: predecir el abandono escolar mediante el uso de datos administrativos de estudiantes y métodos de machine learning” cuyo propósito consistió en identificar los determinantes subyacentes de la deserción y los estudiantes que están en riesgo de abandonar la escuela; conocer las causas de la deserción es el presupuesto básico para reducir exitosamente. (Berens et al., 2016)

El estudio usó como metodología un sistema de detección temprana EDS (early detection system) en universidades de tamaño mediano en el estado federal de Renania del Norte-Westfalia, Alemania; en el desarrollo del estudio se trabajó con una muestra aproximada de 29.700 estudiantes y 116 programas de pregrado de dos universidades, una de ellas pública y la otra privada; 23.000 estudiantes y 90 programas de licenciatura corresponden a la universidad pública y los demás a una universidad privada de ciencias aplicadas.

Para realizar el estudio se seleccionaron y aplicaron varios métodos tales como: modelos de regresión, métodos de machine learning y un algoritmo de impulso denominado “algoritmo AdaBoost” que combina los diversos métodos empleados; también se emplearon algoritmos de machine learning que no se basan en la construcción de modelos complejos y que se autoajustan cada vez que hay nuevos datos disponibles. El uso combinado de varios métodos reduce las desventajas inherentes al uso de cualquier método individual.

Los resultados indican que el 74% de los abandonos se presenta en la universidad pública y el 72% de los abandonos se identifican correctamente al final del primer semestre en la universidad privada de ciencias aplicadas, además, la precisión del EDS aumenta a medida que los nuevos datos de desempeño de los estudiantes están disponibles al final de cada semestre consecutivo; después del cuarto semestre, el EDS predice correctamente el 80% de la deserción en la universidad pública y el 83% de los estudiantes que abandonaron la universidad privada de ciencias aplicadas; sin embargo, la confirmación de estudios anteriores y los datos de desempeño en particular en las primeras etapas son importantes para predecir el desgaste de los estudiantes, mientras que los datos demográficos tienen un valor predictivo limitado una vez que los datos de desempeño están disponibles.

Los autores concluyen que la deserción universitaria es un tema importante para la política educativa; esta debe detectarse tempranamente para prevenir sus consecuencias a través del diagnóstico y para reducirla al máximo mediante la intervención. Se sabe que, la deserción escolar es costosa para todas las partes involucradas; en este sentido, el estado invierta gran cantidad de recursos que se desperdician y los estudiantes pierden tiempo y esfuerzo en mantenerse en el sistema sin que se logre certificar, de igual forma, un estudiante que deserta tiene un valor económico limitado para el sistema universitario cuando no están acompañado por un certificado de graduación.

La ventaja del sistema de detección temprana utilizado es que después de haber identificado a los estudiantes en riesgo, puede servir como base para un sistema de intervención temprana para prevenir el abandono escolar o incluso acelerar la decisión del alumno de abandonar sus estudios y de esta manera los costos públicos y privados asociados con la deserción pueden reducirse mediante la implementación de este sistema como punto de partida

para asignar el apoyo de intervención a los estudiantes en riesgo y para probar la efectividad de la intervención en el estudiante. La predicción de las consecuencias de la deserción escolar es un factor crítico, una alternativa para evitarla es usar un algoritmo que permite predecir la deserción de estudiantes por semestre y facilitar la creación de estrategias para disminuir la deserción escolar haciendo un diagnóstico que permite realizar una intervención oportuna.

4. MARCO TEÓRICO

El pensamiento espacial o razonamiento espacial constituye un componente esencial del pensamiento matemático, en cuanto tiene que ver con la solución de problemas que implica la interacción y el establecimiento de relaciones con los objetos y con el entorno que rodea al sujeto. Dada su importancia en el dominio matemático y geométrico, en adelante se presentan algunas posiciones teóricas de autores que han avanzado en la precisión y definición este concepto.

Para la educación, el individuo que no desarrolla el pensamiento espacial tiene dificultad para interpretar y comprender el mundo físico, manifestando un bajo nivel de interés matemático y falta de destreza numérica y conceptual; de tal forma que, no intuye un entorno propio (Gardner, 2011).

En el caso específico de Colombia, más concreto aún, en las Orientaciones curriculares para el campo de Pensamiento Matemático de la alcaldía mayor de Bogotá se cita a (Vasco, 2006). El cual define el pensamiento espacial como el conjunto de procesos cognitivos mediante los cuales se construyen y manipulan las representaciones mentales de los objetos del espacio, las relaciones entre ellos, sus transformaciones, y sus diversas traducciones o representaciones materiales, contempla las actuaciones del sujeto en todas sus dimensiones y relaciones espaciales para interactuar de diversas maneras con los objetos situados en el espacio, desarrollar variadas representaciones y, a través de la coordinación entre ellas, hacer acercamientos conceptuales que favorezcan la creación y manipulación de nuevas representaciones mentales.

Piaget por ejemplo: realizó de manera sistemática una serie de trabajos orientados a comprender el desarrollo cognitivo del ser humano, en algunos de su trabajos también estudió el

concepto de espacio, al respecto señala que, el espacio no viene dado “a priori”, es decir, el sujeto no nace con la idea de espacio ni surge de la mera percepción humana, por el contrario, sostiene que se elabora poco a poco en la medida que la persona actúa sobre el mundo; en otros términos, el espacio se construye gracias al papel decisivo de la actividad del sujeto sobre sí mismo y sobre su entorno (Piaget & Inhelder, 1997).

Piaget agrega que, el conocimiento espacial surge y se desarrolla en tres grandes periodos o estadios: 1)- sensoriomotor, 2)- operaciones concretas y 3)- operaciones formales. En el primer periodo, el niño elabora el concepto de espacio topológico como resultado de la actividad sensorio-motriz; en un segundo momento, el niño representa mentalmente el espacio, tal representación le permiten flexibilizar, coordinar y revertir las imágenes espaciales para convertirlas luego en operaciones y actuaciones sobre el mundo externo, finalmente el niño se aparta de la interacción concreta con objetos reales y alcanza el estadio de las propiedades espaciales que involucran la métrica, esto es, el conocimiento formal de la geometría; ese largo proceso de elaboración implica un producto final de larga y ardua construcción evolutiva del concepto de espacio.

Otros teóricos entre ellos Galvez (1994), propone una nueva perspectiva del pensamiento espacial en la que expone la importancia del sujeto y de su percepción teniendo en cuenta el desplazamiento, la localización y la ubicación de las relaciones básicas entre el sujeto, el objeto y el espacio (Galvez, 1994)

En el mismo sentido, el MEN en su propuesta de Renovación Curricular enfatiza en el estudio activo de los sistemas geométricos como herramientas de exploración y representación del espacio, debido a que en los sistemas geométricos se hace énfasis en el desarrollo del pensamiento espacial (MEN, 1998). En su propuesta curricular el MEN

retoma en términos generales los presupuestos de Piaget en materia de pensamiento espacial, en este sentido, considera que la construcción del pensamiento espacial se entiende como un proceso cognitivo de interacciones que avanza desde el espacio intuitivo o sensorio – motor hasta un espacio conceptual o abstracto relacionado con la capacidad de representar internamente el espacio (MEN, 1998).

De otro lado, Gardner (2011) se aparta un poco de la tendencia Piagetiana y en vez de pensamiento espacial, define la inteligencia espacial como “la habilidad para resolver problemas o para elaborar productos que son de importancia en un contexto cultural o en una comunidad determinada” (Gardner, 2011) él insiste en que la inteligencia espacial es la capacidad de pensar en tres dimensiones: permite percibir imágenes externas e internas, recrearlas, transformarlas o modificarlas, recorrer el espacio o hacer que los objetos lo recorran y producir o decodificar información gráfica. Hace notar mediante esta inteligencia que, si bien entre espacio y mundo visual parece haber una correlación directa en la inmensa mayoría de seres no es menos cierto que, el espacio tiene una connotación equivalentemente significativa en el mundo visual.

Gardner señala siete inteligencias que implican siete habilidades diferentes para resolver situaciones problemáticas y que son las que permiten la conformación de un grupo de trabajo eficiente al ser observadas detalladamente, es decir, que la resolución de cualquier problema, en este caso de manejo espacial, requiere poner en juego algunas de tales inteligencias.

La propuesta de investigación que se adelanta toma el modelo teórico de Piaget y los planteamientos del MEN. Con respecto al modelo teórico de Piaget relacionado con el manejo del concepto de pensamiento espacial se profundiza en relaciones espaciales.

4.1. Relaciones espaciales según Piaget

Piaget (1948) establece tres aspectos fundamentales en el desarrollo del conocimiento espacial en los niños que siguen vigentes en la comprensión de lo espacial, al respecto plantea:

1. El espacio no surge de la mera percepción sino de la actividad sensoriomotriz que luego será representada mentalmente, lo que permitirá flexibilizar, coordinar y reversar las imágenes espaciales para convertirlas en operaciones, actuaciones sobre el mundo externo. Lo que implica un producto final de larga y ardua construcción evolutiva que inicia en el nacimiento y termina en la adolescencia.
2. Las relaciones espaciales se clasifican en topológicas, proyectivas y euclidianas; las relaciones topológicas consideran el espacio dentro de un objeto o figura particular y comprenden relaciones de proximidad, separación, orden, cerramiento y continuidad, mientras que los espacios proyectivos y euclidianos tienen en cuenta las relaciones entre los objetos de acuerdo con sistemas proyectivos o con ejes coordenados (ver figura 2).
3. Es importante tener en cuenta que Piaget enfatiza el hecho de que el desarrollo de estas relaciones es inverso al desarrollo epistemológico, pues el niño elabora primero el espacio topológico de manera activa y representativa, mientras que las relaciones proyectivas y euclidianas se desarrollan paralelamente, logrando el equilibrio de las euclidianas más tarde (Piaget, 1948).

4.1.1. Relaciones espaciales:

En referencia a las relaciones espaciales Piaget (1947) plantea:

1. El espacio topológico elemental se refiere a relaciones y posiciones relativas que existen entre los objetos como vecindad, aproximación, separación, seriación, envolvimiento, continuación dando lugar a nociones de cerca, lejos, arriba, abajo, adelante, atrás.
2. El espacio proyectivo considera los objetos y sus representaciones teniendo en cuenta las relaciones entre los objetos, desde su inicio cuando se establece como se representa el niño la perspectiva de un objeto simple y luego cómo aborda la investigación de la comprensión de la perspectiva de un grupo de objetos.
3. El paso del espacio proyectivo al espacio euclidiano establece relaciones entre las operaciones proyectivas y euclidianas teniendo en cuenta las relaciones entre los objetos y sus representaciones de acuerdo con sistemas proyectivos o ejes coordenados. (Piaget, 1947).

En el planteamiento de su teoría, Piaget establece cuatro etapas, que integradas con los períodos mencionados cumplen con la función de la sistematización de los resultados experimentales y que demuestran que el niño adquiere la noción del espacio con cierta lentitud debido a las referencias sensibles que posee. La tabla 1 recoge estas etapas.

Tabla 1.
Etapas del Proceso de Conocimiento del Espacio según Jean Piaget

Etapa	Proceso
Primera I	Desde los tres años hasta los cuatro años y medio. Es posible entrevistar a los niños.
Segunda II	Desde los cuatro años y medio hasta los seis – siete años Comienzo de las operaciones concretas.
Tercera III	Desde los seis – siete años hasta los once – doce años Completamiento de las operaciones concretas
Cuarta IV	De los doce años a los quince años Desarrollo de las operaciones formales

Fuente: Teoría de Piaget sobre el Desarrollo del Conocimiento Espacial (1947)

4.2. Características del Pensamiento espacial según el MEN

Con relación a los planteamientos del MEN (1998) este sostiene que el concepto de pensamiento espacial está condicionado por las características cognitivas individuales y el entorno físico, cultural, social e histórico, en un proceso de influencia mutua construyéndose a partir de las actuaciones del sujeto y de su interacción con los objetos situados en el espacio. Cuando el estudio de las propiedades espaciales involucra la métrica, este se convierte en conocimiento formal de la geometría que, a su vez, se relaciona con el arte, la decoración, el diseño, la construcción de objetos artesanales y tecnológicos, la educación física, los deportes, la danza; con la observación y reproducción de patrones (como ocurre en las plantas, animales u otros fenómenos de la naturaleza) y con otras formas de lectura del espacio como los mapas y representaciones a escala de sitios o regiones en dibujos y maquetas, (MEN, 1998). Tales conocimientos, aplicados en un campo pragmático, corresponden a diversas profesiones u oficios: arquitectos, ingenieros, astrónomos, artistas, bailarines, pintores, tecnólogos, entre otros.

4.2.1. Características cognitivas

A nivel nacional el MEN (1998), plantea en los Lineamientos Curriculares de Matemáticas que una característica fundamental del pensamiento espacial es la exploración activa del espacio tridimensional en el mundo externo y en la imaginación, con el fin de facilitar la realización de representaciones, estáticas y dinámicas, de los objetos en el espacio. Para llegar a este nivel son importantes las representaciones planas de las formas y las relaciones tridimensionales, las cuales se clasifican en:

4.2.2. Características individuales

Gardner (2011) en su teoría de las múltiples inteligencias considera que el desarrollo del pensamiento espacial es esencial para el pensamiento científico, ya que es usado para representar y manipular información en el aprendizaje y en la resolución de problemas. El manejo de la información espacial facilita la resolución de problemas de ubicación, orientación y distribución de espacios. El autor estima que la mayoría de las profesiones científicas y técnicas, tales como el dibujo técnico, la arquitectura, las ingenierías, la aviación, y muchas disciplinas científicas como química, física, matemáticas, requieren personas que tengan un alto desarrollo en su nivel de pensamiento espacial (Gardner, 2011).

4.2.3. Características del entorno físico, cultural, social e histórico

Las Orientaciones curriculares para el campo de Pensamiento Matemático (2007) establecen que el desarrollo del pensamiento espacial en el estudiante es un proceso lento y se nutre de las experiencias obtenidas del mundo material, simbólico, social y cultural, pero en este caso, las experiencias con el mundo de los objetos se constituyen en un soporte intuitivo mayor.

Los estudiantes construyen un espacio práctico como resultado de las acciones que hacen sobre y con los objetos del mundo. Desplazan objetos y se desplazan de un sitio a otro. Para dar cuenta de su localización y de los objetos construyen sistemas de referencia. Al comienzo, de niños, únicamente están en capacidad de usar sistemas referidos al propio cuerpo o a objetos que están en el mismo lugar y cercanos a ellos, por eso estos sistemas son muy locales y fragmentados.

Poco a poco se hacen capaces de usar objetos muy conocidos (la casa en la que viven, el parque, la escuela) para utilizarlos como referencia en espacios más amplios. De igual forma se

transforman y representan los objetos y se modifican sus posiciones, aunque se desplacen.

Algunas acciones hechas sobre los objetos modifican su forma y su tamaño. Hay otras acciones que mantienen la forma, pero no el tamaño u otras que mantienen ambas. Aunque de pequeños los estudiantes no sean conscientes de estos hechos, los utilizan en muchas acciones prácticas. Desde pequeños se comparan propiedades de los objetos que dan cuenta de su tamaño o del de algunos de sus componentes, aunque al empezar sea de forma rudimentaria. Poco a poco se introducen procedimientos que permiten independizar paulatinamente las comparaciones de la valoración perceptiva.

4.2.4. Estrategias pedagógicas

El MEN (1998) plantea que la apropiación por parte de los estudiantes del pensamiento espacial requiere del estudio de distintas relaciones espaciales de los cuerpos sólidos con respecto a los mismos estudiantes; de cada cuerpo sólido con sus formas, caras, bordes, vértices, superficies, regiones y figuras planas con sus fronteras, se deben fortalecer los procesos de localización en relación con sistemas de referencia y del estudio de lo que cambia o se mantiene en las formas geométricas bajo distintas transformaciones. El trabajo con objetos bidimensionales y tridimensionales, sus movimientos y transformaciones permiten integrar nociones sobre volumen, área y perímetro, lo cual a su vez posibilita conexiones con los sistemas métricos o de medida y con las nociones de simetría, semejanza y congruencia, entre otras.

De igual forma el MEN (1998) presenta a la geometría activa como una alternativa para refinar el pensamiento espacial, puesto que se constituye como una herramienta privilegiada de exploración y representación del espacio que parte de la actividad del alumno y su confrontación con el mundo dándole prioridad a la actividad sobre la contemplación pasiva de figuras,

símbolos y operaciones sobre las relaciones y elementos de los sistemas y a la importancia de las transformaciones en la comprensión aun de aquellos conceptos que a primera vista parecen estáticos.

La geometría activa plantea que el estudiante debe moverse, dibujar, construir, producir y tomar de estos esquemas operatorios el material para la conceptualización o representación interna que va acompañada en un principio por gestos y palabras del lenguaje ordinario, hasta que los conceptos están incipientemente contruidos a un nivel suficientemente estable para que los alumnos mismos puedan proponer y evaluar posibles definiciones y simbolismos formales.

Vasco (2006) determina que la geometría activa consiste en:

“la exploración de la figura mediante el movimiento, empezando por el propio cuerpo, (como cuando el niño recorre la frontera de una figura) y pasando por el que se aplica a los objetos físicos, para estudiar los efectos que se producen en la figura que comportan y las relaciones entre productos de estos movimientos y de manera muy parcial, entre los mismos movimientos” (Vasco, 2006).

Según las Orientaciones curriculares para el campo de Pensamiento Matemático (2007) se determina que para que el contexto influya benéficamente como recurso del proceso enseñanza es necesaria la intervención permanente del docente, quien es el encargado de modificarlo y enriquecerlo con el propósito de que los estudiantes aprendan que los objetos están dispuestos en el mundo y que ellos deben establecer relaciones entre dichos para dar cuenta de sus posiciones y de los cambios de éstas, comprendiendo que este sistema de relaciones es lo que hace referencia a la noción que los sujetos construyen de espacio.

Tal forma de entenderlo es importante para la enseñanza porque destaca el carácter constructivo del pensamiento espacial. Por una parte, pone de relieve que este no es resultado de impresión de imágenes fruto de experiencias con los cuerpos, de sus formas y de sus posiciones, sino de operar con las relaciones que se establecen entre ellas y, por otro lado, que para movilizar el desarrollo del pensamiento espacial, la enseñanza debe orientar su esfuerzo en enriquecer la experiencia y la reflexión de los estudiantes con el espacio.

Vasco (2006) plantea que desde muy pequeños los estudiantes intentan construir objetos conocidos y presentes, llegando a representar objetos ausentes mediante modelos físicos (tipo maquetas) y/o dibujos. En un comienzo estas representaciones son incompletas e imprecisas, pero, poco a poco se logran mejorar. Estas adquisiciones hacen relativamente eficientes a los estudiantes para vivir en el espacio, así sea local, sin necesidad de mayor instrucción formal; este es un espacio práctico o empírico, pero cualquiera de las dos permite realizaciones con cierto grado de complejidad, que van mucho más allá de desplazarse entre dos sitios conocidos, sin perder y de poder imaginar el camino (Vasco, 2006).

Por ejemplo, los estudiantes pueden imaginar la acción de cortar una tabla de forma cuadrada por una de sus diagonales y anticipar la forma y el tamaño de los dos pedazos. Para resolver este problema a nivel práctico no se necesita conocer el nombre de la forma de la tabla y menos de la línea de corte; simplemente el apoyo intuitivo de muchas acciones ofrece la evidencia necesaria para anticipar este resultado, e incluso da la certeza de que se cumple para todos los casos. Este espacio práctico se va complejizando con mayor capacidad de representarlo y de coordinar acciones para obtener resultados más avanzados. Este es un proceso en el cual la enseñanza escolar juega un papel importante, ya que apoya al estudiante en la apropiación de las herramientas de pensamiento que brinda la geometría.

Un carpintero, incluso si es iletrado, da muestras de una profunda capacidad de representarse mentalmente el mueble que va a construir, del número de piezas que lo conformarán, de sus formas y sus tamaños, aunque si le pidiera que lo dibujara, su “croquis” poco respetaría los principios de la perspectiva y de las proporciones; la progresiva geometrización de este espacio permite construir un espacio abstracto y conceptualizado, que produce entidades teóricas representadas simbólicamente, para las cuales los estudiantes no pueden construir imágenes mentales.

De lo anterior se desprende que el niño llega a la institución con construcciones importantes sobre el espacio, pero el papel de la institución consistirá en reconocerlas y enriquecerlas. Pero habría que preguntarse si lo que está ofreciendo la institución logra potenciar el pensamiento espacial de los alumnos, tanto como es posible y deseable (Vasco, 2006).

De igual forma al presentar la propuesta de Renovación Curricular, enfatiza a la geometría activa como el camino correcto del estudio de los sistemas geométricos como herramientas de exploración y representación del espacio, ya que en los sistemas geométricos se hace énfasis en el desarrollo del pensamiento espacial.

Más adelante manifiesta:

Esta construcción se entiende como un proceso cognitivo de interacciones, que avanza desde el espacio intuitivo o sensorio – motor (que se relaciona con la capacidad práctica de actuar en el espacio, manipulando objetos, localizando situaciones en el entorno y efectuando desplazamientos, medidas, cálculos espaciales, etc.), a un espacio conceptual o abstracto relacionado con la capacidad de representar internamente el espacio, reflexionando y razonando sobre propiedades geométricas abstractas, tomando sistemas de referencia y prediciendo los resultados de manipulaciones mentales (MEN, 1998, p. 40).

Con el propósito de operacionalizar la propuesta el Ministerio de Educación sugiere que el currículo debe estar basado en tres grandes aspectos, que son:

- Los procesos generales que tienen que ver con el aprendizaje, tales como el razonamiento; la resolución y planteamiento de problemas; la comunicación; la modelación y la elaboración, comparación y ejercitación de procedimientos.
- Los conocimientos básicos que permiten procesos específicos que desarrollan el pensamiento matemático y con sistemas propios de las matemáticas, propuestos así: pensamiento numérico y sistemas numéricos; pensamiento espacial y sistemas geométricos; pensamiento métrico y sistemas de medidas; pensamiento aleatorio y los sistemas de datos; pensamiento variacional y sistemas algebraicos y analíticos.

Cuando habla sobre el pensamiento espacial y sistemas geométricos como elemento fundamental de los sistemas geométricos plantea: “Desde un punto de vista didáctico, científico e histórico, actualmente se considera una necesidad ineludible volver a recuperar el sentido espacial intuitivo en toda la matemática, no sólo en lo que se refiere a la geometría”. (MEN, 1998)

También considera el MEN que las características cognitivas individuales, la influencia del entorno físico, cultural, social e histórico siempre condicionan el proceso de construcción del espacio. Por lo que el estudio de la geometría en la escuela debe ser fundamental y favorecer estas interacciones.

Para darle sentido al aprendizaje de las matemáticas, dice el MEN, se deben tener en cuenta los ambientes que rodean al estudiante, al igual que las variables como las condiciones

sociales y culturales locales e internacionales, las diversas interacciones, sus intereses, sus creencias, sus condiciones económicas y las del grupo social de la institución.

4.3. Teorías sobre el pensamiento espacial

El pensamiento espacial o razonamiento espacial es la habilidad que tienen la mayoría de las personas de visualizar lo inexistente, crearlo y poder manipularlo en el espacio real. Este pensamiento constituye un componente esencial del pensamiento matemático y se refiere a la percepción intuitiva o racional del entorno propio de los objetos que hay en él. El desarrollo del pensamiento espacial, asociado a la interpretación y comprensión del mundo físico, permite desarrollar interés matemático y mejorar estructuras conceptuales y destrezas matemáticas.

Existen tres grandes autores del Pensamiento Espacial (Piaget, 1947; Galvez, 1994; Gardner, 2011) y cuyas teorías se analizaron (Ver figura 1) para el desarrollo del proyecto, optando por la Teoría de Piaget sobre el desarrollo del pensamiento espacial.

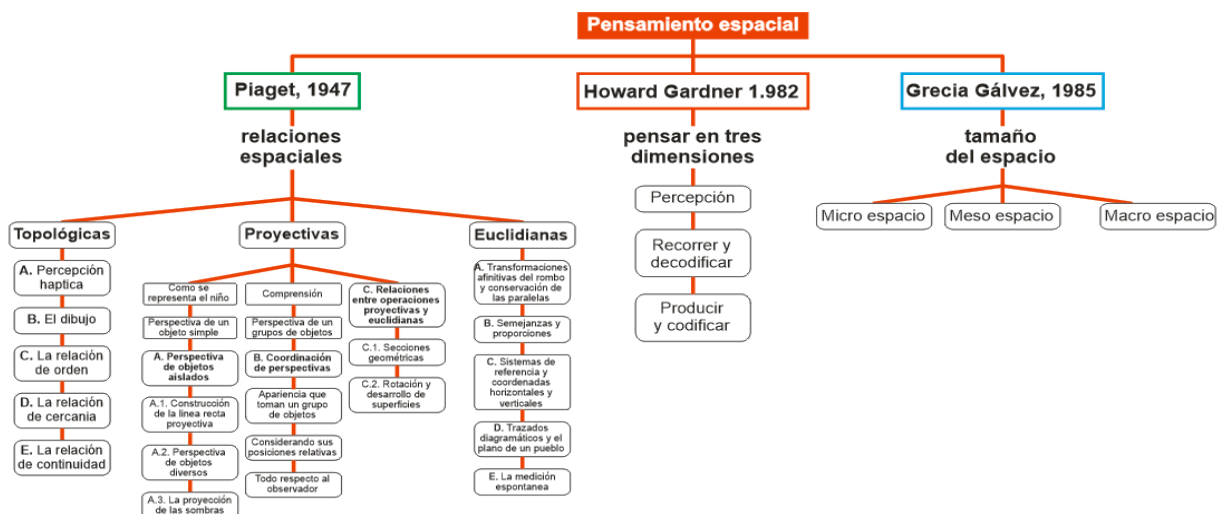


Figura 1. Teorías sobre Pensamiento Espacial

Entre los teóricos del Desarrollo del Pensamiento Espacial, Piaget (1947) estableció que los sujetos desde el momento de su concepción empiezan a establecer una relación directa con el espacio. Este desarrollo que en principio es “rudimentario” se va construyendo progresivamente a partir de las experiencias de desplazamiento del sujeto. Establece como surgen las relaciones topológicas, proyectivas y euclidianas.

Explica cómo el espacio no se define solo por percepción sino que se va desarrollando poco a poco, en donde interviene fundamentalmente la actividad del sujeto, por lo que establece que los tres tipos de relaciones topológicas, proyectivas y euclidianas se van desarrollando durante tres grandes períodos o estadios: período sensorio motor, periodo de las operaciones concretas que comprende el subperíodo preoperativo y otro de operaciones concretas propiamente dichas y el período de las operaciones formales, los mismos estadios que el desarrollo de la inteligencia.

Este planteamiento conlleva a establecer una relación directa entre la edad y la actividad del sujeto, planteando que las relaciones topológicas se desarrollan desde el nacimiento hasta siete años que cubre el período sensorio motor y el subperíodo preoperativo.

A partir de los siete años hasta los once el sujeto va considerando progresivamente las relaciones proyectivas, basado en las relaciones topológicas, y es a partir de los doce años que su pensamiento operativo concreto le facilita la flexibilidad y reversibilidad del espacio permitiéndole examinar las relaciones entre objetos desde fuera de su punto de vista.

Por eso, el estudio aquí presentado se centra en el estudio de las relaciones proyectivas. Los estudiantes intervenidos para las pruebas se encuentran en ese rango de edad (12 – 14 años) lo que establece un parámetro de referencia que permite el análisis de dicho desarrollo, observando si el estudiante relaciona el objeto estudiado con otros objetos; se descentra de su

punto de vista y lo relaciona con otros; es capaz de predecir qué aspecto tendrá tal o cual objeto visto desde diferentes ángulos.

A la luz de los principios de la Teoría de Piaget (1947) y de los resultados de las pruebas aplicadas es importante confirmar que si el sujeto no desarrolla oportuna y eficazmente estas relaciones, su paso al desarrollo de las relaciones euclidianas, que requieren del desarrollo de las operaciones formales, presentará grandes dificultades en el desarrollo del conocimiento, lo que explica los múltiples inconvenientes de los estudiantes en el estudio de áreas del conocimiento que así lo requieren.

4.3.1. Teoría de Piaget y el espacio Proyectivo

El propósito del estudio realizado es detectar el desempeño que presentan los estudiantes de grados noveno, décimo y once de la Institución Educativa Departamental General Carlos Albán, en el desarrollo del espacio proyectivo del pensamiento espacial. Como esto implica variables como los aspectos cronológicos, sociales, económicos, culturales, históricos, cognitivos y de aprendizaje se ha considerado la importancia de la ‘Teoría de Piaget sobre el desarrollo del pensamiento espacial’ (Piaget & Inhelder, 1997) por lo que se profundiza sobre ella, para lograr el objetivo propuesto.

En la Tabla 2 se observa la explicación de la génesis del espacio proyectivo en el niño.

Tabla 2.

El espacio proyectivo según Jean Piaget

El Espacio Proyectivo		
Apartados	Proceso	
Primero:	Construcción de la línea recta proyectiva	Dibujar una línea recta supone tener en cuenta unas relaciones proyectivas o unas relaciones euclidianas.

Representación, por parte del niño, la perspectiva de un objeto simple. Segundo:	Perspectiva de objetos diversos	Investiga la forma como el niño relaciona objetos aislados, teniendo en cuenta sus desplazamientos y la referencia del observador.
Investigación de la comprensión de la perspectiva de un grupo de objetos.	La proyección de las sombras	Imaginar las posiciones relativas de dos objetos frente a factores externos como una pantalla y la sombra que ella proyecta. Este proceso está compuesto por tres etapas. Etapa I: (Hasta 4 años) los niños ni siquiera comprenden las preguntas que se les hacen. Etapa II: (4 – 7 años) no son capaces de imaginar las transformaciones, por el cambio de posición y considerando como único su punto de vista, llamado por Piaget ‘error egocéntrico’. Sub-etapa II B: Intenta dejar a un lado el egocentrismo. No se da cuenta que los cambios de vista llevan a una inversión de las relaciones izquierda – derecha, delante – atrás. Etapa III: Con la aparición del pensamiento operatorio concreto descubre que la variación de las relaciones depende del punto de vista del observador La comprensión de ellas se da desde la geometría proyectiva como desde la euclidiana.
Tercero: Estudio sobre las relaciones entre las operaciones proyectivas y euclidianas, utilizando para ello secciones geométricas y rotación de superficies.	Secciones Geométricas	La sección euclidiana se da cuando se imagina cómo queda una figura geométrica cortada por un cuchillo. La proyectiva lleva a imaginar la representación de la sección de una figura geométrica proyectada ante una pantalla luminosa Desarrollar en plano una serie de objetos, utilizando la técnica dibujo/y/o reconocimiento.
Rotación y desarrollo de superficies		

Fuente: Teoría de Piaget sobre el Desarrollo del Conocimiento Espacial (1947)

El conocimiento espacial presenta tres grandes períodos o estadios:

El primero de ellos es el período sensorio motor que va desde el nacimiento hasta el año y medio o dos años, durante el cual el pensamiento comienza a interiorizarse, el niño va elaborando un conocimiento del espacio empezando por las relaciones topológicas que presentan

dificultad creciente. En este momento no existe coordinación entre los distintos espacios sensoriales, no percibe los objetos como permanentes, ni los tamaños y las formas como constantes.

Durante los períodos comprendidos entre los 4 a 5 meses y los 10 a 12 meses, la visión y la prensión se coordinan construyendo esquemas manipulativos, iniciando en el niño las relaciones euclidianas lo que se manifiesta al adquirir el objeto un tamaño constante sin depender de la distancia y, proyectivas cuando el tamaño es constante independientemente de la posición.

En el segundo año de vida el niño es capaz de relacionar unos objetos con otros en el espacio descubriendo, como ejemplo, que puede alcanzar un punto en el espacio por dos caminos distintos. El segundo período, de las operaciones concretas, comprende un subperíodo preoperativo, que va desde los 10 – 12 meses hasta los siete u ocho años de edad, en donde el niño reelabora representativamente las adquisiciones realizadas en el estadio anterior comenzando por las relaciones topológicas y continuando con las proyectivas y euclidianas; y el de las operaciones concretas propiamente dicho, que va de los siete u ocho años hasta los once o doce años, y en que se considera progresivamente las relaciones proyectivas y euclidianas ya que puede flexibilizar y revertir el espacio.

Una vez superados en el tiempo los estadios anteriores el niño avanza al estadio de las operaciones formales, que inicia a los once – doce años y culmina en la adolescencia, estadio en donde las operaciones espaciales ya pueden ser separadas de la acción real, lo que les permite considerar un universo de posibilidades espaciales.

De lo propuesto, para el espacio proyectivo por Piaget “la noción de un sólido desarrollado no es una consecuencia directa de la percepción” por lo que esta deriva de una acción no realizada entre la percepción del objeto y su representación en el plano, presente en los

niños pequeños por la ausencia de las operaciones que no les permite observar dicha imagen activa y reversible, capaz de anticipar una acción, que sí está presente en el pensamiento de los niños que están complementando el desarrollo de las operaciones concretas (Piaget & Inhelder, 1997).

En “La representación del espacio en el niño”, Piaget se ocupa de conocer como surgen en el desarrollo ontogenético las relaciones espaciales topológicas, proyectivas y Euclidianas. (Piaget & Inhelder, 1997).

En la segunda obra, *La Geometría Espontánea en el Niño* (1948) se estudia la génesis de la geometría euclidiana, esto es, como surgen en el niño la conservación y la medición de la longitud, la superficie y el volumen (Piaget, 1948).

Para Piaget el estudio del espacio proyectivo está dividido en tres apartados: en el primero explica cómo se representa el niño la perspectiva de un objeto simple, como la construcción de la línea recta; en el segundo apartado investiga sobre la comprensión de un grupo de objetos determinando, por ejemplo la forma de un palo o un disco en diferentes posiciones en el espacio; y en el tercer apartado establece las relaciones entre relaciones proyectivas y euclidianas utilizando ejemplos de secciones geométricas y rotación de superficies en donde el niño debe dibujar o reconocer la forma que toman los objetos colocados en diferentes inclinaciones entre una lámpara y una pantalla vertical.

De igual forma, establece que la diferencia más significativa entre estas relaciones es la forma en que los objetos, figuras o partes se relacionan entre sí. Para él las relaciones topológicas son internas dentro de la figura u objeto, constituyendo relaciones bicontinuas simples entre dos o más figuras, mientras que las proyectivas y euclidianas tienen en cuenta las relaciones entre los objetos de acuerdo con sistemas proyectivos o ejes coordenados.

Los experimentos que le permiten determinar el origen de las operaciones proyectivas y euclidianas los realiza con secciones geométricas y de rotación y desarrollo de superficies, para lo cual recurre al ejemplo de cómo la comprensión de las secciones geométricas en dirección a las relaciones proyectivas se manifiesta cuando el sujeto se imagina la representación de la sección de una figura geométrica proyectada ante una pantalla luminosa.

La importancia de la teoría de Piaget radica en:

Es el único autor que ha estudiado seriamente la evolución del espacio.

En una revisión sistemática de la literatura posterior a la primera parte del siglo, se ha confirmado la trascendencia de su teoría, por el hecho de que la mayoría de los datos sobre el desarrollo del pensamiento espacial, incluyendo la de los medios a gran escala, puede ser cómodamente explicada dentro de este punto de vista teórico. La mayoría del trabajo hecho dentro de un marco de referencia estrictamente conductista es fragmentaria y no puede trazar de forma comprensiva la secuencia del desarrollo. (Hart & Moore, 1971) . Esta teoría ofrece el único modelo teórico que relaciona la evolución del proceso cognitivo con el desarrollo de la cognición espacial.

No obstante, algunas de las dificultades más significativas en el desarrollo del pensamiento espacial, basados en Piaget y ratificadas por Vasco (2006), son:

- La representación de figuras tridimensionales en dibujos planos.
- Las construcciones con regla y el compás que ofrecen una frialdad estética de la figura que solo permite reconocer las características de solo ese objeto y no poder imaginarlo para todas las de su tipo.
- Ausencia de ayudas didácticas que le permitan al estudiante manipular los objetos, para analizarlos desde su perspectiva e ir descubriendo sus relaciones con el espacio.

Estas dificultades son justamente, las que se presentan en los resultados arrojados en las pruebas de pensamiento espacial realizadas por los estudiantes de noveno, décimo y once del colegio Carlos Albán.

4.3.2. Teorías de Howard Gardner y Grecia Gálvez

Gardner (2011) en el libro *Inteligencias Múltiples, la Teoría en la Práctica*, describe la inteligencia como “la habilidad para resolver problemas o para elaborar productos que son de importancia en un contexto cultural o en una comunidad determinada.” (Gardner, 2011).

Para ello insiste en que la inteligencia espacial es la capacidad de pensar en tres dimensiones, permite percibir imágenes externas e internas, recrearlas, transformarlas o modificarlas, recorrer el espacio o hacer que los objetos lo recorran y producir o decodificar información gráfica. (Gardner, 2011).

Hace notar mediante esta inteligencia que, si bien entre espacio y mundo visual parece haber una correlación directa en la inmensa mayoría de seres, no es menos cierto que el espacio tiene una connotación equivalentemente significativa en el mundo visual.

El análisis de la Teoría de Gardner conlleva a pensar en inteligencias diferenciadas en las personas, grados y combinaciones diferentes de acuerdo con sus capacidades. Dentro de sus aportes se encuentra la opción de tenerlas presentes al estudiar las habilidades mentales que posee un grupo de trabajo según las inteligencias que presentan sus integrantes.

Considera siete inteligencias que implican siete habilidades diferentes para resolver situaciones problémicas y que son las que permiten al ser observadas detalladamente, para la conformación de un grupo de trabajo eficiente.

En sus aportes teóricos Galvez (1994), establece cómo funcionan las relaciones didácticas en el aprendizaje de las matemáticas. Plantea la importancia de distinguir la incidencia de la variable tamaño del espacio, diferenciando los problemas que se presentan en el microespacio, el mesoespacio y el macroespacio (Galvez, 1994).

El microespacio “Corresponde a un sector del espacio próximo al sujeto y que contiene objetos accesibles tanto a la visión como a la manipulación. (...) es el espacio del objeto, frente al cual se sitúa el sujeto, pero desde afuera” (Galvez, 1994).

El mesoespacio “Es una parte del espacio accesible a una visión global, obtenida a partir de percepciones sucesivas, pero con desfases temporales mínimos. Contiene objetos fijos no manipulables” (Galvez, 1994).

Explica que el macro espacio “... es imposible de percibir globalmente, requiere que haya desplazamiento por parte del sujeto y que este vaya integrando, con continuidad, diferentes visiones obtenidas por desplazamiento sobre la superficie terrestre, lo que demanda una conceptualización, esto es una representación global.” (Galvez, 1994).

En 1994 propone una nueva perspectiva importante a considerar en la construcción del espacio: el tamaño del espacio. Hablando del microespacio y el macroespacio, en donde expone la importancia del sujeto y de su percepción teniendo en cuenta el desplazamiento, la localización y la ubicación de las relaciones básicas entre el sujeto, el objeto y el espacio. (Galvez, 1994).

4.4. Minería de Datos y Machine Learning

Según (Mc Cormick y Salcedo,2017; Tamoghna, Raghav, y Dipanjan, 2018) la minería de datos es el conjunto de técnicas y tecnologías que permiten explorar bases de datos amplias, para encontrar patrones que aporten información valiosa para la toma de decisiones.

Teóricos como Ramírez, Hernández, y Ramírez (2004) en su libro Introducción a la Minería de Datos coinciden en afirmar que el proceso de Minería de Datos considera cuatro etapas:

1. Determinación de los objetivos. consiste en identificar el tipo de información que se desea extraer de la base de datos.
2. Procesamiento de los datos. En esta etapa se selecciona, limpia, enriquece, reduce y transforma la base de datos que ha facilitado el interesado.
3. Determinación del Modelo. Esta etapa permite escoger el algoritmo a utilizar con el fin de estudiar a profundidad una tarea específica, a través de los siguientes tipos de actividades que realiza normalmente cada algoritmo. Hay algoritmos para predecir un atributo discreto, para predecir un atributo continuo; para predecir una secuencia; para buscar grupos de elementos comunes en las transacciones y para buscar grupos de elementos similares.
4. Análisis de los resultados. Estudio de los resultados obtenidos una vez aplicado el algoritmo a la base de datos limpia. (Ramírez et al., 2004)

El desarrollo científico y tecnológico ha llevado al hombre a buscar alternativas que sobrepasen su condición humana con el fin de solucionar problemas de áreas como la medicina, la investigación, la educación, la economía y lo social entre muchas otras áreas, siendo el computador una de sus primeras producciones hasta la invención de Analytical Engine by

Babbage y el primer programa de computadora que fue escrito por Ada Lovelace en 1842, momento en el que la personas comenzaron a pensar que podría haber un momento en que las computadoras o las máquinas fueran realmente inteligentes (Sharma, Bali, & Sarkar, 2017).

Estos desarrollos han propiciado cambios en los paradigmas sociales, que, influenciados por las tecnologías de la información, ofrecen nuevos roles y áreas de investigación y ocupación laboral que exigen nuevos términos o formas de concebir las metodologías de trabajo. La continua y exponencial cantidad de información que se genera a diario lleva a aplicar una serie de técnicas y estrategias que posibilitan su manejo y análisis; áreas como las matemáticas y la estadística, las ciencias de la computación y el conocimiento del dominio se han integrado para tal fin y se reconoce hoy como Ciencia de Datos (Kakade & Ozdemir, 2018).

Para obtener conocimientos de los datos, se debe poder utilizar la programación de computadoras para acceder a los datos, entender las matemáticas detrás de los modelos que se derivan y, sobre todo, entender el lugar de los análisis en el dominio en que se encuentran.

En la figura 2 se detalla la interacción y relación entre estas disciplinas de las cuales emergen nuevas áreas como el Machine Learning y la investigación tradicional.

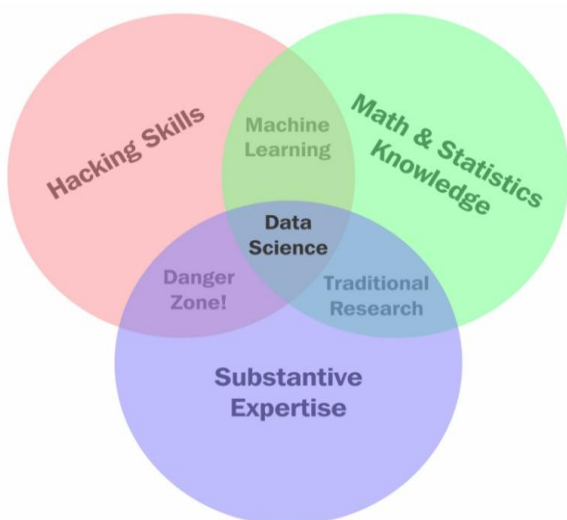


Figura 2. Diagrama de Ven Data Sciencie, (Kakade & Ozdemir, 2018)

Kakade y Ozdemir (2018) resaltan que la importancia de la ciencia de datos es la posibilidad de tomar decisiones acertadas cuando se ha desarrollado un proceso de extracción, transformación y análisis de la información; el cuestionamiento principal de la tesis de (Kakade & Ozdemir, 2018) es ¿es mejor crear un PDF de información o una aplicación en donde puede escribir números y obtener una predicción rápida? No obstante, pese a estar a favor de las formas algorítmicas para el estudio de datos, los autores aseguran que, aunque se tenga el mejor algoritmo para predecir el cáncer con una precisión de más del 99%, según datos de pacientes con cáncer, si no se entiende cómo aplicar este modelo en un sentido práctico para que los médicos y enfermeras puedan usarlo, su uso podría ser inútil.

Para Tamoghna et al. el Machine Learning (ML) es el desarrollo más avanzado de este sueño y su progreso es tal que es por sí mismo un área de investigación de la Inteligencia Artificial el crecimiento exponencial de información generada por los humanos, la naturaleza, el contexto social ha llevado al desarrollo de máquinas capaces de procesar estos volúmenes de información con el fin de poder analizarlos, predecir su comportamiento y definir sus usos; ya que es “un programa informático que aprende de la experiencia (**E**) con respecto a alguna tarea (**T**) y alguna medida de rendimiento (**P**), si su rendimiento en (**T**), medido por (**P**), mejora con la experiencia (**E**)”. (Tamoghna et al., 2018).

Se puede decir entonces que el Machine Learning es una disciplina del ámbito de la Inteligencia Artificial en donde se crean sistemas que aprenden automáticamente. Es decir, identifica patrones complejos en millones de datos a través de una máquina que aprende, que en este caso es un algoritmo encargado de revisar los datos y puede predecir comportamientos futuros de manera automática.

El Machine Learning se utiliza para el manejo de altos volúmenes de información que permiten detectar patrones de comportamiento como resultado de la aplicación del algoritmo, lo que no sería posible sin él. Permite que las empresas e instituciones pasen de un estado reactivo a un estado proactivo al crear una base de datos con toda la información pertinente por voluminosa que sea.

Es decir que el Machine Learning (ML) tiene como objetivo desarrollar técnicas que les permita a las computadoras “aprender” o sea crear programas capaces de generalizar comportamientos a partir de una información suministrada mediante una base de datos depurados con antelación. Cuando se tiene un alto volumen de datos es difícil que una sola persona analice, saque conclusiones y prediga, entonces es ahí donde aparecen los algoritmos que pueden detectar patrones de comportamiento, partiendo de las variables presentadas y descubriendo las que están perjudicando al sistema en estudio.

4.4.1. Fases del Proceso del Machine Learning

Pérez y González (2007) plantea que para aplicar el Machine Learning hay que tener en cuenta las fases que requiere su proceso:

Entender el Problema. Se debe tener claro el problema a resolver hay que preguntar una y otra vez ¿por qué? hasta quedar satisfecho pues es más fácil imaginar las respuestas a las preguntas que van surgiendo a medida que se trabaja el problema (Pérez y González, 2007).

Entender los datos. Es importante realizar un análisis exploratorio de datos para familiarizarse con ellos. Se puede acudir a gráficos, correlaciones y estadísticas descriptivas para entender la historia que están contando los datos y determinar si éstos son suficientes para construir un modelo.

Definir un Criterio de Evaluación. Con el fin de saber si el modelo de Machine Learning es el adecuado y de calidad se debe conocer con anticipación cómo se va a evaluar. Este criterio normalmente trata normalmente de una medida de error.

Evaluación de la Solución Actual. Obtener mejores resultados suele ser la mejor motivación para utilizar Machine Learning y la de reemplazar un trabajo manual aburrido por un trabajo automatizado. Evaluar la solución actual, si la hay, con el rendimiento del modelo de ML determina si vale la pena usarlo.

Preparar los Datos. Es una fase que supone un gran esfuerzo, ya que hay que enfrentar situaciones diversas con ellos:

Se puede presentar que los datos estén incompletos, para lo cual existen algunas soluciones favorables o menos favorables frente a las cuales el profesional debe decidir. Puede eliminarlos corriendo el riesgo de no tener suficientes datos; imputarlos con un valor razonable, que tenga sentido; imputarlos con un modelo de machine learning que permita definir cuál es el valor que hace falta y; no hacer nada y usar alguna técnica de Machine Learning que maneje datos incompletos.

Combinar datos de varias fuentes. Normalmente hay que combinar datos de fuentes diferentes como de una base de datos, de una hoja de cálculo, de ficheros, ya que es importante que Machine Learning pueda considerar toda la información.

Darle el formato adecuado a los datos. Cuando se van a utilizar las librerías de Machine Learning hay que dar formato a los datos. Normalmente se espera que los datos tengan forma de matriz o de tensor (generalización de una matriz).

Calcular características relevantes. Cuando se ofrece al Machine Learning características relevantes de los datos en lugar de presentarle datos puros. Ejemplo: es mejor conocer la temperatura en grados Celsius que saber cuántos miligramos se ha dilatado el mercurio.

Normalización de datos. Normalizar implica poner los datos en una escala similar.

Construir el modelo. Hay que elegir qué tipo de técnica de Machine Learning se va a usar. El algoritmo aprende automáticamente a obtener los resultados adecuados con los datos que se han manejado. Siempre habrá un error.

Análisis de Errores. Es importante analizarlos para entender qué hacer para mejorar los resultados. Las opciones que se pueden tener en cuenta para este caso son: usar un modelo más complejo; usar un modelo más simple; definir si se requieren más datos o más características; desarrollar una mejor comprensión del problema.

Modelo integrado de un sistema. Cuando se comprueba que el error obtenido es mejor que el error de la solución actual se debe integrar el modelo de ML al sistema ya que se necesita repetir de forma automática la preparación de datos; que el modelo se comunique con otras partes del sistema; que los resultados del modelo se usen en el sistema; monitorear automáticamente los errores del modelo; revisar si los errores del modelo crecen con el tiempo. (Pérez & González, 2007).

Es importante aclarar que a pesar de las bondades del Machine Learning, existen algunas dificultades para su uso como: los altos costos generados por el almacenamiento de los grandes volúmenes de datos en equipos servidores; en algunas ocasiones las matemáticas no pueden representar lo que se quiere hacer pues los algoritmos son bastante complejos; la comprensión de los resultados no es tarea fácil, la interpretación o decodificación de los resultados no logra descifrar adecuadamente la respuesta.

Las bondades del Machine Learning permiten su aplicación a contextos educativos para identificar y potenciar procesos de aprendizaje-enseñanza, como identificar en grandes poblaciones dificultades que se presentan al aprender un tema, o para identificar las características de ciertas fortalezas cuando se enseña o se aprende. Es de anotar que el Machine Learning tienen en su haber diferentes técnicas, metodologías y algoritmos sofisticados que han ido evolucionado con el tiempo y que deben ser detalladas para entendimiento del documento, según (Géron, 2019).

El Machine Learning requiere para su aplicación, como se ha dicho, de algoritmos que permitan obtener respuestas. A continuación, se describen los algoritmos principales que se tendrán a consideración para el desarrollo de esta investigación:

Tabla 3.
Tipos de Algoritmos según Categorías

Categoría	Descripción	Objetivo	Algoritmos
Clasificación	Identificar a qué categoría pertenece un objeto	Detección de spam, reconocimiento de imágenes.	SVM, nearest neighbors, random forest, gradient boosting tree
Regresión	Predecir un atributo de valor continuo asociado a un objeto.	Respuesta a medicamentos, precios de las acciones	SVR, ridge regression, Lasso
Agrupamiento	Agrupación automática de objetos similares en conjuntos.	Segmentación de clientes, agrupación de resultados del experimento.	k-Means, spectral clustering, mean-shift
Reducción de dimensiones	Reduciendo el número de variables aleatorias a considerar.	Visualización, Mayor eficiencia.	PCA, feature selection, non-negative matrix factorization.

Selección de modelos	Comparar, validar y elegir parámetros y modelos.	Precisión mejorada a través del ajuste de parámetros	Grid search, cross validation, metrics.
Procesamiento	Extracción y normalización de características.	Transformación de datos de entrada, como texto para usar con algoritmos de machine learning.	Preprocessing, feature extraction.

Cada algoritmo implica el uso de herramientas específicas. La tabla 4 recoge las que se van a considerar en este trabajo.

Tabla 4.
Herramientas y Librerías del Algoritmo

Herramientas y Librerías	Descripción
Python	Lenguaje de programación orientado a objetos, popular por su potencia y versatilidad en diferentes campos, como la investigación, señales, automatización, desarrollo convencional entre otros.
Jupyter Lab	Entorno integrado de desarrollo para investigación.
Pycharm	Entorno integrado de desarrollo para Python.
Sklearn	Librería Python para el trabajo con Machine Learning, integra múltiples utilidades algorítmicas sofisticadas para el trabajo con datos e inteligencia artificial.
PyMc3	Librería de redes Bayesianas, cadenas de Markov y algoritmos Montecarlo.
Pandas	Librería Python para trabajar con conjuntos de datos, estos pueden ser cargados tipos de datos llamados datasets y permiten hacer todo tipo de operaciones con estos: limpiar, transformar, calcular, cruzar etc.
Numpy	Librería de Python necesaria para trabajar con cálculos matemáticos de alto nivel.

Scipy	Librería de Python necesaria para trabajar con cálculos matemáticos de alto nivel, incluye módulos para estadística, interpolaciones, señales etc.
Django	Marco de trabajo para el desarrollo de aplicación web basadas en Python.

La variedad de algoritmo necesita para su aplicación de uno o varios tipos de aprendizajes, como:

Aprendizaje supervisado. Este tipo de aprendizaje involucra alta participación humana en términos de monitoreo, selección de nuevos atributos, entrenamiento y nuevas puestas en producción. El aprendizaje supervisado utiliza un set de datos que es dividido en 2 partes: una de entrenamiento, para generalizar el conocimiento que se espera pueda predecir el modelo resultante y los datos restantes se toman como test y se usan para verificar y contrastar que tan bien aprendió a generalizar el conocimiento el modelo final.

Aprendizaje no supervisado. Esta clase de algoritmos busca identificar estructuras en los datos. No se tiene la respuesta conocida para cada caso por lo que el algoritmo debe encontrar las relaciones entre las variables involucradas. No se busca la representación de los datos.

Aprendizaje semi supervisado. Estos algoritmos son un híbrido de aprendizaje supervisado y no supervisado. En este caso, el algoritmo trabaja con pequeñas cantidades de datos de entrenamiento etiquetados y más de datos sin etiquetar. De este modo, se hace uso creativo de los métodos supervisados y no supervisados para resolver una tarea determinada.

Aprendizaje por refuerzo. El Algoritmo recibe algún tipo de valoración acerca de la idoneidad de la respuesta dada. Ejemplo: darle una orden al perro de sentarse, si lo hace se le premia y si no lo hace se le castiga.

Aprendizaje en lotes. También se denomina como aprendizaje fuera de línea. Este tipo de aprendizaje es utilizado cuando se tiene un conjunto de datos de entrada y se quiere correlacionar con un conjunto de datos de salida, es decir encontrar una conexión entre estos conjuntos, que en la mayoría de los casos son conjuntos de datos diferentes por lo que deben ser normalizados para obtener el resultado. Ejemplo: la correlación que existe entre la estatura y el peso de una persona.

Aprendizaje en línea: En este caso el aprendizaje no se detiene una vez que los datos están disponibles, sino que los datos se introducen en el sistema en mini-lotes y el proceso de aprendizaje continúa con nuevos lotes de datos.

4.4.2. Ámbitos de Aplicación del Machine Learning.

En la actualidad los ámbitos que están aplicando el Machine Learning son:

Compras online. El online advertising que orienta en dónde poner un anuncio para que tenga mayor visibilidad en función del usuario de la web.

Detección de fraudes en transacciones. Al monitorear las características de los pagos como cantidades en cada transacción y rapidez con que se hacen se diseñan algoritmos que determinan las características y patrones genéricos que solo aparecen en este tipo de fraudes.

Predicción de fallos en equipos tecnológicos. Mediante soportes tecnológicos como Oracle Data Miner y metodológicos como CRISP – DM y entornos y lenguajes de programación R y Python.

Decisiones del sector de Recursos Humanos. La inteligencia artificial ha transformado los procesos de reclutamiento de personal describiendo los puestos de trabajo de manera neutral en cuanto al género y prediciendo recomendaciones de formación que los empleados puedan recibir.

Selección de clientes potenciales de acuerdo a sus comportamientos en las redes sociales.

Predicción del tráfico urbano. Es el caso presentado por el grupo Engine-4 para reducir el tráfico en áreas urbanas mediante el uso de cámaras e inteligencia artificial.

Determinar el mejor momento para publicar tuits, actualizaciones en Facebook o enviar las Newsletter.

Realizar pre diagnósticos médicos a partir de los síntomas del paciente.

Cambiar el comportamiento de una app móvil para adaptarla a las costumbres y necesidades del usuario.

Determinar la mejor hora para llamar a un cliente. (CleverData, 2014).

La evolución de la ciencia y la tecnología trae consigo la modificación de estructuras en todas las áreas de la sociedad siendo este uno de los casos, en donde los procesos propios de la investigación encuentran en su haber nuevas herramientas que permiten desarrollar tareas de tipo investigativo mucho más eficientes y con mayor alcance en cuanto al análisis y producción se refiere.

4.4.3. Simulación educación

Es recurrente el uso de las simulaciones a nivel de datos en procesos de investigación y desarrollo ya que han permitido la experimentación a bajo costes y recrear situaciones y contextos posibles definiendo sus condiciones, caracterizar y perfilar sus procesos, estrategias y

procedimientos con el objetivo de descubrir múltiples efectos que no pueden ser observados directamente en experimentaciones reales.

Bravo, Ortega, y Prieto (2009) consideran que:

La simulación es la imitación del modo de funcionamiento u operación de un sistema real que involucra la generación de una historia artificial del sistema, y la observación de esa historia artificial para obtener interfaces relativas a las características del funcionamiento del sistema real. (Bravo et al., 2009).

Sin embargo Cataldi, Dominighini, y Lage (2013), advierten que “la importancia de las simulaciones, desde el punto de vista educativo, reside en hacer partícipe al usuario de una vivencia que es fundamental para el desarrollo de los hábitos, destrezas, esquemas mentales, etc., que pueden influir en su conducta” (Cataldi et al., 2013). No la simulación en sí; es decir, sin acción del aprendiz sobre los datos y resultados, la simulación no tiene valor funcional en el mundo real.

Bravo et al. puntualizan que la simulación puede tener varios objetivos: la predicción, el entrenamiento, la resolución, el aprendizaje, pero en todos los casos es fundamental el objetivo didáctico ya que siempre se aprende de los resultados.

Salas y Zuleta (1995) relacionan dos usos para la simulación: para adelantar procesos de enseñanza – aprendizaje y para la evaluación, para lo cual explican:

...Durante la enseñanza – aprendizaje, los diversos tipos de simulación disponibles pueden utilizarse no sólo para el mejoramiento de las técnicas de diagnóstico, tratamiento y resolución de problemas, sino también para mejorar las dificultades psicomotoras y de relaciones humanas, donde en ocasiones pueden ser más eficaces que muchos métodos tradicionales, todo lo cual está en dependencia

fundamentalmente de la fidelidad de la simulación. La simulación posibilita que los educandos se concentren en un determinado objetivo de enseñanza, permite la reproducción de un determinado procedimiento o técnica y posibilita que todos apliquen un criterio normalizado. (Salas & Zuleta, 1995).

Como se observa los diferentes aportes sobre la aplicación de la simulación en educación van dirigidos a los estudiantes en sus procesos de enseñanza – aprendizaje, pero en las últimas décadas se ha empleado la simulación en predicciones de comportamientos.

4.5. CRISP – DM

De lo anterior, se hace importante resaltar y describir a CRISP-DM: según Fuentes (2018) CRISP-DM es un marco y/o una metodología popular para el desarrollo de análisis predictivos, convirtiéndose en el estándar de todo lo concerniente con la minería de datos. Esta metodología se divide en 6 fases (Ver tabla 5) que para los autores (Hipp, 2000; Wirth, 2000) no tiene una secuencia en el desarrollo de las fases, aunque se sugiere un flujo como el de la figura 3. (Fuentes, 2018)

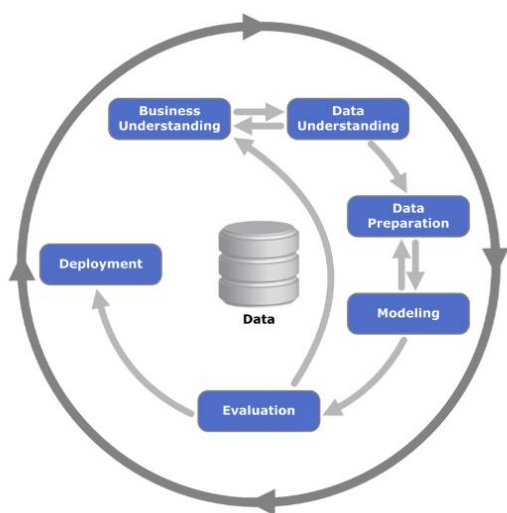


Figura 3. CRISP -DM WorkFlow (Jensen Own, 2016)

Otros autores como Tamoghna, Raghav, y Dipanjan (2018), refieren que la ciencia de datos, la minería de datos y el Machine Learning (ML) son procesos iterativos que buscan extraer información de los mismos datos haciendo de esto una ciencia y un arte que no solo se limita a ejecutar algoritmos sin ningún sentido, por el contrario, esto requiere de otra serie de actividades que dan valor a la información recolectada y por ello requiere un proceso formal que puede ser encontrado en CRISP-D (Tamoghna et al., 2018).

Por otra parte (Gutierrez, 2015; Thanaki, 2018) sugiere una metodología muy similar a la CRISP-DM salvo algunas variantes en las fases donde resalta el cambio de nombre de las mismas manteniendo las mismas actividades a desarrollar en cada una o la agregación de algunas de las actividades como fase, tal es el caso de (Thanaki, 2018).

Ingeniería de características, entrenamiento del modo de línea de base, comprensión de la matriz de pruebas y optimización del enfoque existente (Features engineering, Training the baseline mode, Understanding the testing matrix y optimize the existing approach) entre otros que en conclusión no difiere de lo propuesto por (Jensen, 2016) y lo mencionado por (Hipp, 2000; Wirth, 2000) con respecto a la no rigidez en la aplicación de la metodología CRISP-DM en razón de sus fases que pueden ser alteradas según sea el caso de uso y los objetivos que se pretenden.

Tabla 5.
Fases de la Metodología CRISP – DM

Fases	Objetivos	Actividades	Producto
Comprensión del Negocio	Perfilar de la mejor forma el problema recolectando presentaciones, entrevistas, conocimiento del contexto con el fin de	Determinar objetivos de negocios. Evaluar la situación.	Reporte de levantamiento de la información, definición de

	obtener detallada que permita definir los objetivos de la minería para la resolución del problema.	Determinar objetivos de minería de datos. Producir plan de proyecto.	objetivos y problema
Entendimiento de los datos	Familiarizar al científico con los datos, comprender su naturaleza, formas de recopilarlos, demostración de sus relaciones, encontrar los mejores recursos que nos permitan trabajar con ellos.	Recopilar datos iniciales. Describir datos Explorar datos. Verificar la calidad de los datos.	Reporte del estado de los datos, recolección de estos, particularidades y generalidades.
Preparación de los datos	Construir del conjunto de datos final a partir de los datos en bruto.	Seleccionar datos. Datos limpios. Construir datos. Integrar los datos. Formato de datos.	Conjunto de datos final para el entrenamiento del algoritmo seleccionado.
Modelamiento	Implementar el algoritmo (s) adecuado que permita dar respuesta al problema planteado que luego fue traducido a un problema de minería de datos, este depende del trabajo realizado en las fases anteriores.	Seleccionar la técnica de modelado. Generar diseño de prueba. Modelo de construcción. Evaluar el modelo.	Modelo de Machine Learning
Evaluación	Reportar en el lenguaje del contexto estudiado lo encontrado en el desarrollo del modelo, la	Evaluar resultados. Proceso de revisión.	Reporte del proceso desarrollado: calidad del modelo, calidad de los datos,

	<p>precisión en término de las medidas dispuestas en el caso estudiado describiendo el proceso de cómo se llegó a esto.</p>	<p>Determinar los siguientes pasos.</p>	<p>futuras iteraciones etc.</p>
Desarrollo	<p>Desplegar el producto final en términos del objetivo de la investigación, estos despliegues pueden ser tan simples como entregar un reporte o se pueden convertir en una fase realmente compleja donde la intervención de un equipo de Ingeniería se hace necesario.</p>	<p>Despliegue del plan.</p> <p>Plan de seguimiento y mantenimiento.</p> <p>Producir informe final.</p> <p>Revisar proyecto.</p>	<p>Software en ambiente WEB que permita usar el modelo en un contexto real y productivo.</p>

Fuente: fases de la metodología CRISP-DM de Hipp y Wirth, (2000)

5. Metodología

La metodología de la investigación es de tipo correlacional en la medida en que busca determinar si existe relación entre las variables académicas y sociodemográficas con la predicción del desempeño académico de los estudiantes. Para hallar tal relación se usan técnicas de machine learning.

Tabla 6.

Diseño de la investigación

Variable independiente	Variable dependiente	
Uso de técnicas de machine learning	Predicción del desempeño académico	
Factores académicos y sociodemográficos	Desempeño bajo	Desempeño alto

A continuación, se describe el procedimiento empleado en el desarrollo de la investigación, el cual está integrado por tipo de investigación, población y muestra, técnicas empleadas y etapas e instrumentos utilizados y descripción del ambiente computacional.

En la investigación se usó como enfoque metodológico el modelo CRISP-DM que significa Cross Industry Standard Process for Data Mining. El modelo CRISP-DM actúa como marco de trabajo propio del desarrollo de proyectos que abordan investigaciones basadas en minería de datos y técnicas de machine learning. (Fuentes, 2018).

Los reportes que resultan del trabajo desarrollado con esta metodología se enmarcan en los procesos conceptuales, algorítmicos y estadísticos descritos por James, Witten, Hastie, y Tibshirani (2013) en su libro *An Introduction to Statistical Learning with Applications in R*. y en aplicaciones desarrolladas en contextos educativos cuyos resultados señalan que el uso de

técnicas de Machine Learning favorece la optimización de procesos propios del campo educativo (James et al., 2013; Lawson, 2015; Castellanos, 2017).

Antes de aplicar la técnica de Machine Learning, los estudiantes resolvieron dos pruebas: una prueba simulacro y una encuesta de factores sociodemográficos.

5.1. Prueba Simulacro

Los estudiantes objeto de estudio respondieron una prueba simulacro tomada del modelo de pruebas que aplicó la Universidad Nacional de Colombia en el proceso de admisión para el segundo semestre del año 2010. La prueba estaba conformada por 15 preguntas y se aplicó a los 153 estudiantes de la Institución Educativa Departamental General Carlos Albán del municipio de Albán – Cundinamarca.

5.2. Encuesta sobre Factores Sociodemográficos

Los estudiantes de grado octavo, noveno, décimo y once del año 2019 respondieron una encuesta cuyo propósito fue el de conocer las condiciones sociodemográficas de los estudiantes. La prueba estuvo conformada por 33 grupos de preguntas relacionadas con datos personales, estudios realizados; datos de estudio, datos laborales, salario de los padres; datos de vivienda y comodidades de vivienda (Ver anexo 3).

5.2.1. Población.

La población está conformada por 599 estudiantes de los sectores urbano y rural que hacen parte del Institución Educativa Departamental General Carlos Albán del municipio de Albán – Cundinamarca. El colegio cuenta con tres sedes en el perímetro urbano; la sede principal

en donde funciona el bachillerato con 299 estudiantes y la parte administrativa; la sede de básica primaria Policarpa Salavarrieta cuenta con 199 estudiantes y la sede del Jardín Infantil que atiende 31 niños.

En el sector rural tiene tres sedes: La María con 35 estudiantes; y dos escuelas unitarias Java con 28 estudiantes y Los Alpes con 7 niños.

5.2.2. Muestra.

En el estudio participaron 153 estudiantes de los grados octavo, noveno, décimo y once con edades que oscilan entre los 12 y 18 años. El estudio se realizó en los espacios académicos de la asignatura de Tecnología. Los estudiantes objeto de la muestra pertenecen a cuatro grados: octavo con 26 estudiantes (10 mujeres y 16 hombres), noveno con 49 estudiantes (25 mujeres y 24 hombres), décimo con 45 estudiantes (19 mujeres y 26 hombres) y once con 33 estudiantes (21 mujeres y 12 hombres).

5.2.3. Variable Dependiente

La investigación contiene una variable dependiente: La predicción del desempeño académico en el desarrollo del espacio proyectivo del pensamiento espacial.

5.2.4. Variable Independiente.

La investigación incluye una variable independiente: Aplicación de técnicas de machine learning con múltiples valores que se derivan de: resultados pruebas universidad nacional, competencias socioemocionales, factores familiares, económicos y sociales entre otras. Los

datos que resulten alimentan en parte la base de datos que se usó en el entrenamiento del modelo predictivo.

5.3. Etapas de la investigación

La investigación se desarrolló en tres etapas: 1-) adaptación de la Prueba de Admisión de la Universidad Nacional de Colombia, 2-) diseño de encuesta simulacro conformada por variables sociodemográficas y 3-) diseño del modelo de: procesamiento, sanidad de datos y desarrollo del modelo predictivo con técnicas de Machine Learning.

5.3.1. Primera etapa

Se optó por analizar la Prueba de Admisión Universidad Nacional de Colombia (Ver anexo 1) para estudiar la posibilidad de aplicarla a los estudiantes que hacen parte de la muestra con el fin de detectar con mayor precisión las dificultades que se presenta en el desarrollo de pensamiento espacial de los estudiantes de noveno, décimo y once de 2019. La prueba consta de 15 preguntas de razonamiento espacial acordes con las características del desarrollo de relaciones proyectivas del pensamiento espacial planteadas por (Piaget,1947). Los resultados de la prueba se usarán como una de tantas variables de insumo que alimentan la base de datos para el entrenamiento del modelo predictivo y que luego del análisis permitirán establecer procesos pedagógicos que favorezcan el desarrollo del pensamiento espacial.

5.3.2. Segunda etapa

Se diseñó y elaboró una encuesta sobre factores sociodemográficos a partir del diccionario de variables propuesto por el (MEN, 2006; ICFES, 2016) (Ver anexo 2). La encuesta

tiene el propósito de caracterizar a cada estudiante y los datos que resulten se usan para alimentar la base de datos para el entrenamiento del modelo predictivo.

5.3.3. Tercera etapa

En esta tercera etapa se aplica la metodología CRISP-DM como marco de trabajo para el desarrollo del modelo predictivo que hace parte de esta investigación y que busca reportar sus resultados de manera formal a través de las métricas que contempla el modelo.

La metodología CRISP-DM hace referencia a un diseño metodológico abordado en esta investigación, denominada por los autores Dicovskiy y Pedroza (2018) como una metodología innovadora que se encuentra dentro de los métodos tradicionales de investigación cuantitativa. (Dicovskiy y Pedroza,2018)

En esta tercera etapa se abordan temas como minería de datos, machine learning y metodología crisp-dm; la minería de datos o exploración de datos busca descubrir información en contextos que son objetos de estudio e investigación y donde existe intraversión humana para alcanzar dichos descubrimientos, por su parte el machine learning como disciplina de las ciencias de la computación y rama de la inteligencia artificial hace uso de las bases de datos procesadas con minería de datos implementando algoritmos computacionales que permiten encontrar patrones, probabilidades y conclusiones sobre un conjunto de datos que caracterizan una población o contexto y que además no requiere de intervención humana ni procesamiento de códigos computacionales para su funcionamiento ya que con anterioridad este debe ser entrenado con la información recolectada para que logre generalizar el conocimiento de un contexto, esto equivale a tener información de animales para luego enseñársela a un niño y que este tenga la capacidad de clasificarlos y eventualmente si desconoce alguno pueda inferir de qué tipo de

animal podría tratarse ; normalmente al resultado del proceso anterior se le conoce como modelo predictivo que no es más que un archivo que recibe una información de entrada para luego dar una predicción.

Para lo anterior es importante resaltar que la implementación y desarrollo de este tipo de modelos requiere la aplicación de una metodología que permita garantizar la rigurosidad del proceso, la ejecución, la evaluación y reporte de métricas que garanticen la calidad del modelo predictivo resultante del contexto que se investiga o estudia, por esto encontraremos en crispdm una metodología formal para el desarrollo de la investigación.

Las fases de la metodología que se desarrollará en esta investigación para obtener las métricas de reporte en mención son:

Comprensión del contexto: esta fase consiste en perfilar de la mejor forma el problema mediante la recolección de presentaciones, entrevistas y conocimiento del contexto con el fin de obtener detallada información que permita definir los objetivos de la minería para la resolución del problema.

Interpretación de los datos: se centra en familiarizar al científico con los datos para comprender su naturaleza, las formas de recopilarlos, la demostración de sus relaciones y para que encuentre los mejores recursos que le permitan trabajar con ellos.

Preparación de los datos: esta fase se basa en construir el conjunto de datos final a partir de la minería de los datos en bruto.

Modelación: fase que consiste en implementar el algoritmo (s) adecuado que permita dar respuesta al problema planteado y luego traducido a un problema de minería de datos. El éxito de esta etapa depende del trabajo realizado en las fases anteriores.

Evaluación: se centra en reportar los hallazgos del modelo a través del lenguaje que se utiliza en el contexto estudiado. El reporte muestra la precisión en término de las medidas dispuestas en el caso estudiado y describe el proceso de cómo se concluyó.

Despliegue: fase que consiste en desplegar el producto final en términos del objetivo de la investigación, tal despliegue puede ser tan simple como entregar un reporte o tan complejo que se puede requerir la intervención de un equipo de Ingeniería.

5.4. Diseño de ambiente computacional

Durante la investigación se utilizó un modelo computacional diseñado por los investigadores para evaluar un conjunto de variables académicas y sociodemográficas que afectan el desempeño de los estudiantes en el desarrollo del pensamiento espacial. El ambiente contiene básicamente el modelo predictivo.

El modelo predictivo se diseñó con base en la minería de datos y con el uso de la metodología CRISP-DM. Está integrado por una base de datos, un modelo clasificador, un conjunto de algoritmos computacionales, unas técnicas de Machine Learning y simulaciones.

La minería de datos tiene que ver con toda la información disponible en un dominio de conocimiento específico que se traduce en un conjunto de variables que caracterizan el contexto objeto de estudio y que surten efectos significativos en el comportamiento de un sujeto o de un fenómeno. Con respecto a la metodología CRISP-DM se definió líneas arriba.

La base de datos está conformada por el conjunto de variables académicas y sociodemográficas que caracterizan a los estudiantes del colegio; el modelo clasificador lo integran el conjunto de algoritmos que evalúa las variables de la base de datos y que tiene el propósito de clasificarlas en orden de prioridades de acuerdo con los efectos que surten en el desempeño de los estudiantes. Los algoritmos son programas computacionales diseñados con técnicas de inteligencia artificial que realizan un barrido de la base de datos, evalúan las variables, seleccionan aquellas que afectan el desempeño de los estudiantes y rechazan las que no afectan tal desempeño y las técnicas de Machine Learning se definieron a anteriormente.

Simulación on-line. Esta simulación puede ser integrada en cualquier OVA o AVA para desarrollar otras investigaciones o continuar la que presenta este trabajo. Para el caso en estudio se desarrolló en Python el simulador al que se le integró el modelo de Machine Learning presentado. El simulador recibe un estudiante y devuelve la predicción de manera automática, el diseño de la solución se presenta en la Figura 8 donde se ven tres componentes principales para su funcionamiento.

Aplicación. Se conforma de un Back que es la lógica que corre del lado del servidor y que permite las integraciones, este se desarrolló sobre Django que es un framework para construir aplicaciones web bajo Python; el componente de front es la parte visual que fue desarrollada en HTML, CSS y JS, y permite interactuar con el simulador.

5.5. Modelo de Machine Learning.

Desarrollado con Python que hizo uso de cuatro paquetes principales como: la librería sklearn que agiliza el desarrollo de modelos debido a que tienen integrados algoritmos supervisados y no supervisados que solucionan problemas de Machine Learning; Pandas y generación de estos; PyMC3 librería desarrollada para el campo de la investigación, presenta un

set de algoritmos que trabajan con bayesianos, Montecarlo y cadenas de Markov y agilizan procesos de simulación y/o desarrollo de modelos que requieran estas lógicas y, por último Numpy como librería para cálculos matemáticos de alta precisión y trabajo con matrices, vectores etc.

Base de datos que guardan las simulaciones, reportes y configuraciones del simulador.

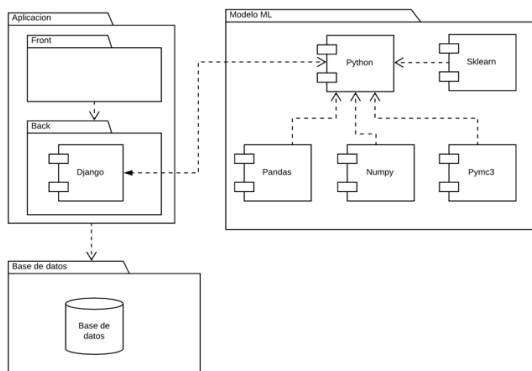


Figura 4. Diagrama de Componentes del Simulador Web

Además, se generó un esquema conceptual de cómo funciona toda la solución para cumplir con el objetivo de este trabajo que se ve en la Figura 5. El esquema muestra un flujo de bloques que empieza de arriba hacia abajo y caracteriza el Simulador (la probabilidad de desempeño de un estudiante categorizados en desempeño alto o desempeño bajo en el desarrollo del espacio proyectivo del pensamiento espacial con cuatro componentes:

- Problemas educativos usando las TIC como herramienta. Componente tecnológico: este contiene la solución tecnológica final que aporta valor como herramienta al desarrollo de actividades pedagógicas con el uso de Machine Learning, Simulación, Minería de Datos y programación; es importante resaltar que es dependiente de los componentes de investigación, disciplinar y pedagógico para su funcionamiento.
- Componente de Investigación: desarrollado bajo metodología CRISP permite engranar los componentes disciplinares, pedagógicos y tecnológicos a fin de generar un proceso

organizado e integrador de cada una de las disciplinas que lo componen y que en conjunto darán respuesta al objetivo y a la pregunta planteada para este trabajo.

- **Componente disciplinar:** en este interviene el conocimiento del experto en la materia, este componente es clave para la definición y caracterización de la intervención tecnología y la aplicación de la metodología.
- **Componentes pedagógicos:** se convierte en la base de todo el proyecto y sobre este se inicia la definición, construcción e integración de lo necesario a nivel metodológico, conceptual y procedimental; este componente define el sentido que el Machine Learning, la simulación y los demás componentes tecnológicos tienen dentro de un contexto enfocado a dar solución.

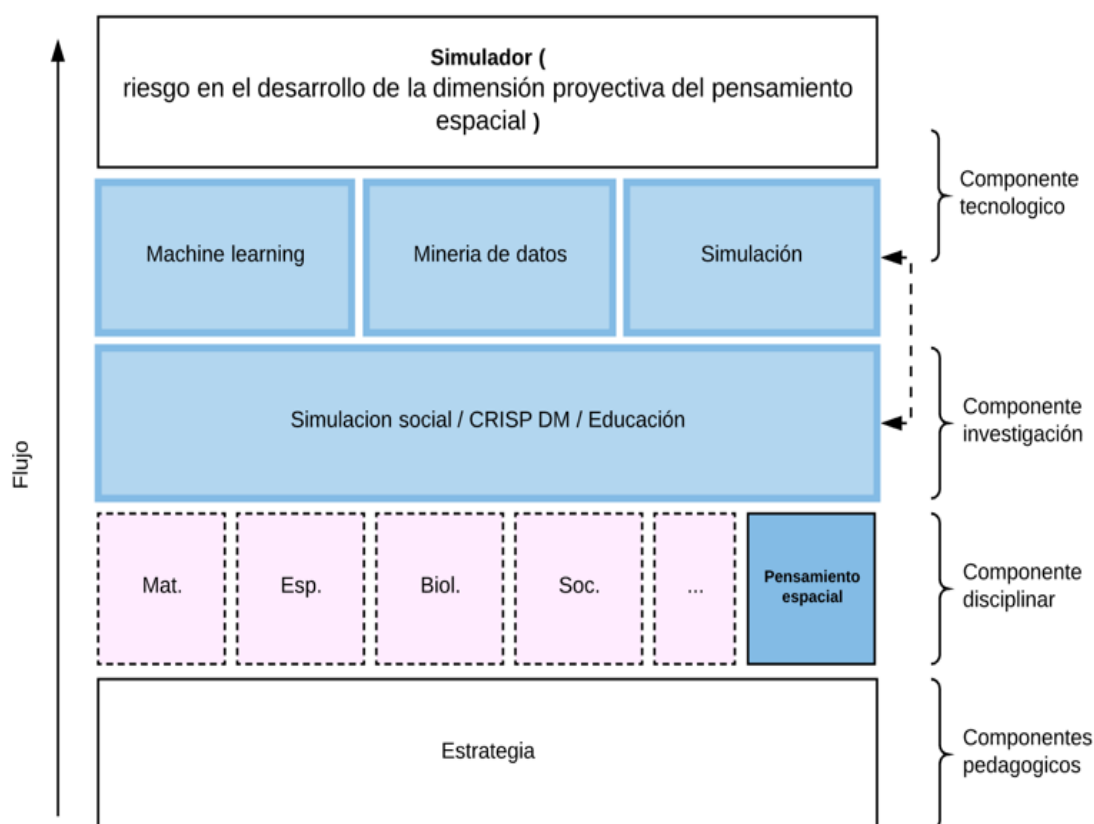


Figura 5. Esquema Conceptual de la Solución.

5.5.1. Simulador Web

Las características del simulador web que integra el modelo de Machine Learning.

Figura 6. Formulario de Ingreso de Datos

La captura de pantalla muestra el formulario base de ingreso de datos, que son procesados por el sistema para ser enviados al modelo de Machine Learning encargado de la predicción en el momento de que el estudiante se matriculado o registrado en la institución.

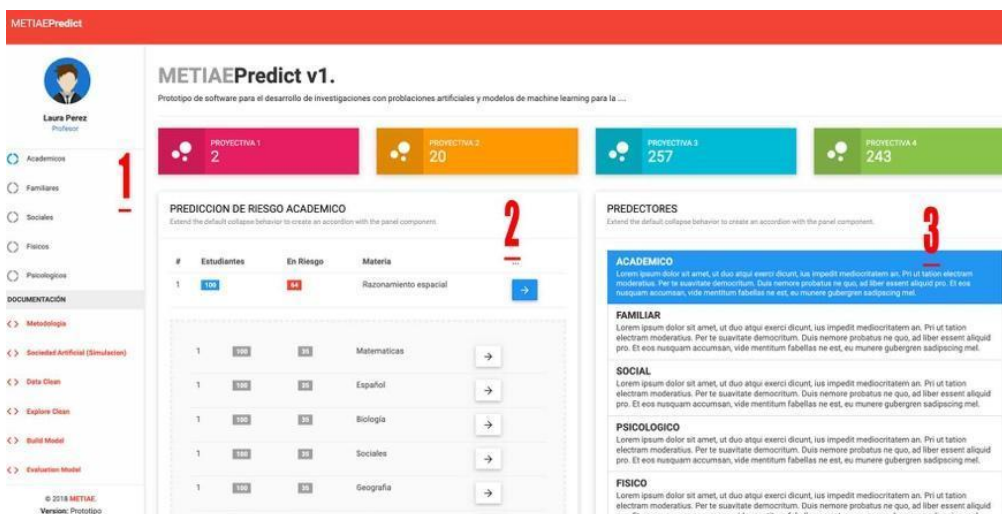


Figura 7. Resumen de Estudiantes Clasificados.

En la Figura 7 muestra un resumen de la clasificación de los estudiantes. El capturador de datos en su sección 1 permite agregar más predictores en futuras investigaciones, además en esta sección se encuentran vinculados los notebook de investigación que se usaron para el desarrollo del modelo de Machine Learning y la simulación. También muestra en la sección 2 un listado de categorías por materia o área disciplinar con sus respectivas predicciones según sea el caso, pero es de anotar que para este trabajo solo está habilitado el del de desarrollo del espacio proyectivo. Al dar click en esta sección irá al listado con el detalle de clasificación de los estudiantes y en la sección 3 da una explicación de uso y posibles predictores que se podrían agregar.

The screenshot displays the METIAEPredict web application interface. At the top, there is a navigation bar with three colored buttons: a green button labeled '0-60 REPRUEBA', an orange button labeled '61-80 CON DIFICULTAD', and a pink button labeled '81-100 APRUEBA'. Below the navigation bar, the interface is divided into three main sections, each marked with a red number: 1, 2, and 3.

Section 1 (left) is titled 'DETALLE, PREDICCIÓN Y RUTAS SUGERIDAS' and shows the student's profile for 'MARIA PEREZ'. It includes a table of features:

FEATURE	Value
Nombre	Maria Perez
Edad	24
Ciudad	Bogota
Curso	11

Section 2 (middle) is titled 'MARIA PEREZ' and shows a table of subject recommendations:

Area	Segmento	Recomendaciones
Razonamiento espacial	EN RIESGO	[+]
Matematicas	EN RIESGO	[+]
Español	EN RIESGO	[+]
Biología	EN RIESGO	[+]
Sociales	EN RIESGO	[+]
Inglés	EN RIESGO	[+]
Geografía	EN RIESGO	[+]

Section 3 (right) is titled 'RUTAS SUGERIDAS' and lists four projective routes (PROYECTIVA 1 to 4) with placeholder text.

Figura 8. Detalle del estudiante.

Esta sección permite ver en detalle las características de clasificación del estudiante, además de poder ver un historial de clasificación si existieran más predictores en el ambiente. Este pantallazo permite observar en la sección 1 la información usada por el modelo para clasificar; en la sección 2 el resultado de la clasificación y en la sección 3 las recomendaciones para que el docente las use en el proceso de seguimiento.

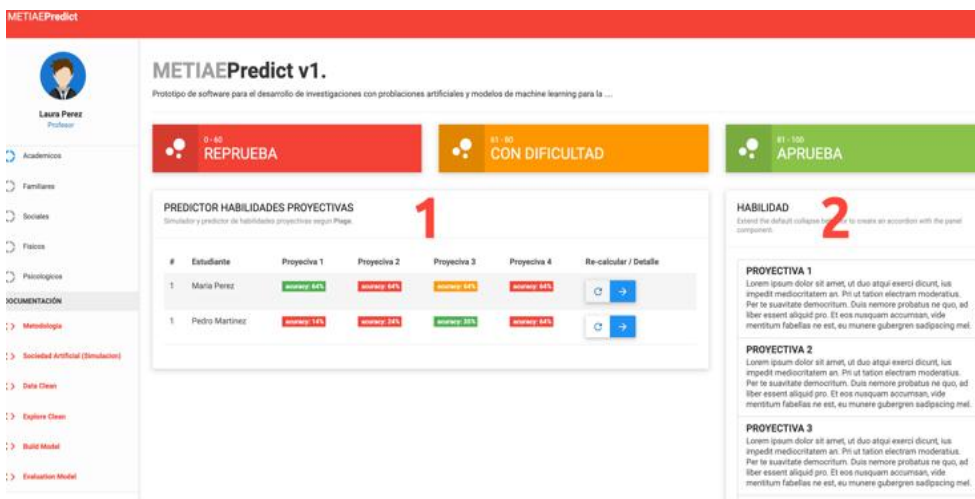


Figura 9. Listado de Estudiantes.

La figura anterior muestra en la sección 1 el listado de estudiantes clasificados, entrega el nombre del estudiante, la predicción del modelo y los valores de probabilidad devueltos por el modelo en el momento de la clasificación; la sección 2 muestra información de lo que está evaluando como guía para el usuario docente.

Lo anterior concluye el desarrollo de los principales pasos propuestos por la metodología CRISP-DM, evidenciando la importancia del uso de técnicas de Machine Learning para la solución de problemas educativos que para el desarrollo de este trabajo de investigación abordó la predicción del desempeño (desempeño alto o bajo) de un estudiante en el desarrollo del espacio proyectivo del pensamiento espacial, lo que se configura no sólo como un reporte de investigación sino que se materializa en una herramienta que promueva y fortalezca los procesos educativos e investigativos además de dar nuevas perspectivas o ramas de investigación e innovación para el campo de las tecnologías de la información aplicadas a la educación.

Se espera que el modelo reporte la predicción del desempeño académico en el desarrollo del espacio proyectivo del pensamiento espacial de los estudiantes a partir de las siguientes métricas: 1-) los mejores y peores puntajes de la muestra objeto de estudio, 2-) la clasificación de

los sujetos por puntaje desde el más alto hasta el más bajo, 3-) el ponderado de puntajes, 4-) la relación entre el número de predicciones correctas y el número total de muestras y 5-) la medida de rendimiento que indica el grado de separación entre un puntaje y otro y las distinciones entre clases.

La aplicación sistemática de una metodología como la CRISP-DM donde el Machine Learning se caracteriza en diferentes fases de la aplicación para dar respuesta a la pregunta de investigación, donde el desarrollo de herramientas permiten hacer productivos este tipo de modelos de Machine Learning siendo de vital importancia para el investigador pues éste se provee y provee al lector interesado en el área de novedosas y sofisticadas formas de trabajar con los datos y experimentar con los estudiantes otros procesos de aprendizaje – enseñanza sobre los que se quiera ahondar; por lo anterior se sugiere a los lectores apropiarse del material generado en este trabajo para aplicarlo a otras áreas disciplinares y compartir sus resultados para dar mayor sustento teórico a la aplicación del Machine Learning en la educación a fin de convertir las instituciones educativas en centros inteligentes de formación. Es de aclarar al lector que para hacer más eficiente el modelo que aquí se presenta o el desarrollo de otros, deberá recolectar y clasificar una mayor cantidad de datos.

6. Análisis de Resultados

Los resultados de la investigación se presentan en dos partes: primero se muestran los resultados que se obtuvieron a través de dos instrumentos: la prueba del simulacro y la encuesta sobre factores sociodemográficos. Estos resultados dan cuenta del desempeño académico en el desarrollo del espacio proyectivo del pensamiento espacial y de las características sociodemográficas de cada estudiante. Estos resultados hacen parte del insumo básico que alimenta la base de datos del modelo predictivo.

6.1. Prueba simulacro grado octavo

Los resultados de los 26 estudiantes de grado octavo señalan que 24 de ellos (88%) se ubicaron en Nivel Bajo y 2 estudiantes (12%) alcanzaron el nivel básico. Estos resultados indican que probablemente hay fallas en los procesos pedagógicos de formación o bien que se da poca importancia al desarrollo del espacio proyectivo de las relaciones espaciales.

Al revisar con más detalle los resultados se observa que el 9,8% de los estudiantes resolvió 2 de las tres preguntas del grupo pp1 que se relacionan con el desarrollo de la perspectiva de objetos diversos; el 4,9% de los estudiantes resolvieron 3 de las cinco preguntas del grupo pp2 que tienen que ver con el desarrollo de las relaciones proyectivas; en cuanto al grupo de preguntas pp3 que permitían determinar la rotación y desarrollo de superficies los resultados obtenidos no son mejores que los anteriores y frente a las preguntas que conformaban el grupo pp4 relacionado con el manejo de secciones geométricas y rotación y desarrollo de superficies de manera conjunta que manifestaban el paso de las operaciones concretas a las operaciones formales son significativas en cuanto a la deficiencia que presentada en este aspecto. (Ver figura 10).

	Nro.	p1-1	p1-2	p1-3	pp1	p2-1	p2-2	p2-3	p2-4	p2-5	pp2	p3-1	p3-2	p3-3	p3-4	pp3	p4-1	p4-2	p4-3	pp4	P_TOTAL	RESULTADO	RESULTA %								
800	1	1	0	0	1	1	1	0	1	0	3	1	1	1	0	3	1	0	0	1	8	BAJO	53,3								
	2	0	1	1	2	0	1	0	0	1	2	1	1	1	1	4	1	1	1	3	11	BASICO	73,3								
	3	0	1	0	1	0	1	1	0	1	3	1	1	1	1	4	1	1	1	3	11	BASICO	73,3								
	4	1	0	0	1	0	0	0	1	0	1	1	1	0	1	3	1	1	1	3	8	BAJO	53,3								
	5	0	0	0	0	1	0	0	1	0	2	1	1	0	0	2	0	1	1	2	6	BAJO	40,0								
	6	0	0	1	1	0	1	0	1	0	2	1	1	1	0	3	1	1	1	3	9	BAJO	60,0								
	7	0	0	1	1	0	0	0	1	0	1	1	1	1	0	3	1	0	1	2	7	BAJO	46,7								
	8	0	1	0	1	0	1	0	1	0	2	1	0	1	1	3	1	1	1	3	9	BAJO	60,0								
	9	0	1	0	1	0	0	1	0	0	1	0	0	1	1	2	1	1	1	3	7	BAJO	46,7								
	10	0	1	0	1	0	1	0	1	0	2	1	0	1	1	3	1	1	1	3	9	BAJO	60,0								
	11	0	0	1	1	0	0	0	1	0	1	1	1	1	0	3	1	0	1	2	7	BAJO	46,7								
	12	0	1	1	2	0	1	0	1	0	2	1	1	1	0	3	1	0	1	2	9	BAJO	60,0								
	13	0	0	1	1	0	0	0	1	0	1	1	1	1	1	4	0	1	0	1	7	BAJO	46,7								
	14	0	0	1	1	0	0	0	0	1	1	1	1	0	0	2	0	0	1	1	5	BAJO	33,3								
	15	0	0	1	1	1	0	0	0	1	2	1	1	0	0	2	1	1	1	3	8	BAJO	53,3								
	16	0	0	0	0	0	0	0	1	1	2	1	1	1	0	3	1	1	1	3	8	BAJO	53,3								
	17	0	1	1	2	0	0	0	1	0	1	0	1	0	0	1	0	1	1	2	6	BAJO	40,0								
	18	0	0	0	0	0	1	0	1	0	2	1	1	0	0	2	0	1	0	1	5	BAJO	33,3								
	19	0	0	1	1	0	1	0	0	0	1	1	1	0	0	2	0	1	1	2	6	BAJO	40,0								
	20	0	0	0	0	0	1	0	1	0	2	1	0	1	0	2	1	0	1	2	6	BAJO	40,0								
	21	0	0	1	1	0	0	0	0	0	0	1	1	0	0	2	0	0	1	1	4	BAJO	26,7								
	22	0	0	1	1	0	0	0	0	0	0	1	1	0	0	2	0	0	1	1	4	BAJO	26,7								
	23	0	1	1	2	0	0	0	0	1	1	1	1	0	0	2	0	0	1	1	6	BAJO	40,0								
	24	0	0	0	0	0	0	0	1	0	1	1	1	1	1	4	1	1	1	3	8	BAJO	53,3								
	25	0	0	0	0	0	0	0	0	0	0	1	1	1	0	3	1	0	1	2	5	BAJO	33,3								
	26	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	3	4	BAJO	26,7								
	24	18	13	55	23	16	24	11	20	94	2	5	11	18	36	9	10	3	22		53,1	46,9									
	78	% perdida 70,5										130	% perdida 72,3					104	% perdida 34,6					78	% perdida 28,2					% preguntas perdidas	% ACIERTOS
																						Preguntas aprobadas	207	BAJO	24						
																						Total de preguntas	390	BASICO	2						
																							ALTO	0							
																							SUPERIOR	0							

Figura 10. Resultados del Simulacro Grado Octavo - 2019.

6.2. Prueba simulacro grado noveno

Los resultados de los 49 estudiantes de grado noveno se distribuyen de la siguiente manera: 38 que corresponden al 76% se ubicaron en nivel bajo y los 11 restantes (24%) alcanzaron el nivel básico, al respecto, la Teoría de Piaget señala que los niños con edades comprendidas entre 12 y 14 años deben presentar un manejo adecuado de pensamiento concreto; según los resultados la mayoría de este grupo de niños no alcanzan un mínimo desarrollo del espacio proyectivo del pensamiento espacial (Ver figura 11).

Nro.	p1-1	p1-2	p1-3	pp1	p2-1	p2-2	p2-3	p2-4	p2-5	pp2	p3-1	p3-2	p3-3	p3-4	pp3	p4-1	p4-2	p4-3	pp4	P	TOTAL	RESULTADO	RESULTADO %	
1	1	1	1	3	1	1	1	0	1	4	0	1	0	0	1	0	0	0	0	8	8	BAJO	53.3	
2	1	0	0	1	1	1	0	0	0	2	0	1	0	0	1	2	0	0	0	0	5	BAJO	33.3	
3	1	0	0	2	1	1	0	0	0	3	0	0	0	0	1	1	0	0	0	0	6	BAJO	40.0	
4	1	1	1	3	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	6	BAJO	40.0	
5	1	1	1	2	1	1	1	0	1	3	0	1	0	0	1	2	0	0	0	0	7	BAJO	46.7	
6	1	1	1	1	1	1	1	1	1	5	0	1	0	0	0	1	0	0	0	1	8	BAJO	53.3	
7	1	1	1	2	1	1	1	1	1	4	0	1	1	1	1	3	0	1	1	2	11	BASICO	73.3	
8	1	1	1	3	1	1	1	1	1	4	0	0	0	0	0	1	1	0	0	1	9	BAJO	60.0	
9	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	1	1	0	0	1	9	BAJO	60.0	
10	1	1	1	3	0	0	0	0	0	2	0	1	0	0	0	1	1	0	0	0	6	BAJO	40.0	
11	1	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	4	BAJO	26.7	
12	1	1	1	3	0	0	1	0	0	3	0	1	1	1	0	2	0	1	0	1	9	BAJO	60.0	
13	1	1	1	3	0	0	1	0	0	4	0	1	1	0	0	0	0	0	0	1	9	BAJO	60.0	
14	1	0	0	2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	4	BAJO	26.7	
15	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	1	0	0	0	1	8	BAJO	53.3	
16	1	1	1	3	1	1	1	1	1	4	0	0	1	1	1	1	0	0	0	2	10	BASICO	66.7	
17	1	1	1	3	1	1	1	1	1	4	0	0	1	1	1	0	0	0	1	1	9	BASICO	66.7	
18	1	1	1	3	1	1	1	1	1	4	0	0	1	1	1	0	0	0	1	1	9	BAJO	60.0	
19	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	1	0	0	1	2	8	BAJO	53.3	
20	1	1	1	3	1	1	1	1	1	4	0	0	1	0	1	2	0	0	1	1	8	BAJO	53.3	
21	1	1	1	3	1	1	1	1	1	4	0	0	0	0	0	0	0	0	0	0	3	BAJO	20.0	
22	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	3	BAJO	20.0	
23	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	0	0	0	0	0	6	BAJO	40.0	
24	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	0	0	0	0	0	6	BAJO	40.0	
25	1	1	1	3	1	1	1	1	1	4	0	0	1	1	0	1	0	0	1	1	10	BASICO	66.7	
26	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	1	1	0	0	1	10	BASICO	66.7	
27	1	1	1	3	1	1	1	1	1	4	0	0	1	1	0	2	1	0	0	1	10	BASICO	66.7	
28	1	1	1	3	1	1	1	1	1	4	0	0	1	1	0	2	1	0	0	1	10	BASICO	66.7	
29	1	1	1	3	1	1	1	1	1	4	0	0	0	1	1	1	0	0	0	0	8	BAJO	53.3	
30	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	0	0	0	0	0	7	BAJO	46.7	
31	1	1	1	3	1	1	1	1	1	4	0	0	1	0	1	2	0	1	0	1	10	BASICO	66.7	
32	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	2	0	1	0	1	8	BAJO	53.3	
33	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	2	0	0	0	0	4	BAJO	26.7	
34	0	0	0	2	1	1	1	1	1	3	0	0	1	1	0	2	0	0	0	0	7	BAJO	46.7	
35	0	0	0	2	1	1	1	1	1	3	0	0	0	0	0	1	0	0	0	0	5	BAJO	33.3	
36	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	0	0	0	0	0	6	BAJO	40.0	
37	1	1	1	3	1	1	1	1	1	4	0	0	0	0	0	0	0	0	0	0	6	BAJO	40.0	
38	1	1	1	3	1	1	1	1	1	4	0	0	1	1	1	3	0	0	0	0	8	BAJO	53.3	
39	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	0	0	0	0	0	5	BAJO	33.3	
40	1	1	1	3	1	1	1	1	1	4	0	0	1	0	1	2	0	1	1	1	10	BASICO	66.7	
41	1	1	1	3	1	1	1	1	1	4	0	0	1	0	1	0	0	0	0	1	6	BAJO	40.0	
42	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	2	1	0	0	1	10	BASICO	66.7	
43	1	1	1	3	1	1	1	1	1	4	0	0	1	1	0	2	0	0	1	1	7	BAJO	46.7	
44	1	1	1	3	1	1	1	1	1	4	0	0	0	0	0	0	0	0	1	1	7	BAJO	46.7	
45	1	1	1	3	1	1	1	1	1	4	0	0	1	0	1	2	0	1	0	1	7	BAJO	46.7	
46	1	1	1	3	1	1	1	1	1	4	0	0	1	1	1	3	0	0	0	0	10	BASICO	66.7	
47	0	0	0	2	1	1	1	1	1	3	0	0	0	0	0	1	1	0	0	0	6	BAJO	40.0	
48	1	1	1	3	1	1	1	1	1	4	0	0	0	0	0	0	0	1	0	1	3	BAJO	20.0	
49	1	1	1	3	1	1	1	1	1	4	0	0	1	0	0	1	0	0	0	0	0	BAJO	53.3	
	147	% perdida	16.3	245						% perdida	43.7	196			% perdida	65.3	147	% perdida	78.9			51.0	49.0	
			1 Cat.							2 Cat.					3 Cat.					4 Cat.		% preguntas perdidas	% ACIERTOS	
																					Preguntas aprobadas	375	BAJO	38
																					Total de preguntas	735	BASICO	11
																						ALTO	0	
																						SUPERIOR	0	

Figura 11. Resultados Simulacro Grado Noveno - 2019.

6.3. Prueba simulacro grado décimo

De los 45 estudiantes que se encuentran cursando el Grado Décimo 44 (98%) se encuentra en un nivel bajo de desarrollo en el espacio proyectivo del pensamiento espacial y solo 1 estudiante (2%) alcanzó el nivel básico.

Los resultados indican que el 39% resolvió satisfactoriamente de desarrollo de la perspectiva de objetos diversos; el 44% de estudiantes resolvió adecuadamente el nivel de determinación espacial de secciones de un objeto; el 29% resolvió el nivel de determinación espacial de la rotación; el 28% respondió las preguntas relacionadas con el desarrollo de las relaciones proyectivas. Los resultados señalan que nivel de determinación espacial de la rotación

las relaciones proyectivas son procesos no se están dando al interior de la Institución o bien que las estrategias utilizadas no generan los resultados esperados (Ver figura 12).

Nro.	p1-1	p1-2	p1-3	pp1	p2-1	p2-2	p2-3	p2-4	p2-5	pp2	p3-1	p3-2	p3-3	p3-4	pp3	p4-1	p4-2	p4-3	pp4	P	TOTAL	RESULTADO	RESULTADO%		
10	1	0	0	1	0	0	1	1	1	2	1	0	0	0	2	0	0	0	0	5		BAJO	33.3		
1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	40.0	
2	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
3	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
4	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
5	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
6	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
7	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	40.0	
8	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
9	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	40.0	
10	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
11	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
12	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	53.3	
13	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	40.0	
14	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	40.0	
15	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	53.3	
16	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
17	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	6.7	
18	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
19	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	13.3	
20	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	53.3	
21	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
22	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	46.7	
23	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	53.3	
24	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	46.7	
25	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
26	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
27	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	20.0	
28	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	46.7	
29	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5	11	BASICO	73.3	
30	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
31	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	40.0	
32	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
33	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
34	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	20.0	
35	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	46.7	
36	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	40.0	
37	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	20.0	
38	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	13.3	
39	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	53.3	
40	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	26.7	
41	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
42	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	20.0	
43	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	46.7	
44	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
45	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5		BAJO	33.3	
	23	29	25	77	30					26	22	133	25	33		24	43	125	38	28	37	103			
	135% perdida			57	225					% perdida			59.1	180		% perdida			69.4	135	% perdida		76.3		
	1 Cat									2 Cat						3 Cat					4 Cat				
	Preguntas aprobadas																				438	BAJO	44		
	Total de preguntas																				675	BASICO	1		
																						ALTO	0		
																						SUPERIOR	0		
																						% preguntas perdidas	64.9		
																						% ACIERTOS	35.1		

Figura 12. Resultados Simulacro Grado Decimo - 2019

6.4. Prueba simulacro grado once

De los 33 estudiantes del Grado once con edades comprendidas entre 15 y 18 años que presentaron la prueba simulacro, 28 estudiantes (84%) obtuvo un puntaje bajo y 5 estudiantes (16%) alcanzó el puntaje básico. Los porcentajes de respuestas correctas fueron del 87% en el grupo pp1, 71% en el grupo pp2, 25% en el grupo de preguntas pp3 y 14% en el último grupo pp4. Los resultados muestran avances en los niveles de desarrollo de la perspectiva de objetos

diversos y en la determinación de secciones de un objeto y cuestionables logros respecto al desarrollo del conocimiento de las secciones geométricas, la rotación y desarrollo de superficies.

En términos generales los resultados del simulacro muestran claramente que a pesar de cumplir con la variable edad según lo estipulado en la teoría de Piaget, los estudiantes a quienes se les aplicó la prueba presentan dificultades en el desarrollo del conocimiento espacial, lo que sin duda incide en su proceso de aprendizaje no solo a nivel de matemáticas sino en las demás asignaturas que requieren de un manejo adecuado tanto de las operaciones concretas como del nivel de pensamiento de operaciones formales (Ver figura 13).

Nro.	p1-1	p1-2	p1-3	pp1	p2-1	p2-2	p2-3	p2-4	p2-5	pp2	p3-1	p3-2	p3-3	p3-4	pp3	p4-1	p4-2	p4-3	pp4	P. TOTAL	RESULTADO	RESULTADO %	
11	1	1	1	1	3	1	1	1	0	0	3	0	1	1	0	2	0	0	0	0	8	BAJO	53,3
2	1	1	1	1	3	1	0	1	0	1	3	0	0	0	0	0	1	0	0	1	7	BAJO	46,7
3	1	1	1	1	3	1	1	1	0	1	4	0	0	0	0	0	0	0	0	0	7	BAJO	46,7
4	1	0	1	2	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	3	BAJO	20,0
5	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	BAJO	6,7
6	1	1	1	1	3	1	1	1	1	1	5	0	1	0	0	1	1	0	0	1	10	BASICO	66,7
7	1	1	1	1	3	0	1	1	1	0	3	0	0	0	0	0	1	0	0	1	7	BAJO	46,7
8	1	1	1	1	3	1	1	1	1	1	5	0	0	0	1	1	0	0	0	1	9	BAJO	60,0
9	1	1	0	2	0	0	1	0	1	2	2	0	1	0	1	2	1	0	0	1	7	BAJO	46,7
10	1	1	1	1	3	0	1	1	0	1	3	1	1	0	0	2	1	0	0	1	9	BAJO	60,0
11	1	1	1	1	3	1	1	1	1	0	4	0	1	0	0	1	0	0	0	0	8	BAJO	53,3
12	1	1	1	1	3	1	1	1	1	1	5	0	1	0	0	1	1	0	0	1	10	BASICO	66,7
13	1	1	1	1	3	1	1	1	0	1	4	0	0	0	0	0	1	0	0	1	8	BAJO	53,3
14	1	1	1	1	3	0	0	1	0	1	2	0	0	0	0	0	1	0	0	1	6	BAJO	40,0
15	1	1	1	1	3	1	1	1	1	1	5	0	1	0	0	1	1	0	0	1	10	BASICO	66,7
16	1	1	1	1	3	0	1	1	1	1	4	0	0	0	0	0	0	0	0	0	7	BAJO	46,7
17	1	1	1	1	3	1	1	1	1	1	5	0	1	0	0	1	1	0	0	1	10	BASICO	66,7
18	1	1	1	1	3	1	1	1	0	0	3	0	0	0	0	0	0	0	0	0	6	BAJO	40,0
19	1	1	1	1	3	0	1	1	1	1	4	1	1	0	0	2	0	0	0	0	9	BAJO	60,0
20	1	0	1	2	1	1	1	0	1	4	0	0	0	0	0	0	0	0	0	0	6	BAJO	40,0
21	1	1	1	1	3	1	0	0	0	1	2	0	1	1	0	2	0	0	0	0	7	BAJO	46,7
22	1	1	1	1	3	1	1	1	0	1	4	0	0	0	0	0	0	0	1	1	8	BAJO	53,3
23	1	1	1	1	3	0	1	1	0	1	3	1	1	1	0	3	0	0	0	0	9	BAJO	60,0
24	0	0	1	1	1	1	1	0	0	3	0	1	0	0	1	0	1	1	1	2	7	BAJO	46,7
25	1	0	1	2	1	1	1	1	0	4	0	1	1	0	2	0	0	0	0	0	8	BAJO	53,3
26	1	1	1	1	3	1	1	1	0	1	4	0	0	0	0	0	0	0	0	0	7	BAJO	46,7
27	1	0	1	2	1	1	1	1	0	4	0	1	1	0	2	0	0	0	0	0	8	BAJO	53,3
28	1	1	1	1	3	1	1	1	1	1	5	0	1	1	0	2	0	0	0	0	10	BASICO	66,7
29	1	1	1	1	3	1	1	1	0	0	3	0	1	0	0	1	0	0	0	0	7	BAJO	46,7
30	0	0	1	1	1	0	1	0	1	3	0	1	0	0	1	0	1	0	1	1	6	BAJO	40,0
31	0	1	1	2	1	1	1	1	0	4	0	1	1	0	2	0	0	0	0	0	8	BAJO	53,3
32	1	1	1	1	3	1	1	1	1	0	4	0	1	1	0	2	0	0	0	0	9	BAJO	60,0
33	1	1	1	1	3	1	1	1	1	1	5	0	1	0	0	1	0	0	0	0	9	BAJO	60,0
4	7	2	13	9	7	3	17	11	47	30	13	25	31	99	23	31	31	85			49,3	50,7	
99	% perdida		13,1	165			% perdida		28,5	132	% perdida		75,0	99	% perdida		85,9				% preguntas perdidas	% ACIERTOS	
			1 Cat						2 Cat					3 Cat				4 Cat					
																					Preguntas aprobadas	244	
																					Total de preguntas	495	
																						BAJO	28
																						BASICO	5
																						ALTO	0
																						SUPERIOR	0

Figura 13. Resultados Simulacro Grado Once - 2019.

6.5. Desarrollo del modelo

Los datos que se usaron para alimentar el modelo predictivo provenían básicamente de tres fuentes: prueba simulacro, factores sociodemográficos y factores académicos, de lo anterior se realiza un análisis exploratorio a nivel de minería de datos para encontrar información relevante que pueda aumentar la capacidad de predicción del modelo en términos de identificar el desempeño de los estudiantes en el desarrollo del espacio proyectivo del pensamiento espacial.

Los datos que se obtuvieron en la prueba simulacro y en la encuesta de factores sociodemográficos se tabularon en Excel para luego procesarlos y registrarlos como archivo histórico de las actividades de estudiantes y docentes. Estos datos se usaron posteriormente como insumo que permitió el desarrollo del modelo predictivo.

De acuerdo con los planteamientos anteriores se procede a describir el desarrollo del modelo predictivo con base en la metodología que se ajusta al modelo predictivo de Machine learning y a lo que sugieren los teóricos en este caso, en consecuencia, se describe la metodología CRISP-DM haciendo énfasis en cada una de las etapas propuestas.

6.5.1. Comprensión del contexto

En esta etapa se procedió a explicitar las características de la Institución Educativa Departamental General Carlos Albán a partir de las directrices del Ministerio de Educación Nacional de Colombia, tales directrices enfocan en construir un modelo estandarizado para el manejo de datos de los estudiantes. La fuente de datos principal fue el libro en el que se registra el proceso de matrícula, lo mismo que, las pruebas académicas diseñadas y aplicadas por el cuerpo docente con determinada rigurosidad en todas las asignaturas. Las fuentes mencionadas

se convirtieron en soporte académico y formal del proceso y de alguna manera garantizan su validez.

Los datos que se obtuvieron de la prueba simulacro y de la encuesta de factores sociodemográficos se tabularon con el propósito de procesarlos como registro histórico de las actividades de estudiantes y docentes.

Uno de los retos que plantea el modelo predictivo es el de comprender los datos; este ejercicio implicó entender que el set de datos lo conforman un conjunto de 119 variables entre categóricas, numéricas y booleanas que describieron la situación sociodemográfica de los estudiantes de la Institución Educativa Departamental General Carlos Albán, la información fue recopilada por la Institución en el proceso de matrícula de acuerdo con los lineamientos del Ministerio de Educación Nacional.

Para efectos de simplificar el análisis de los datos, estos se exportaron en un documento de Excel con autorización previa de los estudiantes y de la Institución Educativa.

El proceso de exploración de estos datos se realizó con un programa de computador desarrollado en el lenguaje Python, el cual utilizó una librería datos de Pandas de Open Source que realiza el procesamiento, exploración y análisis de los datos diseñados en Python y Jupyter que es un entorno de trabajo orientado al uso de cuadernos de investigación que condensa los procesos ejecutados para la solución o análisis de un problema en donde el código es la base. Este entorno de trabajo se usa comúnmente en el desarrollo de proyectos de Machine Learning, Simulación, Inteligencia Artificial, entre otros (<https://jupyter.org>)

Estos programas hacen un barrido de todas las variables que conforman la base de datos y logran entender un poco la naturaleza, forma, peso correlaciones y calidad de los datos recopilados.

6.5.2. Información de la base de Datos.

Number of variables	119
Number of observations	132
Total Missing (%)	21.3%
Total size in memory	122.8 KiB
Average record size in memory	952.6 B

Figura 14. Información Inicial Data Set.

La Figura 14 muestra un 21.3% de datos perdidos en la base de datos principal, esto implicó evaluar el tipo de dato faltante y reemplazarlo por la media de datos continuos, en este caso, se reemplazó por el de más frecuencia que es el de los datos categóricos nominales y ordinales.

También se evidenció la poca cantidad de observaciones (132) que no fueron suficientes para aplicar algoritmos de Machine Learning en su análisis por lo que se recurrió a realizar simulaciones con algoritmos computacionales y método estadístico Montecarlo Metropolis - Hastings¹ que permite generar muestras de cualquier distribución genérica de datos implementados en la librería PyMC3², Esta librería es un set de utilidades computacionales para programación que facilita el proceso de obtención de una nueva base de datos experimental, tal base de datos presenta la dificultad de ser una caja negra para el investigador, sin embargo, la documentación puede orientar al lector. (Ver Anexo 5 de Documentación del Algoritmo).

¹ Según (Martín, 2018) Metrópolis – Hasting permite obtener muestras de cualquier distribución de probabilidad, dado que se puede calcular al menos un valor proporcional a ella, ignorando así el factor de normalización.

² Es una librería de Python desarrollado para la comunidad científica. Se caracteriza por tener un set de algoritmos prefabricados para el modelado estadístico que se centra en el uso de cadenas de Markov, Montecarlo y Bayesianos

Numeric	63
Categorical	27
Boolean	24
Date	0
Text (Unique)	0
Rejected	5
Unsupported	0

Figura 15. Tipo de Variables.

6.5.3. Interpretación de los datos.

Al realizar la exploración de los datos se encontraron una serie de alertas en términos de datos faltantes o perdidos, presencia de ceros catalogados como nulos que debieron ser corregidos; también se encontraron casos en donde las variables presentaban comportamientos poco favorables para el desarrollo del modelo debido a que los datos no se encontraban balanceados y podrían llevar a sesgar las predicciones; este fue el caso de la variable DANE_RESIDE_DEPTO o DEPART_NACE que se muestra como booleana, pero contienen datos categóricos; sumado a que no se comportaba como una variable normal aceptable según se observa en las Figuras 16, 17 y 18 y en el Anexo 5 de documentación del Algoritmo.

- 6_HA_PRESENTADO_PRUEBAS_SABER has 35 / 26.5% zeros Zeros
- 6_HA_PRESENTADO_PRUEBAS_SABER has 5 / 3.8% missing values Missing
- APELLIDOS Y NOMBRES has a high cardinality: 131 distinct values Warning
- AREA_RESIDENCIA has 5 / 3.8% missing values Missing
- DANE_RESIDE_DEPTO has 3 / 2.3% missing values Missing
- DANE_RESIDE_MCPIO has 3 / 2.3% missing values Missing
- DEPART_NACE has 3 / 2.3% missing values Missing
- DEPTO_RESIDE has 3 / 2.3% missing values Missing
- DISCAPACIDAD has 3 / 2.3% missing values Missing
- ESTRATO has 3 / 2.3% missing values Missing
- ETNIA_PERTENECE has 3 / 2.3% missing values Missing
- ETNICO_MINORITARIO has 3 / 2.3% missing values Missing
- FECHA_NACIMIENTO has 4 / 3.0% missing values Missing
- FECHA_NACIMIENTO has a high cardinality: 126 distinct values Warning
- GENERO has 3 / 2.3% missing values Missing
- GRADO has 3 / 2.3% missing values Missing
- HABITACION_DUERME_HOGAR has 11 / 8.3% missing values Missing

Figura 16. Información de Datos Faltantes o Perdidos.

Value	Count	Frequency (%)
25.0	129	97.7%
(Missing)	3	2.3%

Figura 17. Variable DANE_RESIDE_DPTO.

Value	Count	Frequency (%)
Cundinamarca	115	87.1%
Tolima	3	2.3%
Santander	2	1.5%
Zulia	1	0.8%
Antioquia	1	0.8%
Risaralda	1	0.8%
Maracaibo	1	0.8%
Meta	1	0.8%
Guaviare	1	0.8%
Magdalena	1	0.8%
Other values (2)	2	1.5%
(Missing)	3	2.3%

Figura 18. Variable DEPART_NACE.

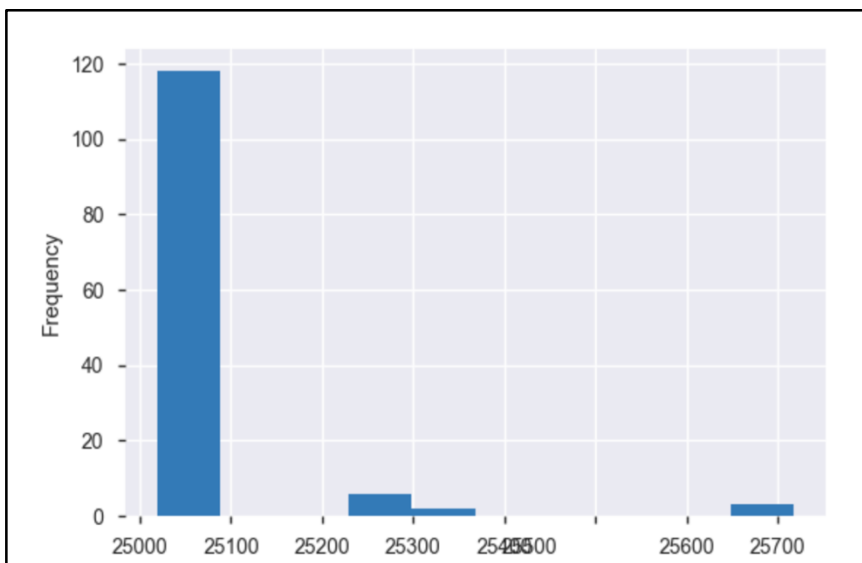


Figura 19. Distribución DANE_RESIDE_DEPTO.

En algunas variables se presentaban datos atípicos que causaban sesgos, en otros términos, su presentación se hacía en diversos formatos que afectaban el proceso de predicción; así, por ejemplo: la variable edad se registraba en diversos formatos; unos colocaban día, mes y año en número, otros en forma de texto o cambiaban el orden, primero registraban el año, luego el mes y finalmente el día etc. Además, los sujetos objeto de la muestra tenían diversidad de edades, unos estudiantes tenían 21 años mientras que, otros tenían sólo 12 años, al promediar las edades se generaba deformidad en los datos; situación similar ocurría con otras variables; así que, se decidió entonces, pasar todos los datos a un solo formato para sanearlos, es decir, para normalizarlos.

La sanidad de datos se encontró en variables como EDAD, Figura 15, que presentaba un comportamiento normal en algunos datos lo que permitía un mejor performance en términos de la importancia que puede tener esta variable para el clasificador en el momento de la predicción, sin embargo, se encontraron algunos datos atípicos que fueron tratados (normalizados) para evitar el sesgo.

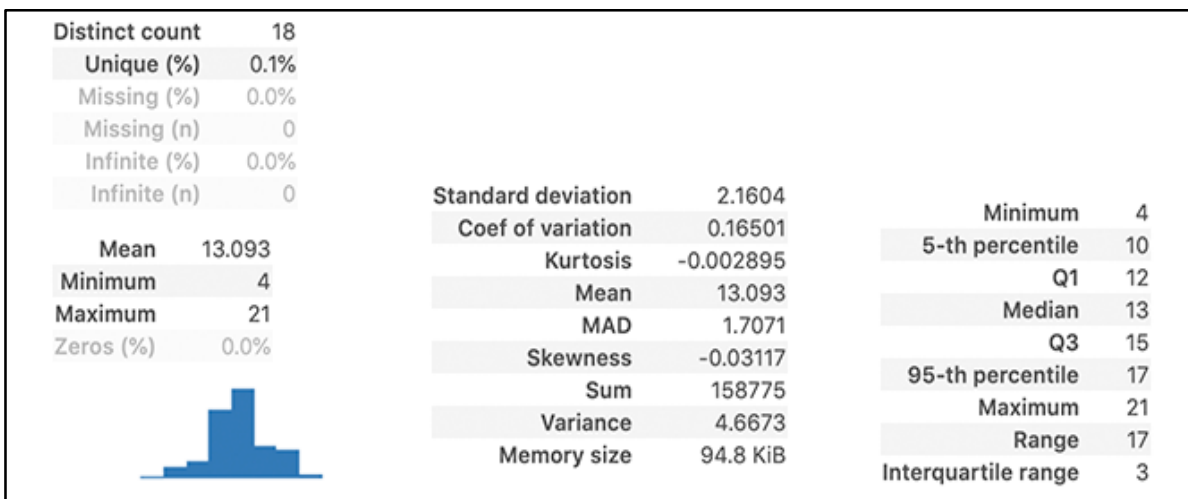


Figura 20. Distribución EDAD.

Por todos los aspectos atípicos que se encontraron en las distintas variables, se sugiere revisar los Anexos 3 de Documentación del Algoritmo donde se muestran en detalle el análisis del total de las variables.

Preparación de los datos. En esta etapa se trabajó con el promedio acumulado de las variables usadas por el clasificador y se organizaron los promedios en orden de importancia, debido a que, los clasificadores utilizados trabajan con árboles y estos son invariantes a las transformaciones de los datos normalizados.

El set de datos obtenido después del análisis exploratorio requirió de conversión de tipos de datos, Figura 21 y el Anexo 5 de Documentación del Algoritmo, solución de datos faltantes compensándolos, en los casos de las variables categóricas y booleanas, con el dato con más frecuencia en la columna y en los datos numéricos con la media.

```
[137]: GENERO                object
      GRADO                object
      FECHA_NACIMIENTO      datetime64[ns]
      MUNICIPIO_NACE        object
      SABER_3               bool
      SABER_5               bool
      SABER_9               bool
      SABER_PRO_11         bool
      SIMULACROS            bool
      DISCAPACIDAD          int64
      DEPTO_RESIDE          object
      DANE_RESIDE_DEPTO     int64
      MCPIO_RESIDE          object
      AREA_RESIDENCIA       object
      NOMBRE_URBANO         object
      NOMBRE_RURAL          object
      NIVEL_EDUCATIVO_PADRE object
      NIVEL_EDUCATIVO_MADRE object
      ESTRATO               int64
      SISBEN                float64
      PERSONA_CONFORMA_HOGAR float64
      NRO_HERMANAS          float64
      NRO_HERMANOS          float64
      MATERIAL_PISOS        object
      HABITACION_DUERME_HOGAR float64
      HOGAR_INTERNET        bool
      HOGAR_SERVICIO_TV     bool
      H_PC                  bool
      H_LAVADORA             bool
      H_HORNO_MICROO        bool
      ...
```

Figura 21. Tipos de Datos de Dataset.

En algunos casos como en la variable REMUNERACIÓN_TRABAJO existía una alta dispersión de los datos continuos que lo conformaban, por lo que se optó por agruparlos en rangos, convertirlos en categóricos y mejorar el dataset con el fin de optimizar el modelo resultante. Otra consideración que se tuvo en cuenta fue el proceso de estandarización de los datos: codificar, quitar caracteres extraños y espacios en blanco etc.

El entrenamiento y desarrollo del modelo predictivo requirió la conversión de datos atípicos para factorizarlos de manera eficiente a fin de obtener un dataset homogéneo, es decir, que se pudiera convertir la codificación de datos categóricos a datos ordinales según lo propuesto por (Fuentes, 2018; Hale, 2018).

En el caso de las variables dependientes fue necesario agrupar los resultados obtenidos en variables muy puntuales que permitieran describir el comportamiento de la variable a predecir, tal fue el caso de la variable pp1 que estaba compuesta por el resultado de tres preguntas cuyo resultado fue promediado bajo la variable RESULTADO_1, esta agrupación permitió definir la variable a predecir de forma más clara.

Simulación. En esta etapa se partió de una implementación simple de Montecarlo en atención a los planteamientos de Martin (2018); como se observa en la Figura 15, para saltar a una implementación con PyMC3 de Montecarlo-Metrópolis-Hastings que permitió que la simulación se adaptara a los diferentes tipos de datos aunque con un poco más de carga computacional lo que hizo lento el procesamiento, al generar la muestra de 5000 observaciones tardó alrededor de 40 minutos (esto depende también de la máquina utilizada para la simulación).

```
[ ]: def metropolis(func, draws=2000):
    trace = np.zeros(draws)
    old_x = 0.5 # func.mean()
    old_prob = func.pdf(old_x)
    for simulacion in range(0, simulaciones):
        for key in data.columns.values:
            if data[key].dtype == 'object':
                artificial_data[key] = np.random.choice(data[key])
            elif data[key].dtype == 'bool':
                artificial_data[key] = np.random.choice([True, False])
            elif data[key].dtype == 'float64':
                if key == 'pp1' or key == 'pp2' or key == 'pp3' or key == 'pp4':
                    aux = float(round(abs(data[key].mean() + data[key].std() * np.random.randn(simulaciones)[0]),2))
                    if aux > 1:
                        artificial_data[key] = 1
                    else:
                        artificial_data[key] = aux
                else:
                    artificial_data[key] = float(round(abs(data[key].mean() + data[key].std() * np.random.randn(simulaciones)[0]),2))
            else:
                artificial_data[key] = abs(math.trunc(
                    data[key].mean() + data[key].std() * np.random.randn(simulaciones)[0]
                ))
        data = data.append(artificial_data, ignore_index=True)
```

Figura 22. Implementación Simple de Montecarlo.

En el caso de la simulación con PyMC, el uso de la librería Metrópolis simplificó el proceso de creación de una muestra, aunque esta librería reporta una gran variedad de datos, en la Figura 23 se destaca la simulación de la variable HABITACION_DUERME_HOGAR que

arroja una distribución normal un poco corrida a la izquierda pero con mayor número de observaciones; la mayor cantidad de observaciones se debe a que toma como muestra el dato anterior y predice los siguientes, esto permite escalar las observaciones minimizando la degradación de los datos. Este caso es muy particular al dar de cómo se esperaría.

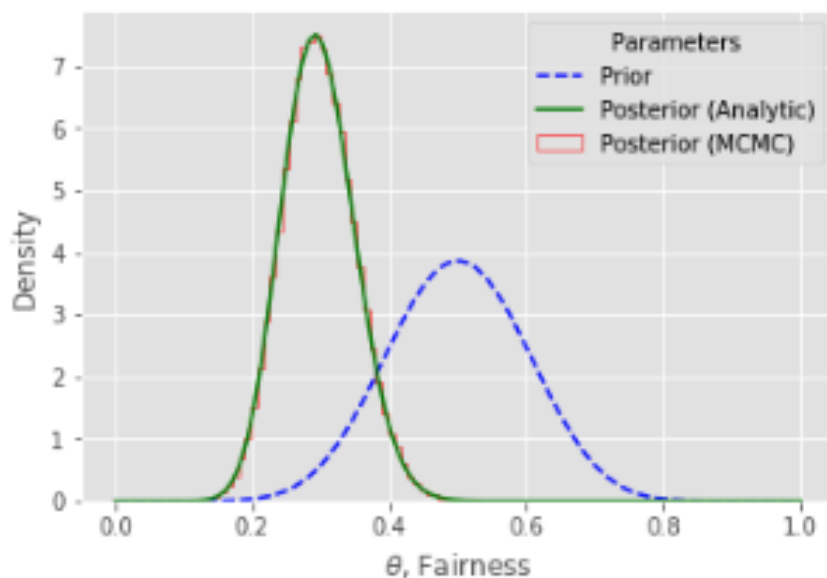


Figura 23. Variable HABITACION_DUERME_HOGAR.

Ingeniería de Características: Ingeniería de Características: esta etapa se trataba de probar los algoritmos encargados de procesar la información a nivel computacional con el propósito de comparar cuál de ellos hacia la selección de los mejores atributos que permitieran aportar poder de predicción al clasificador. El poder del clasificador se mide por la capacidad que tenga de predecir con el mayor nivel de acierto, el comportamiento futuro de un sujeto en un contexto de estudio determinado, en el caso de esta investigación será, la predicción del desempeño en el desarrollo del espacio proyectivo del pensamiento espacial de los estudiantes objeto de estudio.

Se procedió con la prueba del primer algoritmo, el cual realizó una eliminación recursiva de variables que arrojó como resultado un `rfecv.n_features_ = 4`. Este resultado no fue eficiente debido a que entregaba una base de datos de 100 variables, que para el caso era muy amplia; en segundo lugar, se corrió una prueba de eliminación exhaustiva con el uso de librerías de código previamente procesadas y que ofrecen soluciones como `SelectKBest` y `ANOVA F-Value` para la selección de atributos; esta prueba entregó como resultado: `Original number of features=100` y `Reduced number of features=30`, significa que tomó la base de datos compuesta por 100 variables, la redujo a 30 y eligió las variables que consideró más importantes; este algoritmo redujo la dimensión de la base de datos pero, no entregó los pesos específicos en orden de importancia ni el peso de importancia total acumulado que describen qué tan relevante es una variable para poder predecir un comportamiento futuro de un contexto de estudio determinado, en este caso se trata de la predicción del desempeño de los estudiantes en el desarrollo del espacio proyectivo del pensamiento espacial.

Finalmente, el mejor resultado se obtuvo con `Random Forest Classifier` y `Cross Validation` que fueron algoritmos propios de `Machine Learning` que procesan la información a nivel computacional para la selección de los mejores atributos que aportarían poder de predicción al clasificador, el resultado obtenido arrojó 63 variables que aportaron un total del 95% de importancia acumulada como lo soporta (Van der laan, 2011; Boulesteix et al., 2013) (Ver Figura 17).

Del proceso anterior se seleccionó el modelo “`SelectFromModel`” que hace parte de un conjunto de herramientas algorítmicas presentes en una de las librerías de código mencionadas y utilizadas para el desarrollo del modelo predictivo que permitió generar una nueva base de datos para el entrenamiento del modelo final; con este algoritmo se obtuvieron las variables y el peso

de importancia de cada una, lo cual posibilitó la interpretación. Ver con más detalle en la Figura 24 y en los (Anexos 5 Documentación del Algoritmo).

En la Figura 24 se ve el corte en un 95%, estos pesos se dan en un rango de 0 a 1 y en la intersección de la curva con la línea horizontal se traza una recta vertical imaginaria que crea el límite que separa las variables que deben ser descartadas; es de anotar que en las iteraciones en el entrenamiento y calibración del modelo podrían eliminarse más variables o recuperar el total de la base de datos original.

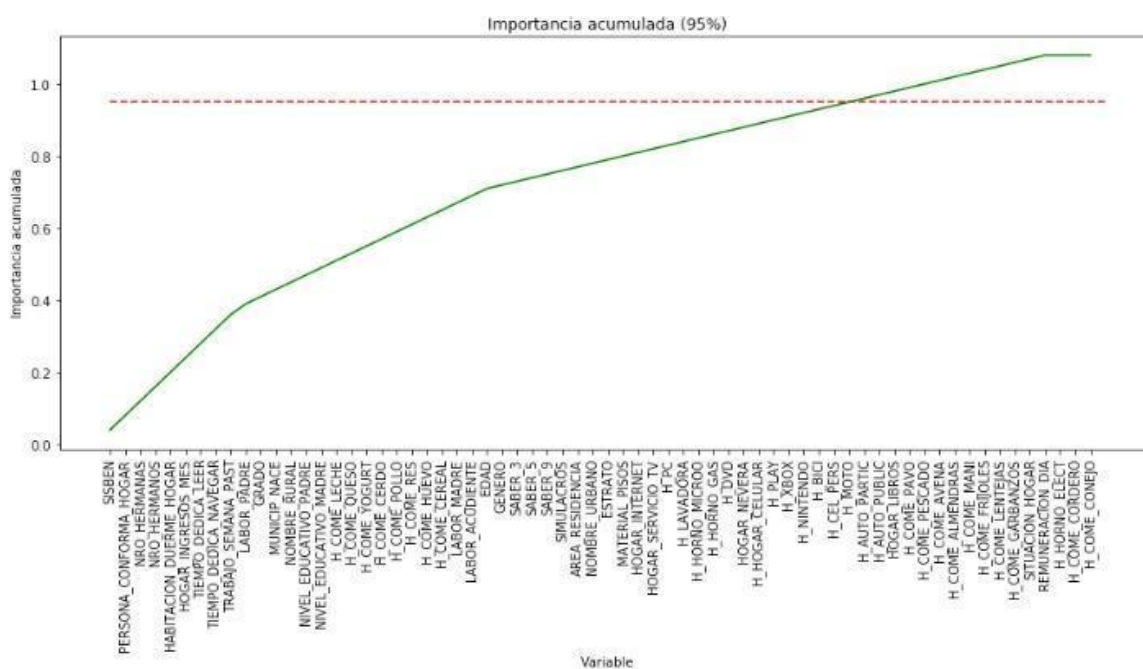


Figura 24. Importancia Acumulada de Variables.

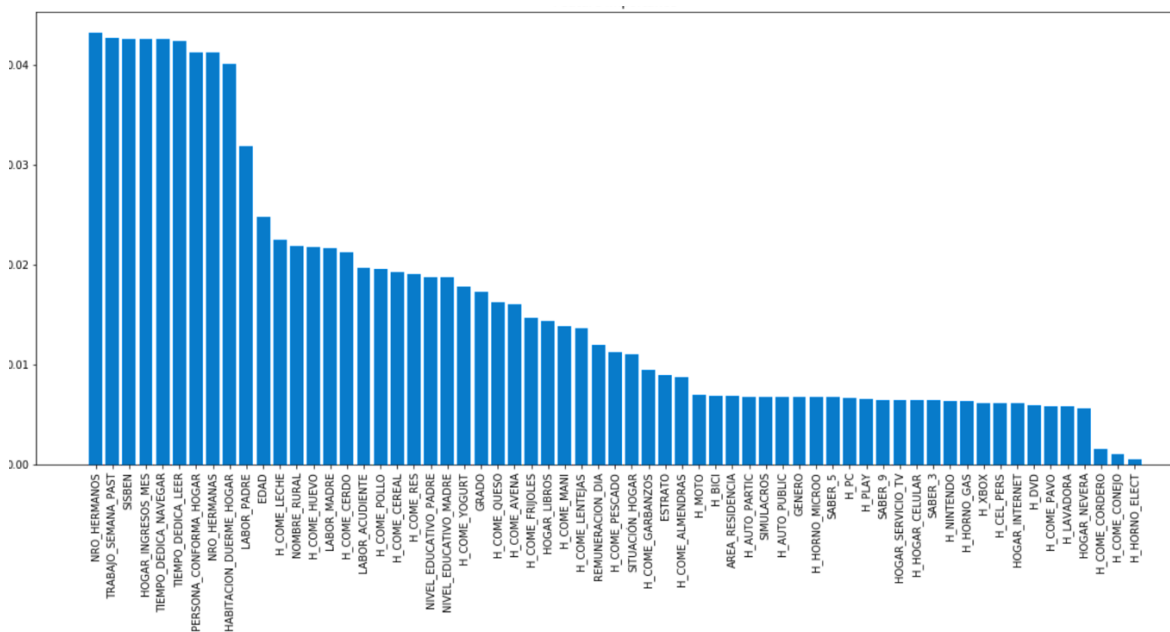


Figura 25. Variables por Importancia.

La Figura 25 muestra el resultado final en la selección de variables ordenadas según sus pesos, fué importante analizarlas junto al experto en el área disciplinar para validar lo reportado por el clasificador automático.

Modelamiento. En términos del desarrollo del modelo predictivo se evaluaron cuatro algoritmos de Machine Learning para encontrar el mejor rendimiento en términos de la calidad de la predicción y la naturaleza de los datos y el problema a resolver.

El problema que se buscaba resolver se configura como una clasificación, puesto que el objetivo era determinar si las técnicas de Machine Learning pueden predecir el desempeño que de los estudiantes en el desarrollo del espacio proyectivo del pensamiento espacial. Teniendo en cuenta las sugerencias de los teóricos se implementaron algoritmos de clasificación como Suport Vector Machine, K Nearest Neighbours, RandomForestClassifier y el GradientBoostingClassifier que se utilizó en el desarrollo del modelo predictivo final, es de

anotar que los algoritmos mencionados hacen parte de un conjunto de técnicas computacionales muy utilizadas cuando se trabaja con Machine Learning, lo anterior se detalla en el Anexo 5; la correcta aplicación de las técnicas de selección de atributos fue importante para proceder con el entrenamiento y pruebas funcionales del modelo predictivo.

Evaluación. El mejor resultado en la predicción y clasificación fue el obtenido con la aplicación del Gradient Boosting Classifier con un ROC (Característica de funcionamiento del receptor) de 0.866 que muestra el rendimiento del modelo en términos de los falsos positivos y verdaderos positivos clasificados en las pruebas que significa explicitar y mostrar cuantos intentos logró predecir de manera correcta y en cuantos se equivocó, como lo muestra la Figura 26.

Es importante resaltar que el área bajo la curva permite comprender el comportamiento del modelo en términos de su precisión, University of Nebraska Medical Center UNMC (2001) y Rice (2005) dan una clasificación del valor resultante así: 0.90 - 1 = Excelente; 0.80 - 0.90 = bueno, 0.70-0.80 = aceptable, 0.60-0.70 = pobre y 0.50-0.60 = malo; por lo que para este caso el resultado de precisión del modelo según la Figura 26 es bueno ya que está entre 0.80 y 0.90.

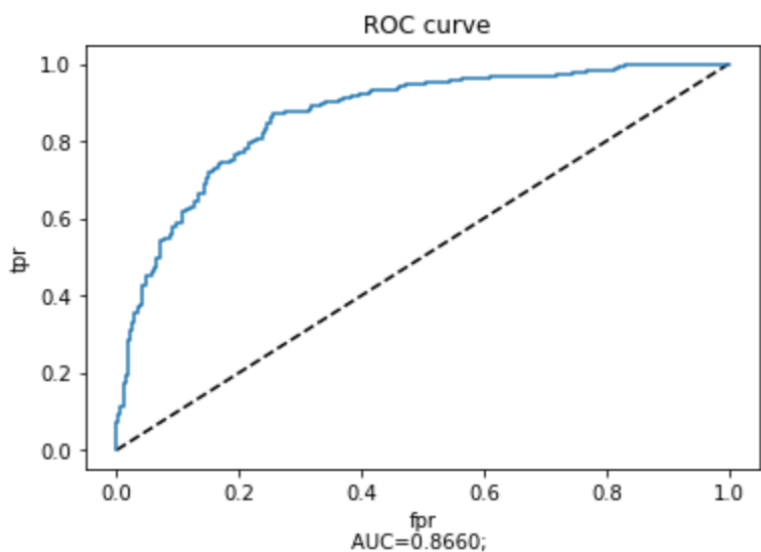


Figura 26. ROC - AUC.

De otra parte, la matriz de confusión de la Figura 27 resume la forma en la que el modelo hizo la clasificación. Para el caso de los estudiantes que están en DESEMPEÑO_BAJO, se tiene que del dataset de test clasificó de manera correcta 163 de 223 sujetos y deja como falsos positivos 66 estudiantes en DESEMPEÑO_ALTO. En DESEMPEÑO_ALTO de 277 estudiantes clasificó de manera correcta 235, pero tuvo 42 falsos positivos que ponen a los estudiantes en un nivel de desempeño bajo del desarrollo del espacio proyectivo del pensamiento espacial.

Ahora bien, si se clasifica el Accuracy reportado por la matriz de confusión que establece la relación entre el número de predicciones correctas y el número total de muestras de entrada y se ubica en la escala propuesta (UNMC, 2001; Rice, 2005) se concluiría que es aceptable, lo que sugiere realizar una nueva iteración en la metodología evaluando nuevamente en cada uno de los pasos; ajustando debilidades encontradas como las pocas observaciones existentes, la agregación de nuevas variables entre otros.

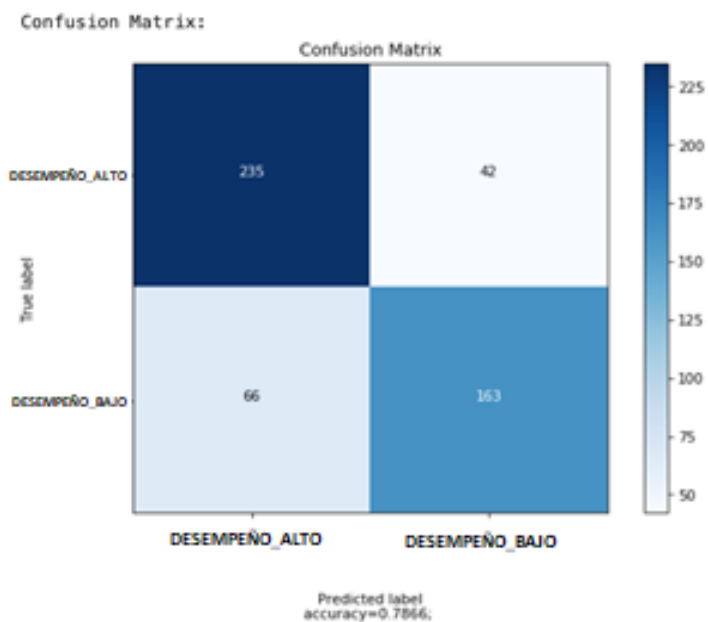


Figura 27. Matriz de Confusión.

Según el Anexo 5 de Documentación del Algoritmo y lo sugerido por (Ramírez, 2018), se tiene el reporte definitivo del modelo obtenido. El recall resultado clave del proceso de desarrollo del modelo que reportó el ratio de positivos reales detectados en el dataset del clasificador, muestra para un grupo de estudiantes DESEMPEÑO_ALTO con un valor de 0.85; este valor es claramente superior al que muestra la clasificación de los estudiantes en DESEMPEÑO_BAJO cuyo desarrollo del espacio proyectivo del pensamiento espacial señala un valor de 0.71 que es considerado por los teóricos como aceptable.

El hecho de encontrar esta diferencia en la forma de clasificar en cada categoría sugiere que se debe realizar la recolección de más observaciones y una nueva iteración para entrenar y calibrar nuevamente el modelo, sin embargo, el promedio del recall reportado es aceptable y deja saber que el modelo propuesto si es capaz de detectar si puede haber desempeño alto del desarrollo del espacio proyectivo del pensamiento espacial en los estudiantes o bien, un desempeño bajo. Estas predicciones le permiten docentes e instituciones tomar decisiones anticipadas para evitar el fracaso académico y mejorar el proceso de aprendizaje de los estudiantes.

Classification Report					
	precision	recall	f1-score	support	
0	0.78	0.85	0.81	277	
1	0.80	0.71	0.75	229	
micro avg	0.79	0.79	0.79	506	
macro avg	0.79	0.78	0.78	506	
weighted avg	0.79	0.79	0.79	506	

Figura 28. Reporte del clasificador.

Despliegue. Si bien la ejecución de la metodología CRISP-DM sugiere una etapa de puesta en producción del modelo predictivo resultante donde el entregable puede ser un reporte; en este trabajo se desarrolló una herramienta que permite simular predicciones con el modelo de Machine Learning resultante, las simulaciones pueden ser on-line (de manera transaccional, uno a uno) o en batch (un grupo de estudiantes) lo que permite al docente cargar un conjunto de estudiantes para obtener la probabilidad de desempeño de un estudiante clasificado en desempeño alto o desempeño bajo, de igual forma es posible integrar el modelo en cualquier sistema como OVAS, AVAS o software administrativo para instituciones educativas.

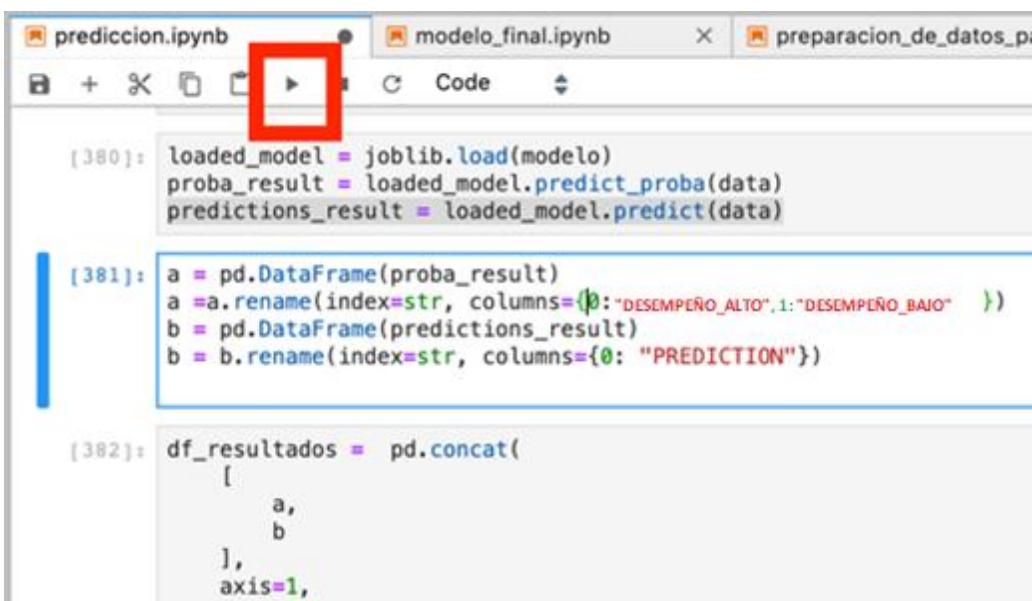
Simulación en Batch. La simulación en batch se puede realizar de manera sencilla, solo se requiere el input de datos en un archivo con formato CSV (Valores separados por comas) que contendrá las entradas correspondientes a las descritas en el Anexo 5 de documentación del Algoritmo y el Jupyter notebook. (Ver Anexo 5 de Documentación del Algoritmo)

La simulación se realizó para con los grados octavo, noveno décimo y once de la Institución Educativa Departamental General Carlos Albán de Albán – Cundinamarca determinado como grupo validador.

Se abre con Jupyter el notebook del Anexo 5 de Documentación del Algoritmo; en él se encuentran las siguientes líneas: *modelo = 'pp1_model.sav'* que corresponde al modelo de Machine Learning resultante, *loaded_model = joblib.load(modelo)* permite cargar el modelo mencionado para ser utilizado, al cargar el modelo podemos usar la función *proba_result = loaded_model.predict_proba(data)* que recibe como argumento los datos de los estudiantes que

queremos simular y nos retornará las probabilidades de desempeño que tendrá cada estudiante en el desarrollo del espacio proyectivo del pensamiento espacial y `predictions_result = loaded_model.predict(data)` devolvió las predicciones que etiqueta a cada estudiante como DESEMPEÑO_ALTO y DESEMPEÑO_BAJO en el desarrollo del espacio proyectivo del pensamiento espacial.

Para ver el resultado es necesario iniciar la simulación dando click en el botón play que está marcado en rojo en la Figura 29.



```

prediccion.ipynb | modelo_final.ipynb | preparacion_de_datos_pa
+ | ✂ | 📄 | ▶ | 🔄 | Code | ⚙

[380]: loaded_model = joblib.load(modelo)
       proba_result = loaded_model.predict_proba(data)
       predictions_result = loaded_model.predict(data)

[381]: a = pd.DataFrame(proba_result)
       a = a.rename(index=str, columns={0: "DESEMPEÑO_ALTO", 1: "DESEMPEÑO_BAJO" })
       b = pd.DataFrame(predictions_result)
       b = b.rename(index=str, columns={0: "PREDICTION"})

[382]: df_resultados = pd.concat(
      [
        a,
        b
      ],
      axis=1,
  
```

Figura 29. Ejecutor Simulación en Batch.

El resultado de la simulación para los 41 estudiantes del grupo comparativo entrega una salida como lo muestra la Figura 30, que puede ser vista en el Anexo 5 de Documentación del Algoritmo; la simulación entregó un total de 20 estudiantes en DESEMPEÑO_BAJO y 21 estudiantes en DESEMPEÑO_ALTO; la columna DESEMPEÑO_ALTO muestra para el primer estudiante de la lista una probabilidad de 0.66 que indica que esté en DESEMPEÑO_ALTO, dicha probabilidad se estableció en el rango de 0 a 1 donde 0 es la probabilidad más baja y 1 la

probabilidad más alta. En la columna DESEMPEÑO_BAJO el mismo estudiante presenta un 0.33 de probabilidad bajo desempeño en el desarrollo del espacio proyectivo del pensamiento espacial. Por último, la columna PREDICCION muestra el resultado final del clasificador que para el caso del estudiante en mención es DESEMPEÑO_ALTO tal que $DESEMPEÑO_BAJO = 0.33 < DESEMPEÑO_ALTO = 0.66$; lo anterior muestra la forma en que el modelo de Machine Learning devuelve los resultados.

```
[ 396 ]: df_resultados
```

[396]:	DESEMPEÑO_ALTO	DESEMPEÑO_BAJO	PREDICCION
0	0,660025	0,339975	DESEMPEÑO_ALTO
1	0,493829	0,506171	DESEMPEÑO_BAJO
2	0,432395	0,567605	DESEMPEÑO_BAJO
3	0,604963	0,395037	DESEMPEÑO_ALTO
4	0,266532	0,733468	DESEMPEÑO_BAJO
5	0,259256	0,740744	DESEMPEÑO_BAJO
6	0,678635	0,321365	DESEMPEÑO_ALTO
7	0,589109	0,410891	DESEMPEÑO_ALTO
8	0,528724	0,471276	DESEMPEÑO_ALTO
9	0,356903	0,643097	DESEMPEÑO_BAJO
10	0,615625	0,384375	DESEMPEÑO_ALTO
11	0,483325	0,516675	DESEMPEÑO_BAJO
12	0,493829	0,506171	DESEMPEÑO_BAJO

Figura 30. Salida de la Simulación para el grupo comparador - 2019.

En la Figura 30, el modelo predictivo muestra que estudiantes están en bajo o alto desempeño y cuáles no, estos resultados permiten al docente identificar los niveles de desempeño de los estudiantes y diseñar estrategias puntuales y personalizadas que contribuyan a que los estudiantes superen sus dificultades y alcancen altos niveles de desempeño. Además, el modelo predictivo permite identificar aquellas variables académicas y socioculturales que mayores afectaciones podrían provocar en el desempeño de los estudiantes, por tanto, se podrían controlar sus efectos (Ver Anexo 4).

7. Discusión y Análisis de Resultados del Uso del Clasificador Final con un Grupo Real de Estudiantes

Con el propósito de comparar los resultados del algoritmo, se tomaron los datos de la encuesta, de las variables sociodemográficas y de la prueba del simulacro de un grupo conformado por 41 estudiantes de la institución pertenecientes a los grados: octavo, noveno, décimo y once. Estos resultados fundamentan la discusión.

La edad de los 41 estudiantes que conforman el grupo comparativo se encuentra en un rango de edad de 12 - 16 años. De ellos el 38% son mujeres y el 62% son hombres.

Los datos obtenidos de la encuesta de factores sociodemográficos muestran que el 54% de los estudiantes viven en el sector rural; en un 80% pertenecen al estrato 1; el 51% de los padres de los estudiantes solo alcanzan nivel primario de estudio, el 20% cuenta con un máximo de 30 libros en el hogar; en un 61% son hogares conformados por 4 a 5 personas; el 58% vive en casas cuyos pisos están contruidos en cemento; el 17% de los papás no trabajan y el 51% de las mamás son amas de casa. El 66% de los padres perciben mensualmente el salario mínimo o menos de él, lo que comprueba que el 16% de los niños viven en hogares con situación económica desventajosa. Con relación a la presencia de electrodomésticos en el hogar el porcentaje de quienes lo poseen es muy bajo. En la mayoría de los casos no tienen acceso a celulares ni internet.

Las familias de los estudiantes es su mayoría consumen productos básicos de la canasta familiar y tienen muy poca opción de consumir productos como la carne o el pescado.

Los resultados de la caracterización socioeconómica dejan ver en particular que, el nivel de estudios de los padres predetermina el resultado académico de sus hijos desde las etapas más tempranas de la enseñanza. Estudios como el realizado por (Moncayo, 2016) señalan que factores socioeconómicos como la educación de los padres, los ingresos o nivel salarial, el

género, las horas de trabajo dedicadas al mejorar el rendimiento académico, el tipo de colegio y los recursos tecnológicos influyen directamente sobre los procesos de aprendizaje de los individuos; otras variables sociodemográficas también determinan su influencia negativa sobre el aprendizaje ya que alteran su estabilidad emocional.

Es bueno enfatizar que la mayoría de los estudiantes del grupo comparativo viven en el sector rural y normalmente, después de la jornada escolar deben llegar a ayudar en las labores del campo, lo que les resta tiempo en sus horas de trabajo sobre su desempeño académico.

Tristemente en nuestro país los colegios públicos adolecen de recursos económicos que les permitan invertir en recursos tecnológicos que beneficien los procesos de aprendizaje de sus estudiantes y cuando existen los recursos son insuficientes para el número de estudiantes.

Los datos que arrojó el simulacro se evaluaron en correspondencia con las preguntas que se categorizaron de acuerdo a las etapas de desarrollo del espacio proyectivo planteadas por Piaget. Los resultados indican que el 24% no presenta un desarrollo adecuado de la perspectiva de objetos diversos; el 40% no maneja adecuadamente las secciones de un objeto; al 75.6% le es difícil identificar la rotación y desarrollo de superficies al igual que al 82.1% no relaciona las secciones geométricas y la rotación y desarrollo de superficies. (Ver figura 31)

Nro.	pp1	pp2	pp3	pp4	P_TOTAL	RESULTADO	RESULTA %			DESEMP	DESEMP	PREDICIÓN	COINCIDE			
										ALTO	BAJO					
2019	1	3	5	2	0	10	BASICO	66,7	→	DESEMPEÑO_ALTO	0,660	0,340	DESEMPEÑO ALTO	SI		
	2	3	5	2	0	10	BASICO	66,7	→	DESEMPEÑO_ALTO	0,494	0,506	DESEMPEÑO BAJO	NO		
	3	3	5	0	0	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,432	0,568	DESEMPEÑO BAJO	NO		
	4	3	4	1	0	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,605	0,395	DESEMPEÑO ALTO	SI		
	5	3	2	0	0	5	BAJO	33,3	→	DESEMPEÑO_BAJO	0,267	0,733	DESEMPEÑO BAJO	SI		
	6	0	3	0	0	3	BAJO	20,0	→	DESEMPEÑO_BAJO	0,259	0,741	DESEMPEÑO BAJO	SI		
	7	3	3	1	0	7	BAJO	46,7	→	DESEMPEÑO_BAJO	0,679	0,321	DESEMPEÑO ALTO	NO		
	8	3	4	1	1	9	BAJO	60,0	→	DESEMPEÑO_ALTO	0,589	0,411	DESEMPEÑO ALTO	SI		
	9	0	0	0	0	0	BAJO	0,0	→	DESEMPEÑO_BAJO	0,529	0,471	DESEMPEÑO ALTO	NO		
	10	3	3	1	1	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,357	0,643	DESEMPEÑO BAJO	NO		
	11	3	3	1	1	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,616	0,384	DESEMPEÑO ALTO	SI		
	12	3	4	1	0	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,483	0,517	DESEMPEÑO BAJO	NO		
	13	3	3	1	0	7	BAJO	46,7	→	DESEMPEÑO_BAJO	0,494	0,506	DESEMPEÑO BAJO	SI		
	14	1	3	0	0	4	BAJO	26,7	→	DESEMPEÑO_BAJO	0,674	0,326	DESEMPEÑO ALTO	NO		
	15	3	5	1	0	9	BAJO	60,0	→	DESEMPEÑO_ALTO	0,534	0,466	DESEMPEÑO ALTO	SI		
	16	3	4	1	1	9	BAJO	60,0	→	DESEMPEÑO_ALTO	0,610	0,390	DESEMPEÑO ALTO	SI		
	17	0	1	2	0	3	BAJO	20,0	→	DESEMPEÑO_BAJO	0,604	0,396	DESEMPEÑO ALTO	NO		
	18	2	4	1	0	7	BAJO	46,7	→	DESEMPEÑO_BAJO	0,570	0,430	DESEMPEÑO ALTO	NO		
	19	2	3	1	2	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,617	0,383	DESEMPEÑO ALTO	SI		
	20	3	1	2	0	6	BAJO	40,0	→	DESEMPEÑO_BAJO	0,338	0,662	DESEMPEÑO BAJO	SI		
	21	3	3	1	0	7	BAJO	46,7	→	DESEMPEÑO_BAJO	0,759	0,241	DESEMPEÑO ALTO	NO		
	22	3	2	0	1	6	BAJO	40,0	→	DESEMPEÑO_BAJO	0,395	0,605	DESEMPEÑO BAJO	SI		
	23	2	1	0	2	5	BAJO	33,3	→	DESEMPEÑO_BAJO	0,699	0,301	DESEMPEÑO ALTO	NO		
	24	0	1	0	3	4	BAJO	26,7	→	DESEMPEÑO_BAJO	0,489	0,511	DESEMPEÑO BAJO	SI		
	25	2	4	1	0	7	BAJO	46,7	→	DESEMPEÑO_BAJO	0,428	0,572	DESEMPEÑO BAJO	SI		
	26	2	4	2	1	9	BAJO	60,0	→	DESEMPEÑO_ALTO	0,252	0,748	DESEMPEÑO BAJO	NO		
	27	2	4	1	0	7	BAJO	46,7	→	DESEMPEÑO_BAJO	0,553	0,447	DESEMPEÑO ALTO	NO		
	28	1	3	1	0	5	BAJO	33,3	→	DESEMPEÑO_BAJO	0,652	0,348	DESEMPEÑO ALTO	NO		
	29	1	0	1	2	4	BAJO	26,7	→	DESEMPEÑO_BAJO	0,467	0,533	DESEMPEÑO BAJO	SI		
	30	3	4	2	0	9	BAJO	60,0	→	DESEMPEÑO_ALTO	0,352	0,648	DESEMPEÑO BAJO	NO		
	31	3	2	0	1	6	BAJO	40,0	→	DESEMPEÑO_BAJO	0,434	0,566	DESEMPEÑO BAJO	SI		
	32	2	3	2	1	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,646	0,354	DESEMPEÑO ALTO	SI		
	33	3	4	1	1	9	BAJO	60,0	→	DESEMPEÑO_ALTO	0,715	0,285	DESEMPEÑO ALTO	SI		
	34	3	2	2	1	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,616	0,384	DESEMPEÑO ALTO	SI		
	35	3	1	1	1	6	BAJO	40,0	→	DESEMPEÑO_BAJO	0,291	0,709	DESEMPEÑO BAJO	SI		
	36	1	2	1	1	5	BAJO	33,3	→	DESEMPEÑO_BAJO	0,231	0,769	DESEMPEÑO BAJO	SI		
	37	2	2	1	0	5	BAJO	33,3	→	DESEMPEÑO_BAJO	0,652	0,348	DESEMPEÑO ALTO	NO		
	38	3	3	2	0	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,378	0,622	DESEMPEÑO BAJO	NO		
	39	3	4	1	0	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,541	0,459	DESEMPEÑO ALTO	SI		
	40	2	4	0	1	7	BAJO	46,7	→	DESEMPEÑO_BAJO	0,414	0,586	DESEMPEÑO BAJO	SI		
	41	2	5	1	0	8	BAJO	53,3	→	DESEMPEÑO_ALTO	0,471	0,529	DESEMPEÑO BAJO	NO		
	30	82	124	101		337		54,8		45,2						
	24	40,0	75,6	82,1		615		% preguntas perdidas		% ACIERTOS		DESEMPEÑO BAJO	22	56,098	COINCIDENCIA "SI"	23
	1 cat	2 cat	3 cat	4 cat								DESEMPEÑO BAJO	19	43,902	COINCIDENCIA "NO"	18
	Total de preguntas					337	BAJO	39			DESEMPEÑO ALTO		Porcentaje de COINCIDENCIA del algoritmo y el grupo comparador			
	Preguntas aprobadas					615	BASICO	2			DESEMPEÑO ALTO					
							ALTO	0			DESEMPEÑO ALTO					
							SUPERIOR	0			DESEMPEÑO ALTO					

Figura 31. Porcentaje de coincidencia 2019.

De otra parte, se observa en la Figura 31, el porcentaje de coincidencia de salida de la Simulación – 2019 con los resultados arrojados en la Prueba de Admisión Universidad Nacional 2010 (Anexo 1). Los estudiantes objeto de la muestra seleccionados de cada uno de los cursos presenta un 56% de coincidencia entre la simulación y la prueba de admisión, lo que equivale a que 23 de los resultados de la simulación coincidan con los resultados de la prueba de admisión.

De acuerdo con los resultados del simulacro, en términos generales, los estudiantes no han alcanzado el nivel de operaciones concretas que plantea Piaget y que les permite establecer

relaciones espaciales proyectivas, este es un asunto relevante que afecta el desempeño académico de los aprendices.

Los resultados que arrojó el algoritmo y que se contrastaron anteriormente con las pruebas planteadas en este estudio, indican que los estudiantes están en la total capacidad de desarrollar su pensamiento espacial a pesar de no contar con todos los factores favorables, siempre y cuando sean apoyados por procesos pedagógicos y didácticos que se los orienten.

Estos resultados se pueden tomar con la debida prudencia, como indicadores del bajo rendimiento de los estudiantes en las pruebas Saber, una probable explicación es que, de un lado, estos estudiantes carecen de condiciones socioeconómicas que les ayuden en sus procesos de desarrollo del conocimiento y de otro se requiere de mayor compromiso pedagógico en la orientación del proceso de aprendizaje.

Se considera que la orientación docente debe tener muy en cuenta los resultados del modelo predictivo para ajustar sus estrategias pedagógicas a los intereses, requerimientos y características individuales de los aprendices, de manera que sus estrategias permeen los factores pedagógicos del aprendizaje y beneficie directamente a cada estudiante; más aún si se tiene en cuenta que en nuestro país, los docentes presentan altos promedios de actualización y cualificación, igual que manejan variadas estrategias didácticas y han aprendido a manejar los ritmos y estilos de aprendizaje, lo que supondría de algún modo que la formación docente sumada a la posibilidad de contar con modelo predictivos garantiza la posibilidad de mejorar a futuro el desempeño de los aprendices.

8. Conclusiones y Recomendaciones

1. Una vez realizado el estudio y aplicado el algoritmo para dar respuesta a la pregunta planteada se puede evidenciar que en los procesos de aprendizaje existen factores que inciden directamente en el proceso de aprendizaje del estudiante y que no le permiten un desarrollo adecuado de dicho proceso.

2. Los factores asociados al aprendizaje son de diversa índole: familiares, institucionales o personales. Los factores socioeconómicos familiares, generalmente desaniman a los estudiantes a continuar estudios o los obligan a desertar para ir a trabajar y ayudar a su familia.

3. Los escasos o nulos recursos tecnológicos tanto de la institución como del hogar contribuyen a negar la posibilidad de un aprendizaje más efectivo; orientaciones pedagógicas deficientes o superficiales; entornos poco favorables que afectan el desempeño académico en algunos casos y en otros a desertar de su proceso de formación.

4. La aplicación de técnicas de Machine Learning y simulación para el desarrollo de un modelo que permita la predicción del desempeño de los estudiantes de Educación Básica y Media se constituyen en una herramienta eficaz para el docente, ya que pueden clasificar a los estudiantes y conocer con alto grado de precisión las categorías de DESEMPEÑO_BAJO y DESEMPEÑO_ALTO de los aprendices; esta es una ventaja que permite a los docentes crear estrategias en cada una de sus asignaturas para orientar al desarrollo del pensamiento espacial a partir de la Teoría del Desarrollo del Conocimiento Espacial de Piaget.

Estos resultados concuerdan con los datos y referencias bibliográficas de investigaciones de expertos como (Prabha & Shanavas, 2014) quienes manifiestan que la aplicación de estas técnicas permite el avance en el conocimiento científico sobre el aprendizaje y los aprendices,

permitiendo a los educadores seguir el progreso académico de los estudiantes y hacer seguimiento a sus debilidades y fortalezas.

Además del aporte bibliográfico de Prabha y Shanavas (2014) recopilan más de 240 artículos científicos relacionados con la aplicación de técnicas de Machine Learning en la educación, como lo manifiesta García (2016). Quién además agrega que el campo de la educación ha cobrado importancia en el área del Educational Data Mining donde su principal objetivo es aplicar métodos que permitan descubrir conocimiento en datos de ambientes educacionales para predecir e identificar el comportamiento futuro de los estudiantes.

Igual proponen que establecer dichas estrategias determina la capacidad de las instituciones de convertir datos sin procesar en inteligencia accionable con la aplicación de técnicas sofisticadas del área del Machine Learning convirtiendo esto en parte integral de la operación de las instituciones educativas.

5. El desarrollo de este trabajo destaca la importancia de la colaboración interdisciplinar en el desarrollo de este tipo de aplicaciones, ya que el docente se constituye en un agente determinante que conoce el contexto en el que se aplican este tipo de tecnologías y permite que el conocimiento y las aplicaciones técnicas cobren sentido dentro del proceso de aprendizaje. También se destaca el hecho de que los docentes usen la herramienta gestada por este intercambio como una vía que propicia el conocimiento entre diferentes disciplinas y facilita la respuesta oportuna a dificultades propias del área de tecnología; según Lawson (2015) la efectividad de las instituciones educativas en aplicaciones del Machine Learning se construye con el trabajo coordinado entre las directivas y docentes; el objetivo de las tecnologías Machine Learning es el de recopilar, extraer y limpiar los datos generados para mejorar los procesos

predictivos, lo mismo que diseñar estrategias pertinentes y crear espacios de asesoría e innovación.

6. Actualmente existen muchas aplicaciones de minería de datos en educación y trabajos realizados en diferentes partes del mundo que señalan que, la intervención de dichas tecnologías en el manejo de los múltiples datos de los estudiantes brinda soluciones que favorecen los procesos de enseñanza – aprendizaje y previenen situaciones que conllevan a la deserción del aprendiz. La comunidad científica como principal generadora de conocimiento, también recurre a herramientas tecnológicas que les permitan generar la suficiente información para simular fenómenos desconocidos a fin de predecir el comportamiento de cuerpos celestes o definir sus trayectorias para usarlas, por ejemplo, en un alunizaje o un satélite.

7. Los algoritmos, el Machine learning, las simulaciones, la extracción, procesamiento y almacenamiento de datos se configura en una metodología necesaria e imprescindible en el desarrollo de estrategias y conocimientos que promuevan y potencien las actividades en el aula de clase y todos los procesos pedagógicos que esto conlleva.

8. El desarrollo tecnológico no busca reemplazar los docentes, por el contrario, se suma a la ardua tarea de concebir nuevas formas, procesos, procedimientos, métodos, metodologías y estrategias que eleven el nivel de eficacia en el desarrollo de actividades de aprendizaje. La apropiación de la información se convierten en un insumo de mucho valor que sugiere sumar esfuerzos para su análisis y exploración afín de aplicarlo en ámbitos tales como: la detección de problemas de aprendizaje, el perfilamiento de estudiantes, el reconocimiento de estilos cognitivos, el diseño de estrategias de evaluación, sistemas de recomendación de contenidos académicos basados en el estilo cognitivo del estudiantes, el diseño y desarrollo de ambientes

virtuales de aprendizaje inteligentes, tutores inteligentes, el diseño de contenidos y material pedagógico y la predicción de rutas de aprendizaje entre otros.

8.1. Recomendaciones

Es fundamental que el docente se apoye en modelos predictivos que le ayuden a transformar su rol en el de orientador que el estudiante requiere para superar sus dificultades de aprendizaje y continuar con su proceso educativo, para lo cual es importante que se apropie de teorías pedagógicas – didácticas y herramientas tecnológicas que faciliten su labor.

Los grandes volúmenes de información que se generan día a día en el desarrollo de todo tipo de actividades pedagógicas demandan atención de los investigadores para explorar todos aquellos fenómenos que rodean los procesos de aprendizaje - enseñanza y crear nuevos contextos educativos e investigativos.

El uso de las nuevas tecnologías en el campo de la educación implica la adopción por parte de las instituciones y los docentes de la cultura de los datos, es importante que las instituciones definan repositorios que cataloguen y concentren toda la información que generan sus actividades académicas y fijen proyectos puntuales como los mencionados para empezar la exploración, análisis y modelamiento de la información para orientarla hacia el desarrollo de herramientas inteligentes.

Es de vital importancia continuar y promover la aplicación de estas técnicas en el desarrollo de herramientas e investigaciones aplicadas a la educación, dado que el continuo avance de la tecnología genera constantemente conocimientos que configuran nuevos caminos para entender los procesos de enseñanza – aprendizaje y genera valor y nuevos conocimientos y metodologías al docente para que realice una labor eficaz y productiva (Lawson, 2015)

Es importante reflexionar sobre cómo un sistema de predicción de dificultades académicas en los estudiantes podría minimizar el riesgo de fracaso, deserción escolar o mal desempeño; en este sentido, la aplicación de técnicas de machine learning puede gestar modelos predictivos que alerten a los padres, docentes antes de iniciar un periodo académico para que instituciones y hasta los mismos estudiantes puedan contar estas alertas que les permita tomar acciones oportunas y a los docentes un panorama detallado y perfilado de sus estudiantes a fin de asumir la estrategia pedagógica que el caso amerite.

Muchos de los procesos de educación virtual se caracterizan por enfocarse en el aprendizaje solo en usar el ambiente virtual de aprendizaje en la herramienta principal para la instalación de plataformas gestoras de contenidos educativos y los estudiantes por su parte, al utilizarlas siempre se encuentran generalizados en un compendio de materiales que en muchos de los casos no son de su interés, esto debido a que estas plataformas no tienen la capacidad de adaptarse a la necesidad que tienen tanto los docentes como los estudiantes lo que genera una limitante que puede ser superada en términos de un análisis de comportamiento de uso de estas herramientas por parte de los estudiantes y docentes para luego con la aplicación de técnicas machine learning podamos generar modelos que clasifiquen los estudiantes y les sugieran contenidos que puedan ser de su interés, esto podría llegar a atraer la atención de los estudiantes frente a los contenidos que el docente desarrolla, no es necesario el desarrollo de nuevas plataformas, solo basta con tomar los modelos de machine learning generados a partir de los datos analizados e integrarlos en los ambientes virtuales de aprendizaje ya existentes, esto aplicado de manera sistémica y ordenada puede potenciar de manera positiva el desarrollo de diferentes actividades académicas.

Muchas veces, la ausencia de datos trae consigo limitantes en términos de la aplicación de machine learning para la generación de modelos predictivos, no obstante, la simulación de

datos es un aliado estratégico que permite prototipar estas ideas de innovación en el aula; los datos simulados se pueden reemplazar progresivamente por datos reales con los que se vuelven a entrenar los modelos a fin de ir optimizando su poder predictivo; las simulaciones por sí solas facilitan la exploración de fenómenos desconocidos y permiten experimentar con las diferentes posibilidades que se puedan presentar.

Por último, las actividades que se desarrollan en el aula se pueden transformar con la aplicación sistemática de una metodología como la CRISP donde el Machine Learning se constituye en una aplicación que da respuesta a algunas preguntas de investigación, hace productivos este tipo de modelos y convierte las instituciones educativas en centros inteligentes de formación.

Referencias Bibliográficas

- Amadiou, F., Tricot, A., Gog, Paas, & Marine. (2009). Apprendre avec le numérique. Mythes et réalités. *RETZ*.
- Baker, D. (2015). Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Wiley periodicals inc*.
- Bazzan, A., Heinen, M. R., & Constantine, E. (2015). An Agent-Based Simulator for Intelligent Transportation Systems. *Research gate*.
- Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2016). Early Detection of Students at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods. *CEsifo Working Papers*.
- Boulesteix, A., Bender, A., Lorenzo, J., & Strobl, C. (05 de 2013). Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations. *BMC bioinformatics*, 13.
- Bravo, J., Ortega, M., & Prieto, M. (Enero - Abril de 2009). Entornos de Simulación en la Educación a Distancia. (E. U. Calatrava, Ed.) *Revista de Enseñanza Tecnológica*, 9.
- Castellanos, J. L. (2017). Using machine learning based on eye gaze to predict targets: An exploratory study. *Symposium Series on Computational Intelligence (SSCI)*.
- Castro, J. (2004). *El desarrollo de la noción del espacio en el niño de Educación inicial*. (U. d. Andes, Ed.) Táchira, Venezuela.
- Cataldi, Z., Dominighini, C., & Lage, F. J. (2013). Fundamentos para el Uso de Simulaciones en la Enseñanza. *Revista de Informática Educativa y Medios Audiovisuales*, 10, 12.

- Chang, Y. (2014). 3D-CAD effects on creative design performance of different spatial.
- CleverData. (2014). <https://cleverdata.io/que-es-machine-learning-big-data/>.
- Cohen, C., & Hegarty, M. (2014). Visualizing cross sections: Training spatial thinking using interactive animations and virtual objects. *Learning and Individual Differences* .
- Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *Springer*.
- Conati, C., & Mavrikis, M. (2018). AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling. *Cornell University*.
- Consejo Provincial de Educación. (1999). *Acerca de la Enseñanza del Espacio*. Provincia Río Negro, Argentina.
- Dicovski, L. M., & Pedroza, M. E. (2018). Mining Data, an Innovation of Quantitatives Research Methods, in the Measurement of University Academic Performance. *Revista Científica de FAREM-Estelí*, 10.
- Duque, C. (2009). *Geometría intuitiva desde el cuarto de baño*. Obtenido de www.sinewton.org: http://www.sinewton.org/numeros/numeros/70/Experaula_01.pdf
- Durán, E., Costaguta, R., Maldonado, M., Únzaga, S., Chequer, G., Menini, M., . . . Fernanadez, N. (2010). Técnicas de aprendizaje de máquina y personalización en educación. *Workshop de Investigadores en Ciencias de la Computación*.
- Fuentes, A. (2018). *Hands-On Predictive Analytics with Python*. Packt Publishing.

Galvez, G. (1994). La geometría, la psicogénesis de las nociones espaciales y la enseñanza de la geometría en la escuela elemental. En G. Galvez, *Didáctica de las matemáticas. Aportes y reflexiones*. Buenos Aires: Paidós. Educador.

Garbanzo Vargas, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista Educación*.

García, J. (05 de 2016). Líneas de investigación en minería de datos en aplicaciones en ciencia e ingeniería: Estado del arte y perspectivas. *Arxiv, Artificial Intelligence (cs.AI)*.

Gardner, H. (2011). *Inteligencias Múltiples, la Teoría en la Práctica*. España: Editorial Paidós.

Gardner, H. (2011). *Inteligencias Múltiples. La Teoría en la Práctica*. España: Ediciones Paidós.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc.

Giraldo, M. (2017). *Descubrimiento de patrones*. Bogotá: Universidad Nacional de Colombia.

González, J. (2017). *Pensamiento Espacial: Una experiencia de aula apoyada por realidad aumentada y Learning Catalytics, para el desarrollo de habilidades de razonamiento geométrico*. Bogotá: Universidad de los Andes.

Gutierrez, D. D. (2015). *Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R*. Technics Publications.

Hale, J. (12 de 10 de 2018). *Towards Data Science* . (Towards Data Science) Recuperado el 12 de 04 de 2019, de Towards Data Science : <https://towardsdatascience.com/smarter-ways-to-encode-categorical-data-for-machine-learning-part-1-of-3-6dca2f71b159>

- Hart, K., & Moore, G. (1971). *Desarrollo de la cognición espacial*. (P. P. Research, Ed.) Worcester, 7.
- Hederich, C. (2017). Estilística educativa un campo de investigación en educación y pedagogía. *Revista latinoamericana de estudios educativos*.
- Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*.
- ICFES. (2016). *Marco de Factores Asociados. Saber 31, 5º y 9º*. Bogotá.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to statistical learning with Applications in R* (Vol. 8). (Springer, Ed.) New York , Heidelberg Dordrecht London: Springer Texts in Statistics.
- Jensen, K. (2016). *file:///Users/oscardenendezaguirre/Downloads/ModelerCRISPDM.pdf*.
Obtenido de *file:///Users/oscardenendezaguirre/Downloads/ModelerCRISPDM.pdf*:
file:///Users/oscardenendezaguirre/Downloads/ModelerCRISPDM.pdf
- Kakade, S., & Ozdemir, S. (2018). *Principles of Data Science - Second Edition*. Packt Publishing.
- Lawson, J. (2015). *Data Science in Higher Education: A Step-by-Step Introduction to Machine Learning for Institutional Researchers*. CreateSpace Independent Publishing Platform.
- Martín, N. (2008). Development of a fast remedial course to improve the spatial abilities of engineering students. *Journal of Engineering Education*, 505-513.
- Martin, O. (2018). *Bayesian Analysis with Python - Second Edition*. Packt Publishing.

- Matemático, O. c. (2007). Orientaciones curriculares para el campo de Pensamiento Matemático. En S. d. Distrital, *Orientaciones curriculares para el campo*. Bogota. Distrito capital: Secretaria de educación.
- Mavrikis. (2010). *La Inteligencia Artificial en la Educación Necesita Aprendizaje Automático Interpretable*.
- Mc Cormick, K., & Salcedo, J. (2017). *IBM SPSS Modeler Essentials*. USA: Packt Publishing.
- MEN. (1998). *Lineamientos curriculares para el área de matemáticas*. Bogotá. Distrito capital.: Creamos alternativas.
- MEN. (2006). *Estandares básicos de competencias en matemáticas*. Bogotá: Ministerio de educación nacional.
- Mitchell, T., Chen, S., & Macredie, R. (2004). Adapting Hypermedia to Cognitive Styles: Is it necessary? *Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK*.
- Moncayo, C. M. (2016). Determinantes que influyen en el Rendimiento Académico: Un Estudio Aplicado para Colombia a Partir de las Pruebas Saber 11 - 2014. Bogotá, Colombia.
- Morales, C. A. (2012). El desarrollo del pensamiento espacial y la competencia matemática. Una aproximación desde el estudio de los cuadriláteros. *Revista Amazonia Investiga / Florencia, Colombia*.
- Navarro, R. (2009). La capacidad espacial y su relación con la ingeniería. *DYNA*.
- Olaya, C., & Diaz, C. C. (2017). *Social Systems Engineering: The Design of Complexity*. Wiley.

- Ott, M., & Pozzi, F. (2012). Digital games as creativity enablers for children. *Behaviour & Information Technology*.
- Pérez, D., & González, C. (2007). *Minería de Datos. Técnicas y Herramientas*. B & N.
- Phillips, L., Norris, S., & Macnab, J. (2010). Visualization in mathematics, reading and science education. *Springer*.
- Piaget, J. (1947). *La representación del Espacio en el Niño*. Paris: PUF.
- Piaget, J. (1948). *La Geometría Espontánea en el Niño*. Paris: PUF.
- Piaget, J., & Inhelder, B. (1997). *Psicología del niño*. Madrid: Morata.
- Prabha, L., & Shanavas, M. (07 de 2014). EDUCATIONAL DATA MINING APPLICATIONS. *Operations Research and Applications: An International Journal (ORAJ)*, 1.
- Rainey, R., & Talk. (2012). FTC Final Privacy Report Draws a Map to Meaningful Privacy Protection in the Online World. *ELECTRONIC FRONTIER FOUND*.
- Ramírez. (19 de 06 de 2018). *medium.com*. (medium.com) Recuperado el 12 de 04 de 2019, de medium.com: <https://medium.com/bluekiri/curvas-pr-y-roc-1489fbd9a527>
- Ramírez, F., Hernández, J., & Ramírez, M. J. (2004). *Introducción a la minería de datos*. Pearson Educación.
- Rice, M. &. (10 de 2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, 29.

- Salas. (2005). Hallazgos de la investigación sobre la inserción de las Tecnología de la información y la comunicación (TIC) en la enseñanza: la experiencia de los últimos diez años en los Estados Unidos. *Revista Educación.*, 29(2), 53-66.
- Salas, R. P., & Zuleta, P. A. (1995). La Simulación como Método de Enseñanza y Aprendizaje. Educación Médica Superior. *Educ Med Super v.9 n.1 Ciudad de la Habana*, 72.
- Samsudin, K., Rafi, A., & Hanif, A. (2011). *Training in Mental Rotation and Spatial*.
- Sharma, T., Bali, R., & Sarkar, D. (2017). *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*. Apress.
- Sison, R., & Shimura, M. (1998). Student Modeling and Machine Learning. *International Journal of Artificial Intelligence in Education (IJAIED)*.
- Sposito, O., Etcheverry, M., Ryckeboer, H., & Bossero, J. (2008). *Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil*. San Justo – Provincia de Buenos Aires – Argentina: Universidad Nacional de La Matanza.
- Tamoghna, Raghav, B., & Dipanjan, S. (2018). *Hands-On Transfer Learning with Python*. Packt Publishing.
- Thanaki, J. (2018). *Machine Learning Solutions* . Packt Publishing.
- Timarán, R., & Jiménez, J. (2014). Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM. *Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación*.

- UNMC Nebraska Medicine. (7 de 3 de 2001). <https://www.unmc.edu/>. Obtenido de www.unmc.edu: <http://gim.unmc.edu/dxtests/roc3.htm>
- Van der laan, M. J. (2011). Targeted learning. *Springer*.
- Vasco, C. E. (2006). *Sistemas geométricos. Un nuevo enfoque para la didáctica de las matemáticas*. Bogota. Distrito capital.: Ministerio de Educacion Nacional.
- Vázquez, S., & Noriega, M. (2011). Razonamiento espacial y rendimiento académico. *Interdisciplinaria. Revista de psicología y ciencias afines*.
- Vera, Y. (2003). *El uso de la Internet como herramienta educativa y su relación en el rendimiento cualitativo de los alumnos del sexto grado de Educación Básica*. Maracaibo, Venezuela.: Trabajo de Grado, Universidad Rafael Beloso Chacín.
- Wirth, R. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.

Anexos

Anexo 1. Facsímil Prueba de Admisión Universidad Nacional 2010

AVISO LEGAL: ESTE DOCUMENTO NO REPRESENTA LA OPINIÓN O CONSENTIMIENTO DE LA UNIVERSIDAD NACIONAL DE COLOMBIA Y NO PUEDE SER EMPLEADO POR NINGUNA RAZÓN CON FINES COMERCIALES.



UNIVERSIDAD NACIONAL DE COLOMBIA
VICERRECTORÍA ACADÉMICA
DIRECCIÓN NACIONAL DE ADMISIONES

PRUEBA DE ADMISIÓN

SEGUNDO SEMESTRE DE 2010



PROGRAMAS DE PREGRADO

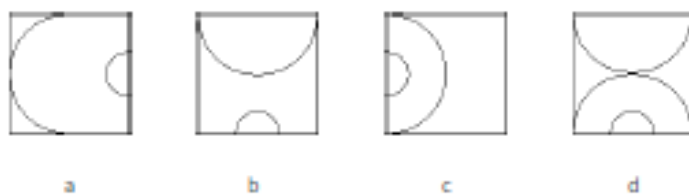
- ARTES
- INGENIERÍAS
- CIENCIAS DE LA SALUD
- CIENCIAS SOCIALES
- CIENCIAS BÁSICAS
- CIENCIAS AGROPECUARIAS

8 DE MAYO DE 2010

ANÁLISIS DE LA IMAGEN

Preguntas 106 a 120

Dadas las siguientes figuras, identifique el orden en que fueron colocadas para obtener la imagen de las preguntas 106 a 108.



106.

- A. b - d - c - a
 B. d - c - a - b
 C. b - c - d - a
 D. a - d - c - b



107.

- A. a - b - c - d
 B. b - d - a - c
 C. d - c - b - a
 D. d - b - c - a



108.

- A. c - b - d - a
 B. d - c - a - b
 C. c - d - b - a
 D. d - a - c - b



En las preguntas 109 a 113 identificar la figura que NO corresponde a la estructura organizativa de la izquierda

109.



A.



B.



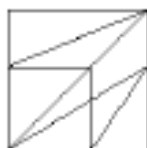
C.



D.



110.



A.



B.



C.



D.



111.



A.



B.



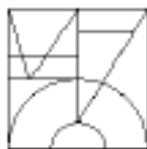
C.



D.



112.



A.



B.



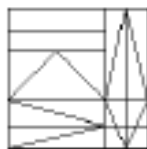
C.



D.



113.



A.



B.



C.



D.

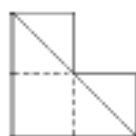


En las figuras 114 a 117 identifique la vista que corresponde a la dirección que indica la flecha en el dibujo de la izquierda

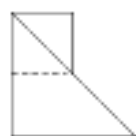
114.



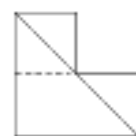
A.



B.



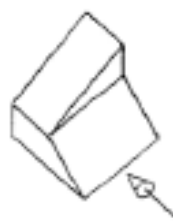
C.



D.



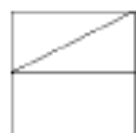
115.



A.



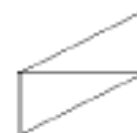
B.



C.



D.



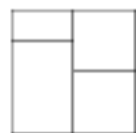
116.



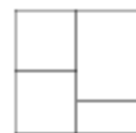
A.



B.



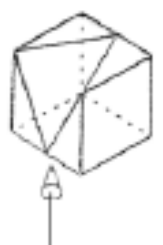
C.



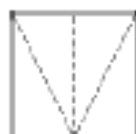
D.



117.



A.



B.



C.

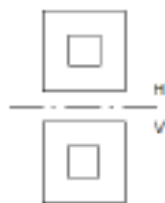


D.



En las preguntas 118 a 120 identifique la vista de perfil que NO corresponde a las proyecciones del plano horizontal (H) y del plano vertical (V) que se muestran a la izquierda.

118.



A.



B.



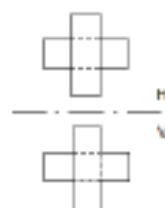
C.



D.



119.



A.



B.



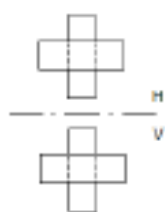
C.



D.



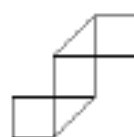
120.



A.



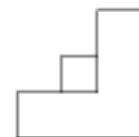
B.



C.



D.



FIN

Anexo 2. Diccionario de Variables MEN - ICFES

Nro.	VARIABLES	DESCRIPCIÓN
1	APELLIDOS Y NOMBRES	Nombre del estudiante
2	TIPO_DOCUMENTO	Tipo de documento del estudiante
3	NRO_DOCUMENTO	Numero de documento del estudiante
4	NACIONALIDAD	Nacionalidad del estudiante
5	GENERO	Género del estudiante: Femenino o masculino
6	GRADO	Grado del estudiante
7	PAIS_RECIDE_ACTUAL	País de residencia del estudiante
8	FECHA_NACIMIENTO	Fecha de Nacimiento del estudiante
9	DEPART_NACE	Departamento de nacimiento
10	MUNICIP_NACE	Municipio de nacimiento
11	SABER_3	¿Ha presentado la prueba SABER 3? (En tercer grado)
12	SABER_5	¿Ha presentado la prueba SABER 5? (En quinto grado)
13	SABER_9	¿Ha presentado la prueba SABER 9? (En grado noveno)
14	SABER_PRO_11	¿Ha presentado la prueba SABER 11? (En grado once)
15	SIMULACROS	¿Ha presentado simulacros?
16	ETNICO_MINORITARIO	¿Cuál es el grupo étnico minoritario al que pertenece?
17	ETNIA_PERTENECE	
18	DISCAPACIDAD	Posee alguna discapacidad
19	DEPTO_RESIDE	Departamento de residencia
20	DANE_RESIDE_DEPTO	Código Dane del departamento de residencia
21	MCPIO_RESIDE	Municipio de Residencia
22	DANE_RESIDE_MCPIO	Código Dane del municipio de residencia
23	AREA_RESIDENCIA	Área de residencia: Rural o urbana
24	NOMBRE_URBANO	Nombre del área si es urbana
25	NOMBRE_RURAL	Nombre del área si es rural
26	NIVEL_EDUCATIVO_PADRE	Nivel educativo más alto alcanzado por el padre
27	NIVEL_EDUCATIVO_MADRE	Nivel educativo más alto alcanzado por la madre
28	ESTRATO	Estrato socioeconómico de su vivienda según recibo de energía eléctrica
29	SISBEN	Número del puntaje del Sisbén.
30	PERSONA_CONFORMA_HOGAR	¿Cuántas personas conforman el hogar donde vive actualmente, incluido usted?
31	NRO_HERMANAS	¿Cuántas hermanas tiene?
32	NRO_HERMANOS	¿Cuántos hermanos tiene?
33	MATERIAL_PISOS	¿Cuál es el material de los pisos que predomina en su vivienda?
34	HABITACION_DUERME_HOGAR	En total, ¿en cuántos cuartos duermen las personas de su hogar?

Nro.	VARIABLES	DESCRIPCIÓN
35	HOGAR_INTERNET	¿Su hogar cuenta con servicio o conexión a internet?
36	HOGAR_SERVICIO_TV	¿Su hogar cuenta con servicio cerrado de televisión?
37	H_PC	¿Posee Computador en su hogar?
38	H_LAVADORA	¿Posee Máquina Lavadora en su hogar?
39	H_HORNO_MICROO	¿Posee Horno Microondas en su hogar?
40	H_HORNO_ELECT	¿Posee Horno Eléctrico en su hogar?
41	H_HORNO_GAS	¿Posee Horno a Gas en su hogar?
42	H_DVD	¿Posee DVD en su hogar?
43	HOGAR_NEVERA	¿Posee Nevera en su hogar?
44	H_HOGAR_CELULAR	¿Cuenta con Celular en casa?
45	H_PLAY	¿Posee PlayStation en su hogar?
46	H_XBOX	¿Posee Xbox en su hogar?
47	H_NINTENDO	¿Posee Nintendo su hogar?
48	H_BICI	¿Posee Bicicleta en su hogar?
49	H_CEL_PERS	¿Posee Celular Personal?
50	H_MOTO	¿Posee Motocicleta en su hogar?
51	H_AUTO_PARTIC	¿Posee Automóvil Particular en su hogar?
52	H_AUTO_PUBLIC	¿Posee Automóvil de uso público en su hogar?
53	H_TAXI	¿Posee Taxi en su hogar?
54	HOGAR_LIBROS	Excluyendo periódicos, revistas, directorios telefónicos y libros del colegio ¿cuántos libros físicos o electrónicos hay en su hogar?
55	HOGAR_INGRESOS_MES	¿Cuál es el total de ingresos mensuales de su hogar, en términos de salarios mínimos (SMMLV)?
56	H_COME_LECHE	¿Cuántas veces por semana toman leche en su hogar?
57	H_COME_QUESO	¿Cuántas veces por semana se come queso en su hogar?
58	H_COME_YOGURT	¿Cuántas veces por semana se toma Yogurt en su hogar?
59	H_COME_CERDO	¿Cuántas veces por semana se come cerdo en su hogar?
60	H_COME_POLLO	¿Cuántas veces por semana se come pollo en su hogar?
61	H_COME_PAVO	¿Cuántas veces por semana se come pavo en su hogar?
62	H_COME_RES	¿Cuántas veces por semana se come carne de res en su hogar?
63	H_COME_CORDERO	¿Cuántas veces por semana se come cordero en su hogar?
64	H_COME_CONEJO	¿Cuántas veces por semana se come conejo en su hogar?

Nro.	VARIABLES	DESCRIPCIÓN
65	H_COME_PESCADO	¿Cuántas veces por semana se come pescado en su hogar?
66	H_COME_HUEVO	¿Cuántas veces por semana se come huevo en su hogar?
67	H_COME_AVENA	¿Cuántas veces por semana se consume avena en su hogar?
68	H_COME_CEREAL	¿Cuántas veces por semana se consumen cereales en su hogar?
69	H_COME_ALMENDRAS	¿Cuántas veces por semana se consumen almendras en su hogar?
70	H_COME_MANI	¿Cuántas veces por semana se come Maní en su hogar?
71	H_COME_FRIJOLES	¿Cuántas veces por semana se come fríjoles en su hogar?
72	H_COME_LENTEJAS	¿Cuántas veces por semana se come lentejas en su hogar?
73	H_COME_GARBANZOS	¿Cuántas veces por semana se come garbanzos en su hogar?
74	LABOR_PADRE	Ocupación u oficio del padre
75	LABOR_MADRE	Ocupación u oficio de la madre
76	LABOR_ACUDIENTE	Ocupación u oficio del acudiente
77	SITUACIÓN_HOGAR	Con respecto al año inmediatamente anterior, la situación económica de su hogar es: igual, mejor, peor.
78	TIEMPO_DEDICA_LEER	Usualmente, ¿cuánto tiempo al día dedica a leer por entretenimiento?
79	TIEMPO_DEDICA_NAVEGAR	Usualmente, ¿cuánto tiempo al día dedica a navegar en internet? Excluya actividades académicas.
80	TRABAJO_SEMANA_PAST	¿Cuántas horas trabajó usted durante la semana pasada?
81	REMUNERACIÓN_TRABAJO	¿Usted recibe algún tipo de remuneración por trabajar?

Anexo 3. Encuesta de factores sociodemográficos.

INFORMACIÓN PERSONAL											
1	Nombres				Apellidos						
2	Tipo de documento del estudiante			C.C.	T.I.	Nro.					
3	Nacionalidad del estudiante				Genero		Masculino	Femenino	Grado		
4	País donde reside actualmente				Fecha de Nacimiento			DIA	MES	AÑO	
5	Departamento de nacimiento				Municipio de nacimiento						
6	Ha presentado pruebas SABER				Saber 3	Saber Quinto	Saber Noveno	Saber Pro 11	Simulacros		
7	¿Pertenece usted a un grupo étnico minoritario?				Mestizos	Mulatos	Criollos	Negros	Indios		
8	Etnia a la cual pertenece el estudiante en caso de que tenga				Arhuaco	Cancuamo	Tucano	Zenú			
	Sikuaní	Pijao	Inga	Páez	Emberá	Afrodescendiente	Palenquero	Wayúu	Gitana	Cubeo	
	Raízal	Pasto	Huitoto	Gua mbiano	Ninguno	Otro grupo	Especifique el nombre de la etnia				
9	Tiene alguna discapacidad				Motriz	Invidente	Condición especial	Sordo	Síndrome de down	Autismo	
INFORMACIÓN DE CONTACTO											
10	Departamento de residencia				Codigo DANE		Municipio		Codigo DANE		
11	Área de residencia		Urbano	Nombre del barrio			Rural	Nombre de la vereda			
INFORMACIÓN SOCIOECONÓMICA											
12	Nivel educativo más alto del padre				Ninguno	Primaria completa	Primaria incompleta	Bachillerato completo			
	Bachillerato incompleto				Técnica o tecnológica incompleta		Técnica o tecnológica completa		Profesional incompleta		
	Profesional completa		Posgrado	Maestría	Doctorado		Otro especifique				
13	Nivel educativo más alto de la madre				Ninguno	Primaria completa	Primaria incompleta	Bachillerato completo			
	Bachillerato incompleto				Técnica o tecnológica incompleta		Técnica o tecnológica completa		Profesional incompleta		
	Profesional completa		Posgrado	Maestría	Doctorado		Otro especifique				
14	Estrato socioeconómico de su vivienda según recibo de energía eléctrica:						1	2	3	4	
15	Puntaje de SISBEN en el que está clasificada su familia				Nivel 1	Nivel 2	Nivel 3	Otro nivel	No esta clasificada		
16	¿Cuántas personas conforman el hogar donde vive actualmente, incluido usted?										
17	¿Cuántas HERMANAS tiene en total?				¿Cuántos HERMANOS tiene en total?						
18	¿Cuál es el material de los pisos que predomina en su vivienda?				Cemento	Gravilla	Ladrillo	Alfombra			
	Madera burda		Madera pulida		Baldosa	Tabla o tablón		Mármol	Tierra	Arena	
19	En total, ¿en cuántas habitaciones duermen las personas en su hogar?										
20	¿Su hogar cuenta con?				Conexión a internet			Servicio cerrado de televisión			
21	¿Cuáles de los siguientes bienes posee su hogar?:				Computador	Lavadora	Horno microondas	Horno eléctrico	Horno a gas		
	DVD		Nevera	Celular en casa	Play station		Xbox	Nintendo	Bicicleta		
	Celular personal		Motocicleta	Automovil de uso particular		Automovil de uso publico		Taxi			
22	¿Cuántos libros físicos o electrónicos hay en su hogar excluyendo periódicos, revistas, directorios telefónicos y libros del colegio?				De 1 a 5		De 1 a 10	De 1 a 20	De 1 a 30		
23	¿Cuál es el total de ingresos mensuales de su hogar?				Menos de \$ 828,116		\$828.116	\$1.656.232	\$2.484.348		
24	¿Cuántas veces por semana comen estos alimentos en su hogar?				Leche		1-2-3-4-5-6-7		Queso		
	Yogurt 1-2-3-4-5-6-7		Cerdo 1-2-3-4-5-6-7		Pollo 1-2-3-4-5-6-7		Pavo 1-2-3-4-5-6-7		Res 1-2-3-4-5-6-7		
	Cordero 1-2-3-4-5-6-7		Conejo 1-2-3-4-5-6-7		Pescado 1-2-3-4-5-6-7		Huevo 1-2-3-4-5-6-7				
25	¿Cuántas veces por semana comen estos alimentos en su hogar?				Avena		1-2-3-4-5-6-7		Cereales 1-2-3-4-5-6-7		
	Almendras 1-2-3-4-5-6-7		Mani 1-2-3-4-5-6-7		Frijoles 1-2-3-4-5-6-7		Lentejas 1-2-3-4-5-6-7		Garbanzos 1-2-3-4-5-6-7		
26	Señale aquella labor o trabajo que realizó su padre durante la mayor parte del último año:										
27	Señale aquella labor o trabajo que realizó su madre durante la mayor parte del último año:										
28	Señale aquella labor o trabajo que realizó su acudiente durante la mayor parte del último año:										
29	Con respecto al año anterior, la situación económica de su hogar es				Muy buena	Buena	Regular	Insuficiente			
30	Usualmente, ¿cuánto tiempo al día dedica a leer por entretenimiento?				1/2 hora		1 hora	2 horas	3 horas		
31	Usualmente, ¿cuánto tiempo al día dedica a navegar en internet?				1/2 hora	1 hora	2 horas	3 horas	4 horas		
32	¿Cuántas horas trabajó durante la semana pasada?				1 hora	2 hora	3 horas	4 horas	5 horas	6 horas	
33	¿Usted recibe algún tipo de remuneración por trabajar?				3 mil	5 mil	10 mil	15 mil	20 mil	25 mil	

Anexo 4. Salida de la Simulación para grupo comparador - 2019

No	DESEMPEÑO_ALTO	DESEMPEÑO_BAJO	PREDICTION
1	0.660024701370156	0.339975298629843	DESEMPEÑO_ALTO
2	0.493829187586022	0.506170812413977	DESEMPEÑO_BAJO
3	0.432395396256957	0.567604603743042	DESEMPEÑO_BAJO
4	0.604962701487575	0.395037298512424	DESEMPEÑO_ALTO
5	0.266532060662326	0.733467939337673	DESEMPEÑO_BAJO
6	0.25925631793271	0.74074368206729	DESEMPEÑO_BAJO
7	0.678634943702626	0.321365056297373	DESEMPEÑO_ALTO
8	0.589109351705634	0.410890648294365	DESEMPEÑO_ALTO
9	0.528723721658217	0.471276278341782	DESEMPEÑO_ALTO
10	0.356903311977864	0.643096688022135	DESEMPEÑO_BAJO
11	0.615624561626286	0.384375438373713	DESEMPEÑO_ALTO
12	0.483324922526222	0.516675077473777	DESEMPEÑO_BAJO
13	0.493829187586022	0.506170812413977	DESEMPEÑO_BAJO
14	0.673564438247135	0.326435561752864	DESEMPEÑO_ALTO
15	0.534306921380225	0.465693078619774	DESEMPEÑO_ALTO
16	0.610255962021231	0.389744037978768	DESEMPEÑO_ALTO
17	0.603773845794992	0.396226154205007	DESEMPEÑO_ALTO
18	0.569805255513094	0.430194744486905	DESEMPEÑO_ALTO
19	0.617166438218591	0.382833561781408	DESEMPEÑO_ALTO
20	0.337832763771719	0.66216723622828	DESEMPEÑO_BAJO
21	0.759459781360571	0.240540218639428	DESEMPEÑO_ALTO

No	DESEMPEÑO_ALTO	DESEMPEÑO_BAJO	PREDICTION
22	0.394873742691963	0.605126257308036	DESEMPEÑO_BAJO
23	0.69917828359153	0.300821716408469	DESEMPEÑO_ALTO
24	0.489060606153219	0.51093939384678	DESEMPEÑO_BAJO
25	0.427677473646176	0.572322526353823	DESEMPEÑO_BAJO
26	0.252307467266515	0.747692532733484	DESEMPEÑO_BAJO
27	0.552616684572955	0.447383315427044	DESEMPEÑO_ALTO
28	0.652446178998759	0.34755382100124	DESEMPEÑO_ALTO
29	0.466892526402838	0.533107473597161	DESEMPEÑO_BAJO
30	0.352076702451075	0.647923297548924	DESEMPEÑO_BAJO
31	0.433665695684753	0.566334304315246	DESEMPEÑO_BAJO
32	0.645808678052515	0.354191321947485	DESEMPEÑO_ALTO
33	0.715319820982095	0.284680179017904	DESEMPEÑO_ALTO
34	0.615624561626286	0.384375438373713	DESEMPEÑO_ALTO
35	0.290580992080344	0.709419007919655	DESEMPEÑO_BAJO
36	0.231343320338925	0.768656679661074	DESEMPEÑO_BAJO
37	0.652446178998759	0.34755382100124	DESEMPEÑO_ALTO
38	0.377809099493105	0.622190900506894	DESEMPEÑO_BAJO
39	0.541420974260034	0.458579025739966	DESEMPEÑO_ALTO
40	0.413785932273932	0.586214067726067	DESEMPEÑO_BAJO
41	0.470775218247379	0.52922478175262	DESEMPEÑO_BAJO

Anexo 5. Documentación del algoritmo

https://gitlab.com/oscfrayle/metiae_notebooks