



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

1 de 1

Neiva, 19 de enero de 2023

Señores

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD SURCOLOMBIANA

Neiva, Huila

El (Los) suscrito(s):

Manuel Antonio Hoyos García, con C.C. No. 1'117.513.191 de Florencia-Caquetá.

Víctor Andrés Castro Dueñas, con C.C. No. 12201781 de Garzón-Huila.

Autor(es) de la tesis y/o trabajo de grado: Manuel Antonio Hoyos García y Víctor Andrés Castro Dueñas

Titulado PREDICCIÓN DE DIFICULTADES EN COMPETENCIAS BÁSICAS, USANDO MÉTODOS MACHINE LEARNING, CON ESTUDIANTES DE DÉCIMO GRADO presentado y aprobado en el año 2022 como requisito para optar al título de Magister en Estudios Interdisciplinarios de la Complejidad;

Autorizo (amos) al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales "open access" y en las redes de información con las cuales tenga convenio la Institución.
- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.
- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, "Los derechos morales sobre el trabajo son propiedad de los autores", los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

EL AUTOR/ESTUDIANTE:

Firma:

EL AUTOR/ESTUDIANTE:

Firma:

Vigilada Mineducación



TÍTULO COMPLETO DEL TRABAJO: PREDICCIÓN DE DIFICULTADES EN COMPETENCIAS BÁSICAS, USANDO MÉTODOS MACHINE LEARNING, CON ESTUDIANTES DE DÉCIMO GRADO

AUTOR O AUTORES:

Primero y Segundo Apellido	Primero y Segundo Nombre
HOYOS GARCIA	MANUEL ANTONIO
CASTRO DUEÑAS	VICTOR ANDRÉS

DIRECTOR Y CODIRECTOR TESIS:

Primero y Segundo Apellido	Primero y Segundo Nombre
MARTINEZ MONCALEANO	CARLOS JAVIER

ASESOR (ES):

Primero y Segundo Apellido	Primero y Segundo Nombre
MARTINEZ MONCALEANO	CARLOS JAVIER

PARA OPTAR AL TÍTULO DE: Magíster en Estudios Interdisciplinarios de la Complejidad

FACULTAD: Ciencias Exactas y Naturales

PROGRAMA O POSGRADO: Maestría en Estudios Interdisciplinarios de la Complejidad

CIUDAD: Neiva

AÑO DE PRESENTACIÓN: 2022

NÚMERO DE PÁGINAS: 148

TIPO DE ILUSTRACIONES (Marcar con una X):

Diagramas_x_ Fotografías_x_ Grabaciones en discos___ Ilustraciones en general_x_ Grabados___
Láminas___ Litografías___ Mapas_x_ Música impresa___ Planos___ Retratos___ Sin ilustraciones___ Tablas
o Cuadros_x_

SOFTWARE requerido y/o especializado para la lectura del documento: Ninguno especial

MATERIAL ANEXO: CD

Vigilada Mineducación

La versión vigente y controlada de este documento, solo podrá ser consultada a través del sitio web Institucional www.usco.edu.co, link Sistema Gestión de Calidad. La copia o impresión diferente a la publicada, será considerada como documento no controlado y su uso indebido no es de responsabilidad de la Universidad Surcolombiana.



PREMIO O DISTINCIÓN (En caso de ser LAUREADAS o Meritoria):

PALABRAS CLAVES EN ESPAÑOL E INGLÉS:

<u>Español</u>	<u>Inglés</u>
1. Complejidad	Complexity
2. Predicciones	Predictions
3. Educación	Education
4. Sistemas inteligentes	Intelligent systems
5. Aprendizaje automático	Machine Learning
6. Interrelaciones	Interrelationships

RESUMEN DEL CONTENIDO: (Máximo 250 palabras)

Valorando el potencial que yace en los datos o la información y considerando el sistema educativo como un complejo constructo de interrelaciones en el que se pueden enmarcar otros sistemas que influyen y múltiples actores que activamente intervienen, se propuso hacer uso de los sistemas inteligentes al interior del campo de la educación para buscar interrelacionar características de singularidad de los agentes principales del sistema educativo: los estudiantes. En ese sentido, la investigación se estructuró como un posible modelo que tiene como objetivo establecer relaciones entre variables sociodemográficas y de desempeños para predecir dificultades en el desarrollo de competencias básicas de estudiantes de décimo grado en dos instituciones del departamento del Huila.

De acuerdo con la complejidad inherente al propósito del trabajo, se integró un algoritmo machine learning que promoviera la posibilidad de encontrar y entender en los centros educativos, las posibles interrelaciones que conllevaron de manera concluyente a posibles predicciones en los desempeños de los estudiantes objeto de estudio. En consecuencia, se logra una caracterización apropiada de las singularidades de los estudiantes, también, una comparación entre los rendimientos de distintos algoritmos de clasificación y la selección del algoritmo *decisiontreeclassifier* como el más pertinente de acuerdo a la naturaleza del trabajo. Por último, se dejó un referente para abordar los problemas en el campo educativo de manera integradora y no a través de la usual segmentación de los mismos, así, se logró suscitar un cambio en las prácticas pedagógicas de docentes para considerar la mejora en el desarrollo de competencias.

ABSTRACT: (Máximo 250 palabras)

Valuing the potential that lies in the data or information and considering the educational system as a complex construct of interrelationships in which other systems that influence and multiple actors that actively intervene can be framed, it was proposed to use intelligent systems within the field of education to seek to interrelate unique characteristics of the main agents of the educational system: students. In this sense, the research was structured as a possible model which has as purpose to establish relationships between sociodemographic and performance variables to predict difficulties in the development of basic skills of tenth grade students in two institutions in the department of Huila.



In accordance with the complexity inherent to the purpose of the research, a machine learning algorithm was integrated that would promote the possibility of finding and understanding in educational centers, the possible interrelationships that conclusively led to possible predictions in the performance of the students under study. Consequently, an appropriate characterization of the singularities of the students is achieved, as well as a comparison between the performance of different classification algorithms and the selection of the *decisiontreeclassifier* algorithm as the most relevant according to the nature of the work.

Finally, a reference was left to treat the problems in the educational field in an integrative way and not through the usual segmentation of them, thus, it was possible to provoke a change in the pedagogical practices of teachers to consider the improvement in the development competences.

APROBACION DE LA TESIS

Nombre Presidente Jurado:

Mag. en Estudios Interdisciplinarios de la Complejidad Edinson Oswaldo Delgado Rivas

Firma:

Nombre Jurado:

Ph.D. en Psicología con Orientación en Neurociencias Cognitivas Alfredis Gonzalez Hernandez

Firma:

PREDICCIÓN DE DIFICULTADES EN COMPETENCIAS
BÁSICAS, USANDO MÉTODOS *MACHINE LEARNING*,
CON ESTUDIANTES DE DÉCIMO GRADO DE LAS
INSTITUCIONES EDUCATIVAS RAMÓN ALVARADO
SÁNCHEZ Y LA MERCED

UNIVERSIDAD SURCOLOMBIANA
Maestría en estudios Interdisciplinarios de la Complejidad
MANUEL ANTONIO HOYOS GARCIA
VICTOR ANDRÉS CASTRO DUEÑAS

CONTENIDO

1. INTRODUCCIÓN.....	7
2. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN	9
2.1 Descripción del Problema	9
2.2 Sistematización del Problema	11
2.3 Enunciación del Problema	11
3. ANTECEDENTES Y JUSTIFICACIÓN	13
3.1 Antecedentes	13
3.2 Justificación	20
4. FUNDAMENTOS TEÓRICOS	23
4.1 Referentes contextual e Institucional.	23
4.2 Marco Teórico.	25
4.2.1 Complejidad, Educación y Sistemas inteligentes	26
4.2.1.1 Complejidad y Educación	28
4.2.1.2 Sistemas inteligentes	30
4.2.1.3 La era de los datos, la información y el <i>machine learning</i>	32
4.2.1.4 Herramientas <i>machine learning</i> aplicadas a la educación	38
4.2.1.5 Algoritmos de clasificación y medición de incertidumbre	41
4.2.2 Importancia y dificultades en la enseñanza-aprendizaje por competencias.	53
4.2.2.1 Competencias básicas, genéricas y específicas	54
4.2.2.2 Importancia de la formación por competencias básicas	56
4.2.2.3 Dificultades en la formación por competencias básicas	58
4.2.2.4 Evaluaciones estandarizadas como sistema de medición de competencias	59
4.2.3 Aspectos socio-demográficos y de desempeño que caracterizan a los estudiantes	62
4.2.3.1 Entorno social, familiar, educativo y su influencia en los estudiantes	63

4.2.3.2 Aspectos específicos del entorno que podrían influir en los desempeños	64
5. OBJETIVOS DE LA INVESTIGACIÓN	68
5.1 Objetivo general	68
5.2 Objetivos específicos	68
6. METODOLOGÍA.....	69
6.1 Tipo y enfoque de la investigación	69
6.2 Universo de estudio, población y muestra	70
6.3 Estrategias Metodológicas	72
6.4 Técnicas e instrumentos de Investigación	75
7. ANÁLISIS DE RESULTADOS	78
8. CONCLUSIONES.....	124
9. BIBLIOGRAFÍA.....	128
10. ANEXOS	140
10.1 Anexo 1: Cronograma de actividades	140
10.2 Anexo 2: Gráfica de las instituciones	142
10.3 Anexo 3: Cuestionario real aplicado	143
10.4 Anexo 4: Dataset	144
10.5 Anexo 5: Codificación de los algoritmos en Python	145
10.6 Anexo 6: Dirección Web del aplicativo generado	146
10.7 Anexo 7: Formato para trabajo interdisciplinar	147

ÍNDICE DE TABLAS

Tabla 1. Antecedentes regionales, nacionales e internacionales	14
Tabla 2. Características generales de las instituciones objeto de estudio	24
Tabla 3. Categorización de herramientas ML y algoritmos respectivos	39
Tabla 4. Matriz de confusión	48
Tabla 5. Valoración de concordancia por el índice k	51
Tabla 6. Rangos de valoración para el AUC – ROC.....	52
Tabla 7. Aspectos de acuerdo al entorno, que pueden influir en el aprendizaje y conductas de los estudiantes.....	64
Tabla 8. Variables socio demográficas consideradas en la investigación.....	66
Tabla 9. Características generales de la población	70
Tabla 10. Fases para el desarrollo del trabajo investigativo	72
Tabla 11. Acciones relacionadas con cada fase/objetivo	74
Tabla 12. Técnicas e instrumentos usados en la investigación	76
Tabla 13. Variables de entrada, convenciones y numeración de sub categorías.....	78
Tabla 14. Variables de salida, convenciones y numeración de sub categorías	80
Tabla 15. Comparación de rendimientos de algoritmos machine learning.....	86
Tabla 16. Nuevas métricas de incertidumbre	92
Tabla 17. Importancia de las características sociodemográficas para cada predicción	103
Tabla 18. Importancia de los predictores para cada una de las salidas.....	104
Tabla 19. Estrategia de intervención en las instituciones educativas	112
Tabla 20. Cronograma de actividades.....	140

103

ÍNDICE DE FIGURAS

Figura 1.	Ubicación geográfica de los municipios del Agrado y Garzón, Huila ..	24
Figura 2.	Ubicación de las instituciones educativas	25
Figura 3.	Relación de fundamentos teóricos	26
Figura 4.	Algunas relaciones dentro de Data Science	31
Figura 5.	Relación entre Dato, Información y Conocimiento.....	34
Figura 6.	Tipos de aprendizaje ML	37
Figura 7.	Diagrama de flujo para el uso de herramientas Machine Learning	38
Figura 8.	Árbol de decisión usando el algoritmo J48 en Weka	43
Figura 9.	Red bayesiana usando Weka	44
Figura 10.	Representación del funcionamiento de una red neuronal.....	44
Figura 11.	Interpretación del RandomForest.	45
Figura 12.	Representación intuitiva de un SVM.	46
Figura 13.	Representación de funcionamiento del K Nearest Neighbors.	47
Figura 14.	Representación del funcionamiento del Gradient boosting tree.....	47
Figura 15.	Expresión para la métrica Accuracy	49
Figura 16.	Expresión para la métrica precisión	49
Figura 17.	Expresión para la métrica recall	50
Figura 18.	Expresión para la métrica F1-Score	50
Figura 19.	Expresión para hallar el índice Kappa.....	51
Figura 20.	AUC – ROC	52
Figura 21.	Expresión para la métrica sensibilidad	53
Figura 22.	Expresión para la métrica especificidad	53
Figura 23.	Zona Residencial de acuerdo al género Biológico.....	70
Figura 24.	Edad de acuerdo al género biológico	71
Figura 25.	Reprobación de acuerdo al género biológico	71
Figura 26.	Relación metodológica.....	75
Figura 27.	Desempeños en competencias básicas I. E. La Merced	82
Figura 28.	Desempeños en competencias básicas I. E. Ramón Alvarado.....	82
Figura 29.	Algunos atributos considerados en la base de datos	83
Figura 30.	Muestra de la base de datos inicial	84

Figura 31.	Dataset preliminar: 92 muestras	84
Figura 32.	Código del aplicativo web: PHP	88
Figura 33.	Screen del aplicativo Web generado	88
Figura 34.	DecisionTreeClassifier usando Python.....	89
Figura 35.	Rendimiento de los algoritmos para la salida Matemáticas	89
Figura 36.	Rendimiento de los algoritmos para la salida Ciencias Naturales	90
Figura 37.	Rendimiento de los algoritmos para la salida Ciencias Sociales	90
Figura 38.	Rendimiento de los algoritmos para la salida Inglés.....	90
Figura 39.	Rendimiento de los algoritmos para la salida Lectura	91
Figura 40.	Nuevos rendimientos de los algoritmos ML.....	93
Figura 41.	Esquema general de un árbol de decisión	95
Figura 42.	Expresión para el índice Gini	95
Figura 43.	Idea intuitiva de elección de un nodo Raíz.....	96
Figura 44.	Función de costo en el algoritmo CART	97
Figura 45.	DT, predicción de desempeños en Matemáticas.....	98
Figura 46.	DT, predicción de desempeños en Ciencias Naturales	99
Figura 47.	DT, predicción de desempeños en Ciencias Sociales	100
Figura 48.	DT, predicción de desempeños en Lectura	101
Figura 49.	DT, predicción de desempeños en Inglés	102
Figura 50.	Relación entre desempeños-Matemáticas	104
Figura 51.	Relación entre desempeños-Lectura.....	105
Figura 52.	Relación entre desempeños-Ciencias Naturales	106
Figura 53.	Relación entre desempeños-Ciencias Sociales	107
Figura 54.	Relación entre desempeños-Inglés	108
Figura 55.	Subdivisión de carpetas del aplicativo.....	109
Figura 56.	Ruta de predicción de dificultades de estudiantes en matemáticas..	110
Figura 57.	Socialización de la investigación con docentes de las Instituciones.	114
Figura 58.	Opiniones sobre mejorar las prácticas pedagógicas	115
Figura 59.	Opinión, idea central del proyecto investigativo	115
Figura 60.	Percepción de la fiabilidad del estudio realizado	116
Figura 61.	Consideración sobre interrelación entre los desempeños de áreas .	116

Figura 62.	Disposición para reorientar prácticas pedagógicas	117
Figura 63.	Valor de utilidad dado a la aplicación	117
Figura 64.	Trabajo interdisciplinar en I.E. La Merced y Ramón Alvarado	120
Figura 65.	Trabajo interdisciplinar pro - desarrollo de competencias básicas....	121
Figura 66.	Trabajo interdisciplinar pro - desarrollo de competencias básicas....	123

1. INTRODUCCIÓN

Vivimos en una era digital, en una era de desborde en información y acceso a la misma. Hace un par de décadas era impensable que el acceso a internet siquiera se daría de manera global. Hoy mismo, justo es a través de la conectividad que es posible agilizar procesos, investigar, trabajar, señalar y hasta comprar elementos para el hogar.

El mundo está en constante evolución y los campos del conocimiento en pro de su propio avance, apuntan al desarrollo de competencias que, sin importar los contextos singulares de cada comunidad educativa, deberían dar resultados similares. En los últimos registros del DANE (Departamento Administrativo Nacional de Estadística) se menciona que en nuestro país más del 40% de la población vive en medio de la pobreza y, aun así, pobres y ricos son evaluados por un mismo sistema y desde una misma perspectiva de valoración. De ahí que dentro de los contextos urbanos o rurales referidos a los estudiantes del Agrado y Garzón – Huila, se estudie un problema relacionado con la necesidad de analizar las dificultades de una muestra de dichos estudiantes; a saber, los que cursan grado décimo, para reflexionar sobre competencias básicas y las posibles relaciones con sus características sociodemográficas.

En el marco del análisis de las posibles dificultades y tomando como referente el trabajo de Barrios-Vargas y Caceres-Soto (2017), se busca entonces hacer uso de la sistematización de las notas y un análisis multivariado en el que, se puedan clasificar ciertas tendencias que permitan definir qué aspectos particulares de los contextos referidos podrían repercutir en las dificultades de desarrollo de competencias. En ese sentido se orienta la elaboración de un entorno virtual institucional que con el uso del concepto de *machine learning* y las herramientas respectivas, permita – entre otras cosas –, reflexionar sobre la importancia de innovar en las prácticas pedagógicas a través de las cuales se enseñan contenidos que deben promover ciudadanos críticos y reflexivos sobre su entorno social, familiar, educativo y cultural, de manera que pueda evidenciarse una mejor comprensión de la brecha social que distancia unos sectores de otros.

De acuerdo con lo anterior, en concordancia con la perspectiva compleja de Morin (1999) y Maldonado (2020) acerca de la educación, se brinda un estudio de las variables que podrían estar asociadas a las dificultades para el fomento de competencias básicas. Esto, haciendo uso de algoritmos *machine learning* que podrían generar predicciones al

clasificar desempeños competenciales, conforme a relaciones con aspectos cualitativos del entorno social, institucional y familiar de los estudiantes dentro del grado décimo de la Institución Educativa La Merced, del Agrado – Huila y de la Institución Educativa Ramón Alvarado Sánchez de Garzón – Huila.

Por último, la investigación promueve un espacio de discernimiento entre docentes - principalmente- y la comunidad educativa en general, para que, conforme a la presentación de los hallazgos, puedan, en sentido consecuente, aportar al enriquecimiento de los planteamientos finales del presente trabajo y cuyo propósito esencial estriba en sentar un referente plausible como propulsor de cambios, mejoras o propuestas innovadoras para que el sistema educativo aplicado al interior de las instituciones educativas valore la singularidad, se reflexione introspectivamente sobre el sendero recorrido durante varias décadas y sobre el sentido de la evaluación como proceso continuo dentro de las prácticas educativas y sistema útil para el progreso académico de las instituciones .

2. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

El uso de pruebas estandarizadas para medir desempeños en los estudiantes y la cuantificación de tales desempeños dentro de las instituciones educativas, podrían dar una idea sesgada del nivel competencial de los estudiantes. A continuación, se describe el problema y la formulación del mismo.

2.1 Descripción del Problema

El problema de investigación se aborda desde la línea: Ciencias de la complejidad en Educación y se plantea desde el análisis multivariado de aspectos cualitativos y cuantitativos.

Generalmente en todas las instituciones educativas, los procesos de evaluación de cada periodo consisten en asignar una apreciación cuantitativa que define el desempeño del estudiante en cada asignatura y es eso lo que determina si el joven cumple con los desempeños establecidos que obedecen a unas competencias básicas para ser promovido al siguiente grado, pero de acuerdo con Méndez y López (2019) los desempeños académicos de los estudiantes se pueden relacionar con diferentes variables no necesariamente cuantitativas, por lo que, este tipo de calificación, podría estar limitada a dar una apreciación de si se cumple o no, con unos objetivos propuestos de manera estandarizada, y no está analizándose la información generada, para mostrar otras consideraciones cualitativas que pueden ayudar al docente y a la institución a reflexionar sobre dificultades inherentes a la singularidad de los estudiantes y sus respectivos entornos, para por ejemplo, mejorar estrategias de enseñanza – aprendizaje que promuevan el desarrollo de competencias.

En este sentido, se podría percibir un amplio potencial en la información referida a múltiples variables como la zona de residencia, la cantidad de hermanos, el tipo de familia a la que pertenece, la lejanía a la institución, el nivel de ingresos familiar, entre otros aspectos y su relación con desempeños académicos en competencias básicas evaluadas en el sistema educativo actual. No obstante, la posibilidad de encontrar tales relaciones, se ven imposibilitadas por medios tradicionales, descriptivos o lineales. Es decir, a pesar de que por muchos años se tiene información sociodemográfica de los

estudiantes, es improbable por simple inspección, relacionar múltiples características con los desempeños o desarrollos competenciales de uno u otro estudiante.

Consecuentemente, desde la perspectiva compleja, se estudian y usan los avances informáticos para llegar a considerar interacciones ocultas y las interrelaciones entre múltiples variables y no entre pocas variables como se suelen analizar los problemas dentro de las orientaciones científicas heredadas de occidente; esto con el fin de no omitir posibles causalidades que deberían generar una comprensión más cercana a lo que podría ser la naturaleza del problema.

Teniendo en cuenta aspectos de interacción propios de la vida y del sistema educativo, se considera que las perspectivas actuales conforme al currículo, la evaluación y las prácticas pedagógicas o didácticas, podrían reformarse conforme se pueda establecer correspondencias, tendencias o correlaciones entre distintos ámbitos que caracterizan el comportamiento y desempeño académico de los estudiantes. Lo anterior, sin dejar de lado referentes que nos han conducido a la estructuración educativa actual, es decir, la evolución cognitiva que desde Piaget (1981) se argumenta y las zonas de desarrollo próximo que Vygotsky (1988) enmarca dando preponderancia a las relaciones externas y al entorno como agente formativo. No obstante, pensar en que hay relaciones conlleva un problema inminente, no solo porque hay muchas variables a tomar en cuenta, sino porque la manera en que se puedan tratar estas para su estudio puede ser errada. Es por ello que dentro de la situación problemática encontrada, también se debe enmarcar el método con el cual podría establecerse una solución.

Se puede establecer entonces que parte de las dificultades yacen en la abundante información relacionada con notas, caracterización y el amplio número de personas que se vinculan a las instituciones con sus respectivas singularidades cognitivas, históricas, culturales, de entorno, lo que obstaculiza realizar un seguimiento a la evolución cognitiva del estudiante y a las posibles causas que, ajenas a las instituciones unas o con los establecimientos inmersos en ellas, conllevan desempeños bajos, básicos, altos o superiores en los estudiantes dentro de las distintas competencias básicas que se esperan desarrollar. Adicionalmente, es de recalcar que las notas se encuentran aisladas

por asignatura y no permiten tener un panorama más amplio que pueda orientar sobre posibles relaciones en cuanto a las dificultades referidas.

De ese modo, el problema estriba en lo que hasta ahora es una situación impredecible, ya que estudiar cómo o cuáles variables podrían influir en las dificultades asociadas al desarrollo de competencias, en entornos educativos diferenciados como es el caso de la institución educativa la Merced en el municipio del Agrado y la institución educativa Ramón Alvarado Sánchez en zona rural de Garzón – Huila, no es algo que pueda analizarse desde o a partir de métodos tradicionales.

2.2 Sistematización del Problema

Conforme al apartado anterior, surgen cuestiones relacionadas con la importancia de analizar los aspectos que podrían generar dificultades en el desarrollo de competencias dentro de los contextos específicos de las dos instituciones mencionadas.

A continuación, se describen tales interrogantes:

¿Cómo analizar las dificultades asociadas al desarrollo de competencias básicas en estudiantes de décimo grado, usando métodos de *machine learning* en las instituciones educativas Ramón Alvarado Sánchez y la Merced? ¿Cómo valorar los resultados de la herramienta *machine learning*, para relacionarlos con las dificultades reales dentro del entorno académico de los estudiantes de décimo grado de las I. E. Ramón Alvarado Sánchez y la Merced, de los municipios Garzón y Agrado (Huila)? ¿Con base a la validación de los resultados arrojados, se puede crear un entorno virtual institucional que permita predecir y disponer de la información relacionada con las dificultades en las competencias básicas de los estudiantes de décimo grado de las I. E. Ramón Alvarado Sánchez y la Merced, de los municipios Garzón y Agrado (Huila)?

2.3 Enunciación del Problema

Teniendo en cuenta el apartado de sistematización, el problema de investigación se centra entonces en presentar un análisis relacionado con los aspectos que podrían generar dificultades en el desarrollo de competencias básicas en estudiantes de grado

décimo, tomando en consideración particularidades de contexto, desempeños académicos y la incorporación de las TIC como herramientas para el cruce de información que pueda concurrir en el establecimiento de posibles relaciones que permitan entender la trascendencia de los entornos y en general de factores sociodemográficos dentro del desarrollo de competencias básicas que genere un punto de reflexión sobre los procesos de enseñanza – aprendizaje a partir de un entorno virtual institucional.

Analizada la necesidad, se busca dar solución a la cuestión:

¿Cómo predecir dificultades en el desarrollo de competencias básicas relacionando aspectos sociodemográficos y de desempeño, mediante el uso de herramientas *machine learning* en estudiantes de décimo grado de las I.E. Ramón Alvarado Sánchez y la Merced, de los municipios Garzón y Agrado en el departamento del Huila?

Conforme se logre dar respuesta al anterior cuestionamiento, se espera generar un producto que, en el ámbito académico, podría ser un entorno virtual que sirva como herramienta para el análisis de información sociodemográfica y de desempeños académicos de los estudiantes para reflexionar sobre las prácticas pedagógicas.

Por otro lado, en términos investigativos, podría ser un artículo generado a partir del análisis multivariado que, desde la perspectiva compleja permita reflexionar en el ámbito educativo, sobre los aspectos incidentes en las dificultades asociadas al desarrollo de competencias en estudiantes de grado décimo.

3. ANTECEDENTES Y JUSTIFICACIÓN

Dentro de las investigaciones que abordan el carácter predictivo a través de herramientas *machine learning* en el ámbito educativo, se destacan los trabajos de: Florez y Pulido, 2020; Peña, et al. 2019 en espacios regionales. Las investigaciones de Méndez y López, 2019; De la Hoz, et al. 2018; Gonzalez y Peñaloza, 2021; Castrillón et al. 2021; Contreras et al, 2020 en espacios nacionales, e internacionalmente, los trabajos de Ramirez, 2021, (España); Rojas, 2013, (Costa Rica); Caraballo, 2020, (Rep. Dominicana); Nieto, et al, 2019, (España); Vilca, 2019, (Perú) y Urreaga, et al. 2020, (España).

Cada una de las anteriores investigaciones han enriquecido y justificado la presente propuesta y orientan, con los aportes hallados, el desarrollo metodológico y epistemológico en torno a la predicción de dificultades en desempeños académicos de estudiantes tomando en consideración variables sociodemográficas.

3.1 Antecedentes

A continuación, se presentan en la tabla 1. Los aportes de cada antecedente hallado en espacios regionales, nacionales e internacionales, distinguiendo en cada fila los espacios dentro de los cuales se aplicaron los trabajos para comparar los múltiples aportes que se dan tanto a nivel metodológico como teórico, adicionalmente, en la última columna, se exponen conclusiones tomadas para el propósito de la presente investigación:

Tabla 1. Antecedentes regionales, nacionales e internacionales

E ¹	Título	Objetivo General	Problema (síntesis)	Metodología	Conclusión/Aporte
REGIONAL	Estrategia Didáctica mediada por Técnicas de Machine Learning para Potenciar la Habilidad de Pensamiento Crítico Interdisciplinar en Procesos Académicos y Actitudinales. (Florez y Pulido, 2020)	Proponer una estrategia didáctica que potencie las habilidades de pensamiento crítico interdisciplinar en los procesos académicos y actitudinales mediante una Unidad Didáctica diseñada a través de la implementación de técnicas de Machine Learning en estudiantes de grado sexto del Colegio Claretiano de Neiva y Colegio Campestre Padre Arturo de Neiva.	La investigación se centra en ¿Cómo potenciar las habilidades de pensamiento crítico interdisciplinar en los procesos académicos y actitudinales mediante una Unidad Didáctica estructurada por la implementación de técnicas de Machine Learning y aplicada en estudiantes de grado sexto?	la investigación tiene enfoque mixto debido a que se consideran tanto aspectos cuantitativos como cualitativos. Éstos junto al uso del <i>machine learning</i> , orientan el trabajo.	El trabajo permite pensar en la posibilidad de: con las herramientas tecnológicas actuales, diseñar bases de datos que permitan, a través de la minería de datos, entre otras cosas, identificar y caracterizar los factores psicosociales, emocionales e intereses cognoscitivos que se pueden relacionar con los estilos de aprendizaje y las inteligencias múltiples de los estudiantes. Lo anterior sustenta la presente investigación y deja la posibilidad de enriquecerse en otros contextos.
	Modelo basado en minería de datos para el análisis de los intereses cognoscitivos y los factores psicosociales que inciden en el aprendizaje de los estudiantes. (Peña, et al. 2019)	Proponer un modelo de caracterización y análisis de los factores psicosociales y los intereses cognoscitivos de los estudiantes del grado sexto de las instituciones educativas Enrique Olaya Herrera y Oliverio Lara Borrero de la ciudad de Neiva, que permita potenciar el aprendizaje y reducir los factores de deserción escolar.	La investigación se centra en ¿Cómo identificar los factores psicosociales, cognoscitivos, intereses y estilos de aprendizaje de estudiantes, con el fin de cualificar las estrategias didácticas que permitan mejorar el rendimiento académico y minimizar los niveles de deserción?	la investigación tiene enfoque mixto y se enmarca dentro de la línea: “sistemas dinámicos complejos en fenómenos naturales, sociales y modelamiento matemático”	La evidencia presentada en la investigación permite valorar el campo del Machine Learning en la Educación, mirando éste como un “mundo nuevo” por descubrir, considerando que, a través de su uso, se promueve en los profesionales e investigadores de la Educación la generación cambios positivos tanto en los estudiantes como en el actuar pedagógico. Se destaca además, desde la investigación, el reconocimiento de necesidades y la posibilidad de fortalecer habilidades para la vida.
NACIONAL	Identificación y predicción de riesgo de deserción de estudiantes en	Analizar por medio de una herramienta machine learning, los factores que afectan la deserción dentro de la facultad de	El trabajo presenta una problemática recurrente en la educación, a saber, la deserción. Se plantea la necesidad de	El enfoque de la investigación es cualitativo en tanto se valoran variables socio-	La investigación establece un antecedente de uso de herramientas machine learning, en ese sentido, se muestra que el modelo implementado puede ser utilizado como herramienta de

¹ Espacio/ámbito de la investigación

E ¹	Título	Objetivo General	Problema (síntesis)	Metodología	Conclusión/Aporte
	académica por medio de modelos basados en machine learning. (Gonzalez y Peñaloza, 2021)	ingeniería, y predecir la probabilidad de deserción que pueda tener un estudiante.	identificar factores que contribuyen a la deserción de estudiantes en educación universitaria y se promueve el uso de métodos <i>machine learning</i> par tal propósito.	demográficas y cuantitativa al buscar encontrar un método de <i>machine learning</i> que brinde un valor de pronóstico matemáticamente pertinente.	predicción de la deserción, lo cual puede extenderse a otros propósitos. Se destaca que éstas herramientas buscan permitir la toma de acciones pertinentes que conlleven (para este caso particular) la retención y permanencia de los estudiantes en la institución.
	Hacia un método de predicción de resultados de evaluación en un contexto de micro aprendizaje. (Sánchez et al. 2020)	Proponer un método para predecir los resultados de la evaluación de los alumnos que interactúan con un sistema de microaprendizaje consciente del contexto.	Dentro del marco de la investigación, se plantea la problemática relacionada con el tratamiento, análisis entendimiento y uso predictivo de los datos dentro de contextos educativos	La investigación toma enfoque cuantitativo y cualitativo en el sentido de querer entender los datos para generar predicciones por medio de un método apropiado.	Resulta favorable para la propuesta, verificar desde ésta investigación, que el modelo Random Forest les proporcionó una precisión del 94%, y que usaron además una técnica de validación cruzada. Con ello, es fundamento importante considerar la posibilidad de que el modelo aplicado, logra predecir con precisión el resultado de la evaluación y relacionado con el contexto del alumno.
	Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. (Contreras et al, 2020)	Plantear la selección de variables que influyen en la predicción del rendimiento en estudiantes de ingeniería industrial de la Universidad Distrital (Colombia) por diferentes metodologías: filtro, envoltura e integrados.	El problema sobre el que toma sentido la investigación se enmarca en la necesidad de solucionar diversos aspectos académicos y administrativos que surgen en la educación y que desde el análisis de datos, pueden tratarse.	Dentro de la síntesis metodológica, los autores sostienen que, pensar en un método de solución a la problemática, se torna tan complejo como la determinación de las variables a considerar. Exponen así un método de 7 pasos.	La investigación permite visualizar la implementación de algoritmos de clasificación a través del lenguaje de programación Python. Con ello, se brinda un referente sobre el uso de: árbol de decisión, K vecinos más cercanos, perceptrón y otros que, dentro de la investigación, fueron comparados para validar el que brindara mejor resultado de predicción. Se recalca, además, que algunas variables que influyen en el rendimiento académico de los estudiantes de ingeniería son: edad, género, puntaje ICFES para aptitud matemática, puntaje global ICFES, valor de matrícula y puntaje ICFES para condición matemática y cohorte. Con lo cual se evidencia la importancia de características cualitativas.

E ¹	Título	Objetivo General	Problema (síntesis)	Metodología	Conclusión/Aporte
	Técnicas de Machine Learning para la predicción de desempeño académico en el Desarrollo del espacio proyectivo del Pensamiento Espacial. (Méndez y López, 2019)	Evaluar la incidencia del uso de machine learning en la predicción del desempeño en el espacio proyectivo del pensamiento espacial.	La investigación se centra en la solución de la 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?	En la investigación se usa el modelo Cross Industry Standard Process for Data Mining (CRISP-DM). Dicho modelo orienta gran parte de las investigaciones basadas en minería de datos y técnicas de <i>machine learning</i> .	Respecto a la investigación de Méndez y López, (2019) se deben destacar varios aspectos que se convierten en antecedente y referente para la propuesta. A saber: Se considera que los factores o aspectos asociados al aprendizaje pueden ser de índole familiar, institucional o personal. Se referencia que, los factores socioeconómicos familiares, pueden desanimar a los estudiantes y repercutir en la deserción escolar. Se destaca el uso de técnicas de Machine Learning (aprendizaje automático) y la simulación para desarrollar modelos que permitan la predicción de desempeños de estudiantes en diferentes niveles de escolaridad. Con base a lo anterior, surge en nuestra era (y con base a la investigación), la posibilidad de usar una herramienta eficaz para el docente y en general en el campo de la educación, para clasificar y conocer con alto grado de precisión las categorías de DESEMPEÑO_BAJO y DESEMPEÑO_ALTO de los estudiantes. Así, de manera consecuente, se brinda la oportunidad de crear estrategias en cada una de las asignaturas académicas
	Metodología de Aprendizaje Automático para la Clasificación y Predicción de Usuarios en Ambientes Virtuales de	Desarrollar una metodología para clasificar y predecir usuarios en ambientes virtuales de educación, estudiando la interacción de los estudiantes con la plataforma y su	El trabajo se centra en la solución de cuestiones como ¿De qué forma clasificar los estudiantes participantes en un ambiente virtual de educación utilizando el aprendizaje automático? ¿Qué	Dentro del enfoque metodológico se realiza el trabajo desde el análisis de clustering y se introduce la lógica Fuzzy para interpretar transiciones entre	Desde la investigación, surgen dos aspectos relevantes como antecedente para la propuesta: * El establecimiento de criterios para el estudio y análisis de procesos de transición de perfiles de los estudiantes de acuerdo al rendimiento alcanzado previamente.

E ¹	Título	Objetivo General	Problema (síntesis)	Metodología	Conclusión/Aporte
	Educación. (De la Hoz, et al. 2018)	desempeño en los exámenes.	factores asociados a un ambiente virtual de educación determinan los niveles de conocimiento de los estudiantes?	niveles de conocimientos, además se aplica la técnica de k-vecinos dentro del aprendizaje automático.	* Con base a la metodología propuesta en la investigación, se aprecia la posibilidad de pronosticar la pertenencia a un conglomerado de un estudiante, de acuerdo a las variables inherentes utilizadas en esta investigación.
INTERNACIONALES	Análisis del machine learning como estrategia didáctica para el mejoramiento de competencias básicas en educación secundaria en entornos virtuales de aprendizaje (Ramirez, 2021. España)	Desarrollar estrategias didácticas con base en el Machine Learning para el mejoramiento de competencias básicas de pensamiento matemático en educación secundaria en entornos virtuales de aprendizaje.	La investigación se centra en la importancia de identificar aspectos o procesos en los que puede ser integrada la IA (Inteligencia Artificial) y específicamente el Aprendizaje Automático (en inglés, Machine Learning) para el mejoramiento de la educación.	La investigación tiene un enfoque cuantitativo, se basa en una secuencia de fases para comprobar hipótesis.	La investigación presenta un antecedente, por un lado, sobre la determinación de variables predictoras que se pueden relacionar significativamente con el alcance de competencias básicas en pensamiento matemático de los alumnos y por otro, sobre la influencia del machine learning en la realización de actividades para el nivel de educación básica. Adicionalmente, se resalta la muestra de un diseño de un entorno virtual de aprendizaje para la implementación de estrategias didácticas con base al uso de Machine Learning. Esto último, orienta acerca de las bondades de un entorno virtual y de la mediación del ML.
	Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea. (Urteaga, et al. 2020. España)	Verificar si es posible predecir deserción solamente con los datos de las interacciones entre estudiantes y la plataforma en las dos primeras semanas de un curso. Ver qué algoritmo produce los mejores modelos predictivos en términos de la métrica de impacto.	La problemática, en sentido general y sintetizado de la investigación es la necesidad de prevenir la deserción para que la educación, tanto online como presencial, alcance su potencial.	La investigación tiene enfoque cuantitativo, puesto que, a través de criterios que enmarcan resultados numéricos, se busca predecir la deserción o no, de algunos grupos de estudiantes.	Resulta positivo desde este referente, valorar el porcentaje de efectividad de algunos modelos predictivos. Adicional, es particularmente interesante vislumbrar que, dentro de los modelos usados, la red neuronal entrenada con retropropagación fue la de mayor efectividad. Así, se concluye que, "las redes neuronales generaron los mejores modelos con diez neuronas ocultas y se verificó la estabilidad de esa solución" (Urteaga, et al. 2020. Pp. 7)

E ¹	Título	Objetivo General	Problema (síntesis)	Metodología	Conclusión/Aporte
	<p>Predicción del nivel de estrés en estudiantes universitarios utilizando técnicas de machine learning. (Vilca, 2019. Perú)</p>	<p>Utilizar los datos psicológicos para analizar y pronosticar el nivel de estrés y así ayudar a mejorar el tratamiento a los profesionales de la salud.</p>	<p>La investigación busca, por medio del aprendizaje automático, predecir el nivel de estrés académico producido por la demanda que se exige en el ámbito educativo.</p>	<p>La investigación tiene enfoque mixto, puesto que, valora características cualitativas que, cuantitativamente, deben relacionarse con un método machine learning que permita realizar predicciones.</p>	<p>La investigación no solo orienta sobre el uso de machine learning para propósitos predictivos en el campo educativo, sino que, además, establece un referente que sugiere aplicar en futuras investigaciones el desarrollo e implementación de una interfaz de usuario que proporcione facilidad de uso mediante un aplicativo web o móvil que promueva en mejor medida la interacción con las personas. Con base a lo anterior, la investigación de Vilca (2019) se convierte en un antecedente que se puede llevar al campo de la educación básica y media, para por ejemplo ofrecer interacción a los docentes con posibles desempeños académicos de estudiantes, de acuerdo al análisis de variables socio-demográficas.</p>
	<p>Decision – Making Model at Higher Educational Institutions based on Machine Learning. (Nieto, et al, 2019. España)</p>	<p>Proponer una metodología detallada usando machine learning cuando se llevan a cabo pronósticos académicos en estudiantes de educación superior</p>	<p>La investigación centra su atención en la necesidad de establecer algoritmos computacionales que permitan mejorar la toma de decisiones en instituciones de educación superior, de modo que, no se deje tanto espacio al azar, experiencias pasadas o a la intuición.</p>	<p>Se presenta un enfoque mixto, en tanto se centra en encontrar un método machine learning que apoye la toma de decisiones relacionadas con aspectos académicos, de deserción, detención de estudiantes, de currículo e incluso con la administración del personal docente.</p>	<p>La investigación deja de manifiesto que el uso de métodos o herramientas machine learning para situaciones de toma de decisiones o de tipo predictivas, está sujeto a la naturaleza del problema que se espera resolver. En ese sentido, la elección de un método no está definida de forma estandarizada. Conforme a ello, se deja un antecedente más sobre la comparación de algoritmos o métodos de machine learning que podrían usarse en investigaciones similares o al menos relacionadas.</p>

E ¹	Título	Objetivo General	Problema (síntesis)	Metodología	Conclusión/Aporte
	Predicción de la dificultad de la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica (Rojas, 2013. Costa Rica)	Plantear un modelo estadístico que permita predecir la dificultad de los ítems de la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica.	El problema de esta investigación radica en la necesidad de validar la idoneidad de la prueba de habilidades cuantitativas y de cómo definir un método que ayude u optimice tal propósito	El enfoque metodológico se centra en aspectos cuantitativos de la prueba y del método a elegir, sin embargo también se valoran aspectos cualitativos cómo la pertinencia o no de algunos ítems de la prueba de habilidades.	La investigación se convierte en un antecedente que direcciona acerca del posible uso del modelo LLTM para efectos de la investigación sobre desempeños y al análisis de sus supuestos considerando las características del problema propuesto. Por otro lado, similar a los antecedentes anteriores, brinda la posibilidad de validar otros modelos como el lineal múltiple y el logístico multinivel que pueden compararse y discernir sobre su pertinencia conforme al objetivo que se tenga.

Fuente: elaboración propia

3.2 Justificación

La presente es una propuesta cuyo propósito esencial estriba en varios aspectos y que a continuación se describen:

De manera inicial, tomando en consideración la trascendencia de las competencias básicas, tanto para las pruebas estandarizadas como para el progreso académico, se resalta que, desde PISA (2003), las competencias comunicativas, matemáticas y científicas se tornan importantes para el respectivo estudio impulsado por la OCDE (Organización para la Cooperación y el Desarrollo Económico) debido a la necesidad para cada persona de la comprensión de aspectos numéricos, variacionales, probabilísticos y demás, en la vida cotidiana y el mundo en general (Recio, 2006), en ese sentido, no solo se busca identificar la relevancia de las competencias matemáticas para la resolución de problemas contextualizados, sino también de las competencias científicas y tecnológicas en el ámbito académico y social. Ahora, de acuerdo con Paredes-Labra et al. (2019), se referencia la trascendencia de la competencia digital y tecnológica, no solo porque es una oportunidad para que los estudiantes se formen en un área que les permita manejar críticamente la abundante información que se maneja actualmente, sino porque también es necesario que desde el personal docente se tengan conocimientos esenciales para el uso de herramientas imprescindibles para solucionar problemas o agilizar la solución de los mismos tanto en ámbitos académicos como sociales, culturales y científicos.

Por otro lado, en sentido general, se da valor al afán global por estandarizar organizaciones académicas y, por lo tanto, dar relevancia al desarrollo competencial que por ejemplo trae como consecuencia la sintonización de estructuras educativas en Europa (Bezanilla, 2003) y que brinda un referente para avanzar en Latinoamérica sobre una puesta común que se mueva alrededor de las competencias, tanto a nivel escolar como universitario. Localmente, hablando de Colombia y los esfuerzos ejercidos por el Ministerio de Educación Nacional (MEN) en pro de la mejora educativa, se destacan los estándares básicos de competencias, los cuales son documentos con una estructuración que trasciende de los contenidos para orientar a los docentes sobre la importancia de

valorar desempeños y, en ese sentido, trabajos como el de Perroun (2008) y Rodríguez et. al. (2007), enmarcan aspectos que justifican la relevancia del estudio competencial.

Consecuentemente, si se evidencia la importancia, se debe pensar en las dificultades que pueden estar asociadas al desarrollo de tales competencias. De esta manera, si se hablara de desarrollos competenciales en matemáticas, se destacan los trabajos de Brousseau (1989), Socas (1997), Posso et al. (2007), Barallobres (2016), no obstante, la investigación considera dificultades desde otros referentes (Velásquez y Hernández (2016); Perrenoud (2008); Fernández et. al. (2016)), que consolidan la oportunidad para pensar en una manera de identificarlas teniendo como insumo las notas o desempeños reales de los estudiantes y relacionarlas con aspectos sociales y demográficos de los mismos, considerando la posibilidad de predecirlas usando una herramienta de *machine learning*.

De acuerdo a lo anterior, en segunda instancia, la presente investigación busca establecer una estrategia que permita predecir dificultades académicas de los estudiantes con base al análisis de sus características personales, familiares, sociales y escolares. De ese modo, se retoma principalmente el trabajo de Méndez y López (2019), como referente teórico y metodológico para la selección de variables sociodemográficas a considerar y el uso de una herramienta *machine learning* específica que sea fiable con base a métricas de incertidumbre y de acuerdo a la naturaleza del problema y el propósito de la investigación. Al respecto, también se destacan trabajos que fundamentan la comparación de herramientas machine learning en el campo educativo, a saber, De la Hoz, et al. (2018), Vilca (2019), Contreras et al. (2020), Urreaga, et al. (2020), Castrillón et al. (2021), Gonzalez y Peñaloza (2021).

Por último, y en concordancia con la prospectiva de la investigación de Vilca (2019), el cómo predecir dificultades en competencias básicas, usando una herramienta *machine learning*, se fundamenta en el desarrollo de un entorno virtual que permita visualizar y analizar de manera descriptiva (como en cualquier plataforma educativa) el comportamiento académico de los estudiantes de grado décimo (15-16 años) en el primer semestre 2022, pero adicionalmente, se espera, dentro del que será un aplicativo web, poder tener la facilidad de predecir dificultades en las ya mencionadas

competencias, relacionando variables de tipo personal, familiar, social y escolar. Así, el aplicativo, permitirá a los diferentes estamentos de las instituciones educativas Ramón Alvarado y La Merced (Del Agrado y Garzón – Huila, respectivamente), acceder a información específica y general de los estudiantes, además de considerar índices de predicción en los posibles desempeños para las competencias en cuestión mediante el uso de la información que se maneja, pero que no se utiliza más allá de las tablas de resumen y el llenado de informes en el conocido POEG (Proyecto de Orientación Escolar Grupal).

En sentido general, la propuesta aporta en términos teóricos, un análisis del desarrollo de competencias básicas y las dificultades inherentes; un marco referencial sobre ciencias de la complejidad y educación, además de un comparativo entre diferentes herramientas *machine learning* que pueden ser usadas para los propósitos educativos que atañen al problema predictivo sobre el que se desarrolla el trabajo investigativo. Dentro del enfoque metodológico, se presenta una orientación en la que se da relevancia a los aspectos cualitativos relacionados con las particularidades de los estudiantes y cuantitativos relacionados con el nivel de efectividad en las predicciones y los desempeños académicos propios de los estudiantes en las áreas que corresponden a las cinco competencias básicas que desde el MEN se preponderan para ser valoradas. Ahora, en sentido práctico, se busca establecer un entorno virtual útil, dinámico y funcional para ser usado por diferentes estamentos dentro de las comunidades educativas de las instituciones ubicadas en el municipio de Garzón y el Agrado, Huila. Con base a este último aspecto, desde la investigación, se promueve la posibilidad de usar la inteligencia artificial para propósitos educativos que pueden ir más allá de lo descriptivo, de lo lineal y de lo clásico, para con base a las ciencias de la complejidad, dentro del margen estocástico intrínseco al sector de la educación y de las características sociales de los estudiantes, se puedan predecir situaciones que permitan tomar decisiones asertivas sobre el mejoramientos de desempeños y desarrollo paulatino de competencias.

4. FUNDAMENTOS TEÓRICOS

Como lo expone Hawking (2003), en su libro *A Hombros de Gigantes*, la ciencia avanza conforme los científicos avanzan, dejan su legado y otros, posteriores, sustentan nuevas tesis con los fundamentos previos. En ese sentido, la presente investigación no nace de la nada ni se orienta desde la subjetividad. Realmente es un aporte al conocimiento, que tiene origen en contextos con problemáticas singulares pero que se pueden asemejar a las de otros lugares, conforme se escudriña, indaga e investiga sobre los ejes centrales que teóricamente dan formalismo al trabajo presentado. Así, en los siguientes apartados, se explicitan los ejes orientativos de la investigación.

4.1 Referentes contextual e Institucional.

Dentro de los marcos legislativos, tanto de nuestra república colombiana, como de la mayor parte de los países que se basan en gobiernos democráticos –incluso en los que no, también–, se promueve la oferta de una educación de calidad, inclusiva y necesaria para la mejora y el progreso en aspectos económicos y culturales. No obstante, en cada país, hay múltiples contextos, diversas circunstancias y, por tanto, multiplicidad de entornos en los que se desenvuelven las poblaciones y se transmite la cultura respectiva. La presente investigación, concurre en el anhelo de aportar conocimiento en el campo educativo, principalmente de la zona sur de Colombia, precisamente, en el departamento del Huila, lugar de tierras fértiles cuya economía se fortalece con el café, el cacao, el arroz y el turismo -principalmente-. Es justamente dentro de contextos agrícolas, que se llega a las instituciones Ramón Alvarado y La Merced, ubicadas en zona rural de Garzón y zona urbana del Agrado, respectivamente. Ambos municipios con más de siglo y medio de fundación e historia, con niveles socioeconómicos que van de estratos 0 a 1 en el caso del Agrado y de 0 a 3 en el caso de Garzón.

Al ser parte del territorio geográfico central del departamento, los dos municipios han sido víctima de conflictos encabezados por grupos insurgentes y los campesinos, principal población de ambos municipios, aún manifiestan padecer amenazas, ataques, hurtos y demás acciones delincuenciales que, estancan y corroen el progreso agrícola debido a la inseguridad que se evidencia en ciertos sitios. A pesar de las circunstancias, el sector agrícola desde la producción de café y cacao y el comercio activo por la

ubicación geográfica (ver figura 1), dinamizan la economía permitiendo que la vida sea amena, tranquila y educativa dentro de ambos lugares.

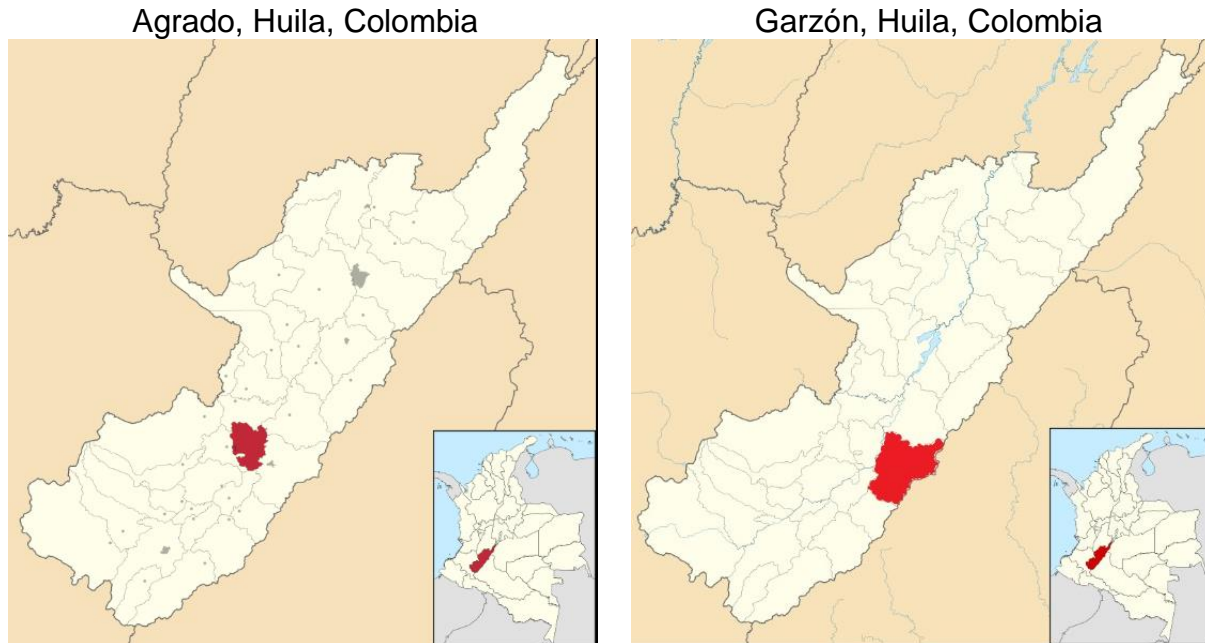


Figura 1. Ubicación geográfica de los municipios del Agrado y Garzón, Huila
 Fuente: https://es.wikipedia.org/wiki/Agrado#/media/Archivo:Colombia_-_Huila_-_Agrado.svg

Respecto a las instituciones, en la tabla 2, iniciando por la Institución Educativa La Merced, ubicada en zona urbana del Agrado, se muestran algunas características generales, seguidamente se enmarca la Institución Educativa Ramón Alvarado, ubicada en zona rural de Garzón (Ver figura 2).

Tabla 2. Características generales de las instituciones objeto de estudio

Institución	Sede	Ubicación	Laboratorios-salas	Especialidad	Número total de estudiantes
La Merced	Principal	Urbana	- Física; - Química; - Inglés; - ADC ² - Sistemas (3) - Electrónica	- Académica - Técnica en electrónica - Técnica en sistemas	650
Ramón Alvarado	Principal	Rural	- Biología; - Química; - Tecnología	- Académica - Técnica en gestión agropecuaria	627

Elaboración propia

² Aula de Desarrollo curricular, cuenta con computadores para cada estudiante, proyector, altavoces, computador para el docente e internet

A continuación, en la figura 2, se muestra la ubicación geográfica de cada institución



Figura 2. Ubicación de las instituciones educativas

Fuente: Google maps

4.2 Marco Teórico.

Dentro de los fundamentos base sobre los que se estructura la investigación, se toman aportes de diferentes autores que van desde la perspectiva compleja en términos de educación, la importancia de las herramientas tecnológicas actuales para incursionar en el tratamiento de problemas educativos y cómo el estudio de variables sociodemográficas y de desempeño en competencias, podrían relacionarse para considerar la importancia de la singularidad de los estudiantes y cómo las pruebas estandarizadas brindan información sesgada sobre las competencias de los estudiantes.

La figura 3 sintetiza la relación de los fundamentos teóricos que orientan el presente trabajo investigativo y plasma una orientación no vertical entre los campos conceptuales referidos.

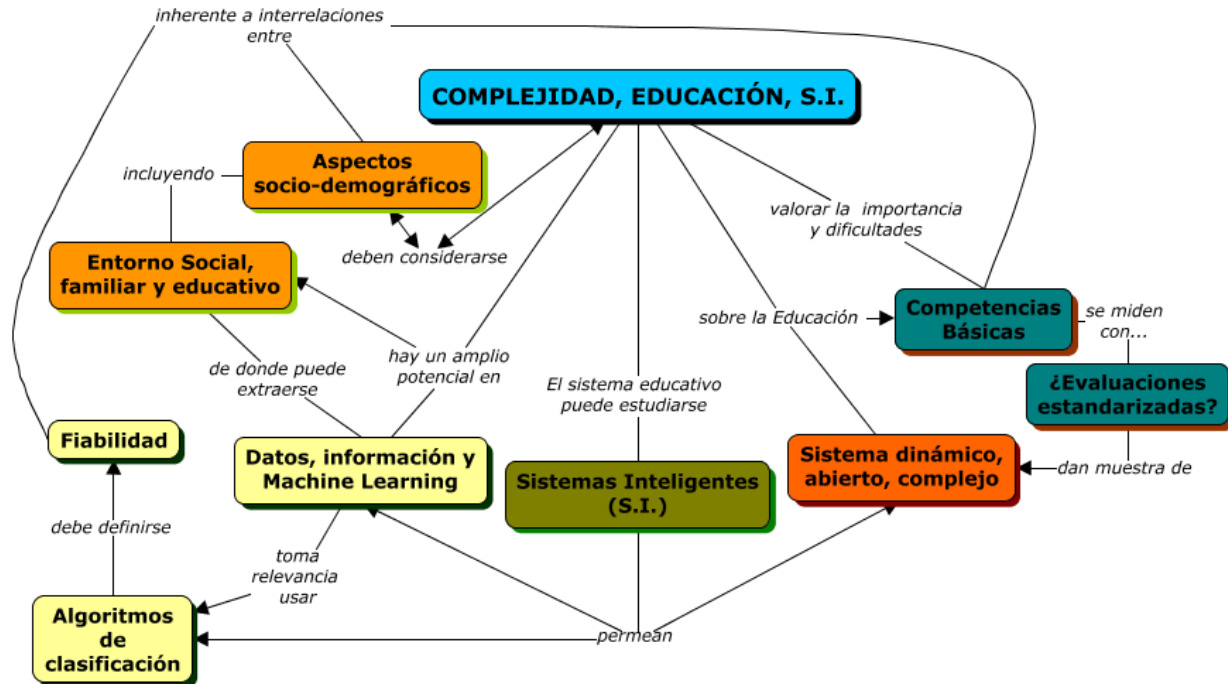


Figura 3. Relación de fundamentos teóricos
Fuente: Elaboración propia

4.2.1 Complejidad, Educación y Sistemas inteligentes

Albert Einstein menciona que, si se mide la inteligencia de un pez, de acuerdo a su habilidad para trepar un árbol, este pasará su vida entera creyendo que es estúpido. De acuerdo al postulado del Físico más influyente del pasado siglo, a continuación, se enmarcan los referentes que sustentan la investigación y que obedecen al paradigma de la complejidad implícito en la educación.

Se considera importante aclarar al menos tres aspectos que pueden generar confusión y que compete a esta investigación enmarcar para fines teóricos y orientativos referidos a la perspectiva compleja. En ese sentido, se debe mencionar que el pensamiento complejo, expuesto principalmente por Edgar Morin alude un método, una estrategia, un enfoque o perspectiva en la que, el mundo visto de manera reduccionista, no tiene coherencia con el afán de la “buena” ciencia, por descubrir la verdad en los problemas que surgen constantemente. Retomando a Maldonado (2012), es el pensamiento complejo una manera integradora de aproximarse al entendimiento del mundo y el

hombre, en la que, el carácter disciplinar de la ciencia es insuficiente para abordar las situaciones que surgen en un universo ineludiblemente cambiante.

Por otro lado, retomando el trabajo de Elizalde (2013), se debe diferenciar el pensamiento complejo del enfoque o pensamiento sistémico, puesto que, es en este segundo que se adopta una cosmovisión relativa al marco de referencia del observador, tal y como se expone desde la relatividad especial de Albert Einstein, la realidad para un observador en reposo, puede ser muy diferente a la realidad de un observador que se mueve de manera constante; podría pensarse entonces en un enfoque impreciso pero que aborda las diferencias y los cambios, las relaciones y la coherencia entre interacciones.

Por último, considerando la relevancia de las perspectivas actuales sobre complejidad para tratar el problema de investigación que compete al presente trabajo, es de recalcar dos aspectos importantes para entender el componente metodológico que orienta el abordaje de las posibles relaciones entre variables sociodemográficas y los desempeños académicos de los estudiantes. Así, evitando reducir o simplificar el problema, se considera en primera instancia, desde la complejidad, el uso de herramientas informáticas –un campo de conocimientos que yace en la cibernética– para posibilitar en términos probabilísticos, hacer predicciones de desempeños, pero, sobre todo, escudriñar posibilidades de relación entre las variables ya mencionadas al inicio de este párrafo; el segundo aspecto estriba en la no teorización, es decir, no se pretende atribuir a un solo campo disciplinar el tratamiento desarrollado, por el contrario, en correspondencia con el aporte de Maldonado (2012), se establecen argumentos para visionar la complejidad como el problema en el que confluyen diferentes ciencias para solucionar los sub problemas que coexisten tanto en la relaciones humanas, sociales, económicas, como en el mundo per se.

Con lo anterior, se esclarece en cierta medida la concepción de la complejidad, no como una ciencia, no como un paradigma, ni siquiera es un método o estrategia; es más que una cosmovisión y menos que una teoría unificada. Se referencia, con base a los ya mencionados autores [Morin (1994), Maldonado (2009, 2012), Elizalde (2013)] la complejidad como un confluir de disciplinas que, sin fronteras arbitrarias, se abren

camino hacia la solución pertinente y coherente de problemas. Problemas como el que se caracteriza en el capítulo 2 del presente trabajo y que se fundamenta en sentido teórico con los apartados que siguen.

4.2.1.1 Complejidad y Educación

De acuerdo con los aportes de Morin (1999) y Maldonado (2017), al considerar la Educación como un sistema altamente complejo, es ineludible tomar conciencia de la necesidad de abordar el sistema de la manera más amplia posible. En este sentido, en concordancia con el libro: Los siete saberes de Edgar Morin, compartido en el marco de la divulgación en pro de la mejora educativa por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, Francia, 1999), en la presente investigación, se confiere la importancia de reconocer lo que el autor denota como error e ilusión dentro de la construcción del conocimiento. Así, desde el propósito de buscar una manera de comprender mejor, aspectos que pueden influir en los desempeños de estudiantes, se otorga valor al hecho de movernos dentro de la ilusión al considerar que se tiene una organización pertinente dentro del sistema educativo, tanto en sentido pedagógico como administrativo y organizacional. De ahí que tome gran significado la, al parecer confusa distinción entre ilusión y error dentro del campo educativo.

Se considera, consecuentemente, que desde la sensibilización y conciencia ante la comprensión por lo que es error e ilusión, desde los roles que engendran los agentes del sistema educativo (comunidad educativa), se podría aportar desde la presente investigación un referente para prever posibles dificultades que se suelen obviar o que no son analizadas en detalle dentro de las instituciones educativas.

Otro aspecto a resaltar por el ya mencionado autor, dentro del mismo libro, es el titulado: principios de un conocimiento pertinente, desde la perspectiva compleja, atendiendo a aspectos de contexto, ergo, de cultura, de necesidades locales, familiares e institucionales. Claro, sin dejar de lado factores globales que afectan a las personas por habitar un mismo planeta, así, es imprescindible reconocer la importancia de orientar las prácticas pedagógicas desde Proyectos Educativos Institucionales (PEI) sensibles a cambios coherentemente relacionados con las posibles variaciones sociales en la comunidad educativa para la cual se crea. De esta manera, la propuesta que se enmarca

en la presente investigación, se inscribe en el reconocimiento de las diferencias, de las singularidades y la necesidad consecuente de tomar tales unicidades como factores que podrían – o al menos deberían – dar una mejor comprensión del cómo se pueden desenvolver los estudiantes en el ámbito académico.

En concordancia con Maldonado (2017), se brinda la posibilidad dentro del carácter académico de la propuesta, de reflexionar sobre la arraigada perspectiva disciplinar dentro del sistema educativo cuyo fundamento mayoritariamente justificado por la cultura occidental, desconoce la interdisciplinariedad e interrelaciones evidentemente benéficas que dieron, dan y deberían seguir dando solución a problemas de todo tipo. Así, se describe en términos generales la necesidad de por ejemplo vincular las ciencias de los datos y la computación, o incluso la inteligencia artificial, para aportar a la mejora de la educación que tal y como lo menciona el autor, debería cumplir una función liberadora más que organizativa desde los regímenes instaurados desde la hegemonía de las disciplinas y sus pedagogías o didácticas inherentes.

En torno a lo anterior, se deja de manifiesto desde los argumentos del autor, que la actual “disciplinariedad” en el sistema educativo colombiano, aunque resulta adecuada en términos organizativos, podría carecer de sentido y significado respecto al propósito de generar aprendizajes útiles para las futuras generaciones, de ahí que, desde la – cada vez más disiente – descentralización del conocimiento, muchas personas encuentren más atractiva la autoformación fuera de las escuelas y universidades. Al respecto, desde el marco operativo de la propuesta, se busca vincular dentro de las posibles predicciones de dificultades académicas, la necesidad de promover valores, a la par de competencias que trasciendan de los límites de las asignaturas estipuladas por los planes de estudio convencionales, es decir, se pretende sentar un referente que promueva la reflexión sobre los actuales currículos que orientan la educación.

Por último, dentro de los fundamentos teóricos en el campo de la educación y la perspectiva compleja, se referencia el trabajo de Dewey (1995), en el que se destaca un apartado de su libro *Democracia y Educación* relacionado con la necesidad educativa para la vida de las personas, mostrándose una vez más el carácter funcional, orientativo y formativo como eje fundamental para la sociedad que debe abordarse desde la

educación. En su obra, no solo se debe recordar que la educación se convierte en el proceso imprescindible para la preservación de la vida per se, sino que debe entenderse como un proceso inexorablemente adaptativo, cambiante y orientativo para el progreso de nuestra especie. Se considera de este modo, la relevancia de indagar e investigar más al respecto de la misma, además de justificar el trabajo alrededor de los agentes y aspectos constitutivos del sistema (estudiantes, padres de familia, sociedad, contexto, Instituciones, currículo, etc.).

Con base a lo anterior, alrededor de una leve vista al trabajo del autor, se considera, dentro de la presente propuesta, importante valorar la pluralidad de los estudiantes e incluso de los docentes, como un factor que contribuye no a entorpecer el avance en la aprehensión de conocimientos, saberes, desempeños o competencias, sino como una oportunidad para mostrar que educar se convierte en “arte” en tanto prescindamos de tecnicismos y las sistematizaciones recurrentes dentro de las instituciones que desconocen la necesidad de atender en mayor medida las diferencias que nutren la cultura, la ciencia y el arte (Cadrecha, 1990).

4.2.1.2 Sistemas inteligentes

Para explorar – leve o profundamente – el análisis de datos y el tratamiento de información en amplias cantidades, se hace necesario retomar algunos conceptos que son importantes para comprender el propósito que subyace en la presente investigación. En ese sentido, desde la Ciencia de los Datos (Data Science), se define el Data Mining (DM) como un proceso que permite extraer información nueva y trascendente de un conjunto de datos que en apariencia no muestra relación alguna (Britos, 2008), con base a esta nueva información se puede entonces generar conocimiento que, para efectos del presente trabajo, deberá mostrar posibles conexiones o dependencias implícitas dentro de las características socio-demográficas de los estudiantes. En general, se definen los Sistemas Inteligentes (SI) como software enfocados en la creación de conocimiento a partir del análisis de datos mediante distintas herramientas o algoritmos. Éstos, junto a la Inteligencia Artificial (IA), que hace referencia a máquinas o sistemas con capacidad de simular el razonamiento humano y el Machine Learning (ML), entendido como un subconjunto o disciplina en el ámbito de la IA debido a que son máquinas programadas

para el reconocimiento de patrones dentro del análisis de amplias cantidades de datos que deberían permitir realizar predicciones (Méndez y López, 2019), son el fundamento actual de la estructuración comercial globalizada que permite un mejor manejo de la información para potenciar su uso en múltiples campos. Campos dentro de los cuales, actualmente la educación no es discriminada.

De acuerdo con Loinaz (2001) y González (2004) el interés por vincular en la educación el uso de nuevas tecnologías, ha permitido un abordaje interdisciplinario dentro del mismo campo, así, el uso de la IA el ML y en general, los Sistemas Inteligentes, han generado la incursión en aspectos que promueven la creación de entornos culturales alrededor de la virtualidad y la digitalización de la sociedad. Adicionalmente, se otorgan libertades ante la necesidad de crear o explorar nuevas maneras de favorecer procesos de enseñanza-aprendizaje. La figura 4, muestra una relación general de los campos mencionados.

Data Science

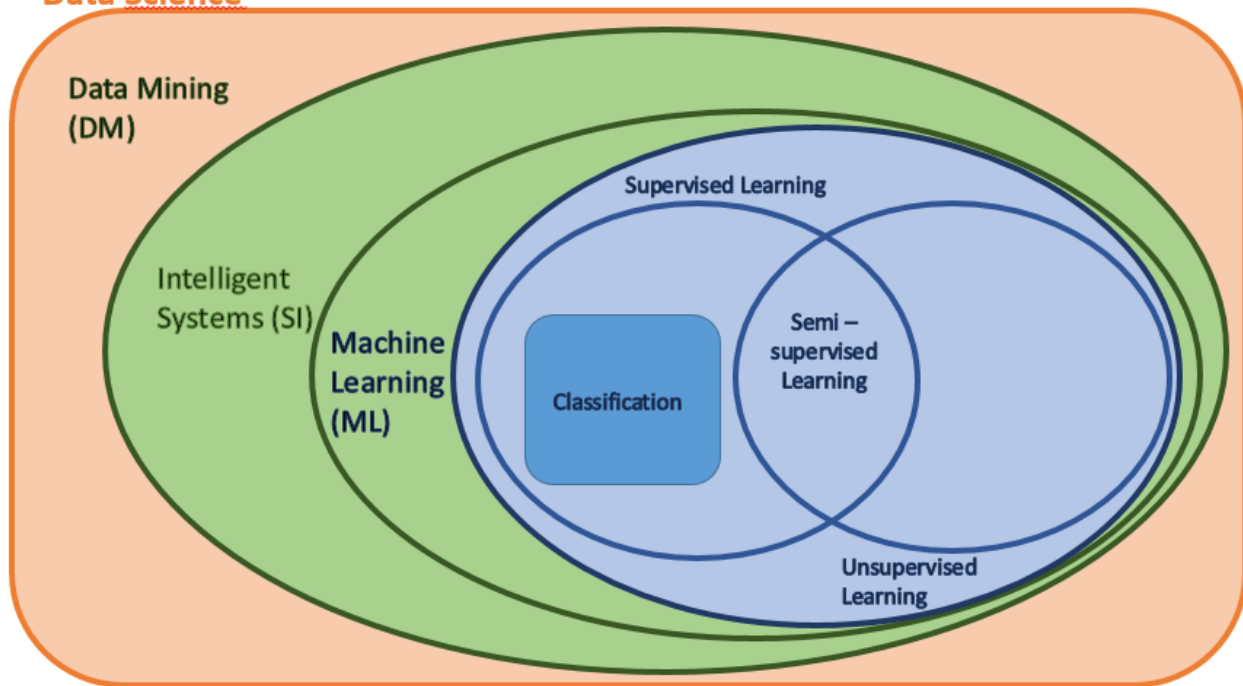


Figura 4. Algunas relaciones dentro de Data Science
Fuente: Elaboración propia

Retomando aspectos teóricos inherentes a este apartado, a continuación, se muestran las características que proporcionan las herramientas sobre las que se basan los sistemas inteligentes. Así de acuerdo con Osterwalder (2004), se tiene:

- Promueven la accesibilidad a la información, es decir, permiten al usuario acceder a datos de manera efectiva.
- Más allá del acceso a los datos, se busca permitir a los usuarios contar con elementos que mejoren el análisis de los datos de acuerdo a sus preferencias o necesidades.
- Su uso debe ser sencillo, de modo que, independientemente de los conocimientos técnicos, las personas que vayan a usar las herramientas, puedan hacerlo sin complicaciones adicionales.

Conforme a las anteriores características, la elaboración del interfaz virtual planteado en este trabajo, obedece a las mismas y, además, de acuerdo con Britos (2008), ya que el nivel de complejidad del problema a abordar es alto, se clasifica dentro de la explotación de información o DM, puesto que, a través del uso de aplicaciones y herramientas, se espera aislar grupos, patrones o identificar elementos comunes que dentro del volumen de la base de datos, se busca, pueda aportar a la predicción de dificultades académicas en los estudiantes de décimo grado de las instituciones educativas, Ramón Alvarado y La Merced.

Al respecto, dentro del Data Mining, en los apartados siguientes se explicitan levemente las tecnologías basadas en análisis estadístico y en sistemas inteligentes, con el objetivo de dejar sentados los fundamentos por los que se usan elementos de ambas. Así, de acuerdo con el trabajo de Mayorga y Obregon (2016), se retoman desde la investigación, referentes de los autores relacionados con lógicas no lineales –o tradicionales– y métodos de agrupación o *clustering*, que constituye la parte predictiva que se obtendría al final del trabajo (Ver metodología).

4.2.1.3 La era de los datos, la información y el *machine learning*

El aumento de la conectividad y la abundancia de dispositivos inteligentes como computadoras y celulares, no solo ha cambiado la manera en que se organizan sociedades, también ha traído consigo un auge sin precedentes de la acumulación de datos, información y generación de conocimiento a partir de los mismos. Se hace importante así aclarar las concepciones respectivas para posteriormente ver cuál es el

papel del llamado machine learning tanto en el ámbito investigativo como en el campo social acaparado en los últimos tiempos.

En ese sentido, de acuerdo con Davenport y Prusak (1999, citado por Carrion, 2017), se debe reconocer que dato, información y conocimiento, no son concepciones congruentes, si bien es cierto, tienen características y relación inherente, cada uno de éstos tiene su particularidad. Es así como los *datos* podrían ser entendidos como simples registros o conjuntos discretos que representan hechos, sobre los que se debe tener muy en cuenta que en sí mismos, no tienen trascendencia o propósito más allá de su acumulada existencia. Por su parte, la *información* es un concepto más dicente, puesto que se podría entender como un mensaje con sentido y significado, los cuales serán interpretados por el receptor, básicamente, la información difiere de los datos debido a que debe tener un propósito inmerso. Con lo anterior se puede afirmar que los datos se convierten en información siempre que éstos sean dispuestos de tal manera que transmitan un mensaje, que permitan ser interpretados y por tanto conlleven juicios de valor sobre su utilidad. Por último, considerando el hecho que los datos pueden estructurarse o transformarse para que sean informativos, es la información también generadora de conocimiento, pero es éste último un concepto que va más allá de transformaciones, arreglos, registros, hechos, mensajes e interpretaciones; se trata de un concepto que abarca experiencias, rutinas u organizaciones, es en esencia la consolidación de esquemas más complejos con carácter útil, informativo, aplicativo y trascendente. Así, el conocimiento es un concepto ineludiblemente inherente al ser humano, puesto que es el conocedor, intérprete, el que hace conexiones y quien valora la funcionalidad respectiva ya sea mediante inferencias, discusiones, pruebas y demás herramientas que considere puedan servir para validar lo que potencialmente pueda ser conocimiento.

Considerando el elemento esquemático y formal del conocimiento con base a datos e información, a continuación, se representa la relación entre los conceptos.

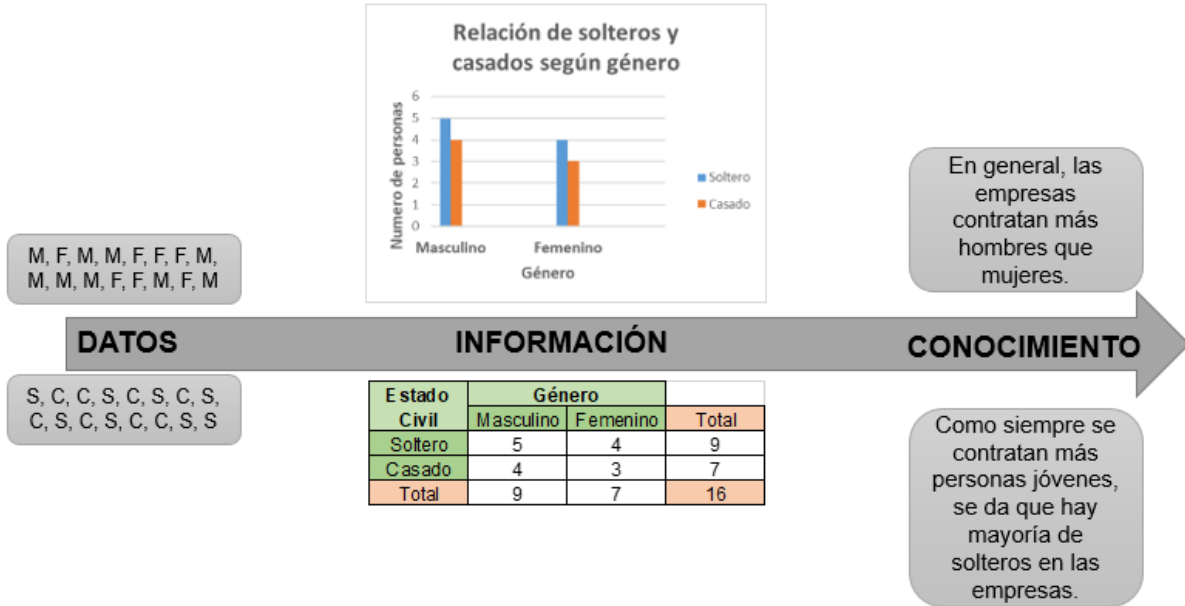


Figura 5. *Relación entre Dato, Información y Conocimiento*
Fuente: Elaboración propia

Con base a lo anterior, se evidencia la relevancia de los datos dentro del proceso de generación de conocimiento, al respecto, es de aclararse que, en la actualidad, debido al ya mencionado auge de la información por temas de conectividad y reorganización de la sociedad, la cantidad de datos abunda y es difícil para el ser humano realizar interpretaciones adecuadas debido a la naturaleza variable de los mismos y la complejidad inherente al contexto o problema del que se toman. Justo en ese ámbito surge el Data Mining y el Machine Learning. Éstos campos traen consigo luz ante la oscuridad aparente que subyace en amplios volúmenes de datos cuya existencia per se, se torna confusa, irrelevante y poco funcional.

Considerando el potencial de los datos y la información, para la generación de conocimiento, dentro de la necesidad de analizar y realizar tratamiento de grandes cantidades de datos, se retoma para esta parte el Data Mining y a continuación, con base al aporte de Ramírez et al. (2004), se explicitan los pasos a seguir para metodológicamente, implementar un modelo o algoritmo que permita extraer relaciones de datos diversos pero que caracterizan a estudiantes. A saber, dentro de las etapas se destaca:

Etapa 1: Determinar objetivos; se trata de identificar el tipo de información que se pretende sacar de la base de datos considerada en la investigación.

Etapa 2: Procesar los datos; consiste en seleccionar, limpiar, enriquecer, reducir, mitigar el ruido³ y transformar la base de datos que se tiene dentro de la investigación.

Etapa 3: Determinar el Modelo. Se considera la etapa más importante – dentro del presente trabajo – porque se trata de escoger el algoritmo a utilizar cuyo propósito es estudiar a profundidad una tarea específica, que debe obedecer a los objetivos esenciales de la investigación. Así, 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.

Etapa 4: Análisis de los resultados; Como en todo trabajo formal, es necesario realizar estudios de los resultados obtenidos una vez aplicado el algoritmo a la base de datos limpia, de modo que pueda validarse la efectividad o conveniencia del mismo dentro de la naturaleza del problema a abordar en la investigación.

Conforme a lo mencionado en la etapa 3, se hace necesario aclarar algunos aspectos a considerar dentro de las investigaciones, así, es de anotar que, dentro de los sistemas inteligentes, se han desarrollado métodos y tecnologías que pueden ser indispensables para múltiples trabajos. En ese sentido, se explicita que algunos métodos o tecnologías usados dentro del DM son de carácter netamente estadístico y entre estos, de acuerdo con Britos (2008), cabe destacar:

Análisis de varianza: útil para datos cuantitativos, es un método en el que se evalúa la existencia de diferencias significativas entre las medias de una o más variables continuas en diferentes poblaciones o espacios muestrales.

Regresión: permite establecer posibles relaciones entre una o más variables y un conjunto de variables predictoras de las primeras. Obedecen a la posibilidad de tendencias o normalizaciones.

³ Entendido éste como la posibilidad de datos incompletos, datos poco claros, entre otros.

Prueba Chi-cuadrado: considerada importante para realizar contraste de hipótesis de dependencia entre variables.

Análisis de agrupamiento: en esencia, resulta útil para la clasificación de una población de individuos que comparten distintos atributos (de distinta naturaleza) en un número determinado de grupos o clases, de acuerdo a criterios de semejanza o diferencias entre estos.

Análisis discriminante: Se diferencia de la anterior, en tanto permita la clasificación de individuos en grupos previamente establecidos, dando la posibilidad de encontrar la regla de clasificación de los elementos de estos grupos; así, se da una mejor identificación de cuáles son las características que definen los agrupamientos.

Series de tiempo: Su importancia estriba en el estudio de la evolución de una variable o característica a través del tiempo para poder realizar predicciones, con base a ese conocimiento y conforme al supuesto de que no van a producirse cambios estructurales o del contexto.

Ahora, dando el valor a los sistemas inteligentes para el tratamiento predictivo de los datos, se considera el Machine Learning más que como un subconjunto de la IA, como una disciplina que propicia la generación de dichos sistemas (Méndez y López, 2019), que aprenden bajo supervisión o sin supervisión para generar respuestas o resultados automáticos, útiles a diferentes campos del conocimiento y con la versatilidad inherente a su capacidad de suplir herramientas a la naturaleza variable de las múltiples investigaciones.

Por último, se hace necesario ubicar –de acuerdo a los fines de este trabajo– la investigación dentro de uno de los mencionados aprendizajes. Así, se destaca que, el aprendizaje no supervisado (*unsupervised ML*) consiste, principalmente en una manera de ver relaciones imperceptibles en las que son los algoritmos los que nos muestran a través del *clustering* qué posibles particularidades hay entre distintas agrupaciones y qué características tomadas dentro de las bases de datos no son relevantes para las tareas buscadas mediante la *dimensionality reduction*.

Por su parte, dentro del aprendizaje supervisado (*supervised ML*) a diferencia del anterior, se debe monitorear, entrenar y seleccionar atributos para que los algoritmos muestren el mejor rendimiento posible (Florez y Pulido, 2020). Dentro de éste tipo, de acuerdo a la naturaleza del problema de investigación, se distinguen *Regression*, si las salidas dentro de las bases de datos son numéricas, o *classification* si las salidas esperadas son categóricas; justo considerando esta última característica, es que el *supervised ML* y específicamente las herramientas de clasificación, se convierten en el referente a proyectar, puesto que se van a revisar, monitorear, interpretar y orientar mejoras en las respuestas dentro de la herramienta a usar, es decir, se elige el algoritmo considerando la pertinencia de que sea éste el que muestre las relaciones que, desde las limitaciones humanas, no se pueden extraer, pero de acuerdo con un buen tratamiento de datos y un buen rendimiento con base a métricas de incertidumbre que den fiabilidad al carácter valorativo inicial de las predicciones previstas sobre desempeños de estudiantes. La figura 6 ilustra la categorización de las herramientas ML y algunos de los algoritmos respectivos.

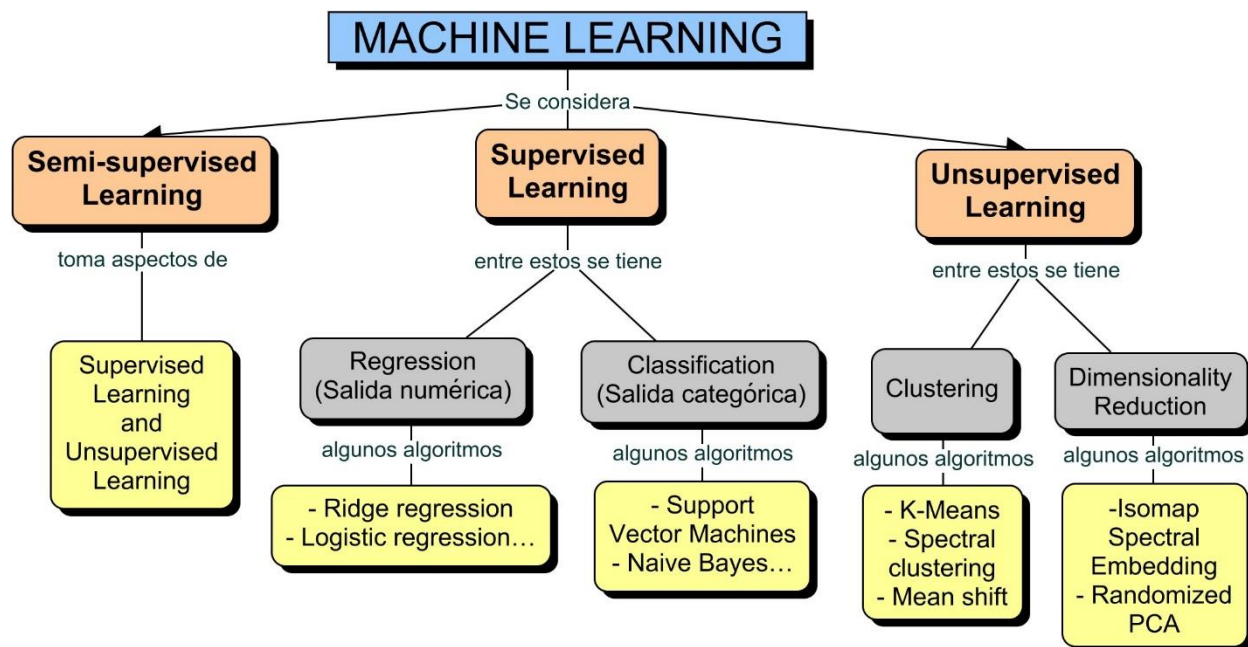


Figura 6. Tipos de aprendizaje ML
 Fuente: Modificado de Beunza, 2020

En la figura 7, además se destaca un diagrama de flujo que podría ser un referente orientativo para el trabajo con herramientas ML de acuerdo a las particularidades de los datos o la investigación per se.

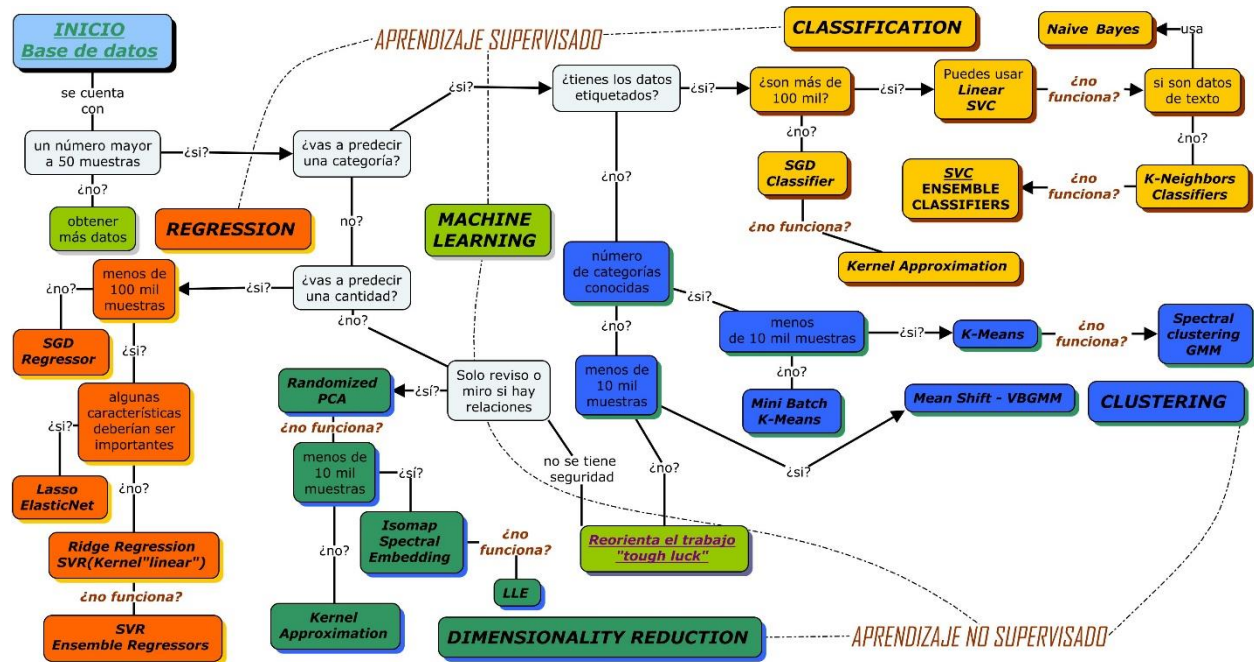


Figura 7. Diagrama de flujo para el uso de herramientas Machine Learning
Fuente: Modificado de Woiski (2015)

En el siguiente apartado se exponen las herramientas ML que posibilitarían la consecución del objetivo de la investigación.

4.2.1.4 Herramientas machine learning aplicadas a la educación

Se debe hacer claridad que el ML surge como disciplina cuya relevancia se centra en propósitos mercantiles y ante la necesidad de sacar provecho a la abundancia de datos que inundan las sociedades actuales. A ese respecto, y en concordancia con Quinlan (1985), en su trabajo *Induction of Decision Trees*, se evidencian sus potencialidades como herramienta práctica para ser aplicada a múltiples campos, en ese sentido, desde Méndez y López (2019), se destaca su utilidad dentro de compras online, para por ejemplo publicar un anuncio que promueva el consumo; detección de fraudes en transacciones, puesto que se pueden generar algoritmos que reconocen patrones dados dentro de estas irregularidades; predicción de fallos en elementos tecnológicos, mediante

la programación de algoritmos y soportes metodológicos; predicción del tráfico urbano, e incluso dentro del campo de la medicina, para realizar pre diagnósticos que consideren los síntomas de un paciente. Los anteriores y muchos más, son muestra de la aplicación del ML y del uso del DM. Dentro del campo de la educación, se busca incursionar considerando la complejidad del sistema per se y la necesidad de aportar mejoras que promuevan una orientación renovada administrativa y pedagógicamente hablando.

Con base a lo anterior, se debe hacer mención de las diferentes herramientas del ML, al para qué de las mismas y algunos de los algoritmos que dentro de la presente propuesta, se podrían considerar para la consecución del objetivo general. En la tabla 3, se sintetiza lo anterior.

Tabla 3. Categorización de herramientas ML y algoritmos respectivos

Categoría	Descripción	Objetivo	Algoritmos
Clasificación	Identificación de la categoría a la que pertenecen los objetos.	Detección de spam, reconocimiento de imágenes, patrones.	SVM, nearest neighbors, random forest, gradient boosting tree, Logistic regression.
Regresión	Predicción de atributos de valor continuos asociados a objetos.	Respuesta a medicamentos, precios de las acciones	SVR, ridge regression, Lasso
Regresión: Pronóstico	Predicción de aspectos cualitativos o cuantitativos en contextos principalmente empresariales	Encontrar patrones para prever comportamientos en ventas, acciones o características climáticas	Temporal series, exponential smoothing, moving averages
Agrupamiento	Agrupación automatizada de objetos similares o con atributo similares, en conjuntos.	Segmentación de clientes, agrupación de resultados del experimento.	K-Means, spectral clustering, mean-shift.
Reducción de dimensiones	Reducción de variables aleatorias por carencia de relevancia o relaciones significativas consideradas.	Visualización, mayor eficiencia.	PCA, feature selection, non-negative matrix factorization.
Selección de modelos	Comparación, validación y elección de modelos.	Mejora en la precisión debido al ajuste de parámetros.	Grid search, cross validation, mertics.
Procesamiento	Extracción y normalización de características.	Transformación de datos de entrada, como texto para usar algoritmos de ML.	Preprocessing, feature extraction.

Tomado y modificado de Méndez y López (2019)

Considerando las variadas herramientas y algoritmos dentro del ML, se hace necesario dentro de la etapa 3 del DM, además, seguir un riguroso proceso de selección, aplicación y validación del ML. Así, es imprescindible dentro de la investigación, seguir las fases propuestas por Pérez y Gonzalez (2007). Éstas se enuncian – de manera sintetizada – a continuación y su adecuada orientación repercute en la efectividad de la elección del modelo, herramienta o algoritmo a usar de acuerdo a la naturaleza de los datos enmarcados en el problema educativo que se aborda.

Fase 1. Entendimiento de los datos y del problema: si se dificulta la comprensión del problema y el pensar en las respuestas a las preguntas que sistematizan la problemática, difícilmente se podrá entender los datos y su suficiencia para considerar un modelo que pueda aportar a la solución.

Fase 2. Criterio de evaluación del modelo y de la solución: se hace necesario considerar el margen de error dentro de los posibles modelos ML a usar, por ello se requiere, en el proceso de selección, considerar el que presente el mínimo error, esto con el objetivo de elegir uno que sea pertinente conforme a la naturaleza del problema. Adicionalmente, posterior a su aplicación, dentro de la solución que ofrece, en pro de la mejora, previamente, es ideal pensar en cómo evaluar el rendimiento para vislumbrar adecuaciones.

Fase 3. Preparación de los datos. De acuerdo con los autores, es una fase de gran importancia debido a las múltiples situaciones que pueden emerger dentro de la investigación. Se hace necesario considerar entonces – entre posiblemente más elementos –:

- Datos incompletos, para lo cual se debe hallar una solución conveniente de cara al problema que se esté trabajando, así, la omisión de algunos, imputarlos o la búsqueda de una herramienta ML para manejar datos incompletos pueden ser soluciones a la situación.
- Combinar datos de varias fuentes, puesto que no siempre las bases de datos se encuentran en un mismo formato o en un mismo “sitio” por lo que se puede

considerar la combinación de estas fuentes para que el modelo ML trabaje de manera óptima.

- Dar formato a los datos, tomando en cuenta que se usarán librerías de ML, es recomendable usar matrices (o tensores) para vincular los datos.
- Calcular características relevantes, puesto que con ello se puede mejorar el rendimiento del modelo ML. Es decir, siempre que sea posible, es mejor ponderar en lugar de suministrar datos puros.
- Normalización de los datos, para se manejen en escala similar.

Fase 4. Construcción y análisis del modelo: Considerando la preparación adecuada de los datos, se requiere ahora, elegir el tipo de herramienta ML que, en definitiva, conforme a la fase 1 y 2 va a usarse para resolver el problema. Posterior a la aplicación del algoritmo, se procede a analizar los errores que se presenten para valorar la adecuación que conlleve a la mejora del mismo. Respecto a esto último, puede considerarse el cambio del modelo, uso de modelos más simples o complejos e incluso, la disposición de más datos. Todo esto está sujeto al marco del problema abordado.

Fase 5. Modelo integrado de un sistema: al considerar, con base al análisis de errores, si el modelo elegido es pertinente, se debe integrar este al sistema sobre el que se esté trabajando (bien puede ser una plataforma, interfaz gráfico, simulador, etc.), con ello se permite al modelo comunicarse con otras partes del sistema y seguir generando respuestas automáticas que deben monitorearse para verificar que no se genere un aumento progresivo en los errores emergentes.

Por último, se debe tener claridad que las fases anteriores son orientativas pero que la efectividad en el uso de herramientas o modelos ML, está sujeta a la comprensión no solo de la problemática, sino del porqué de los algoritmos elegidos y de la interpretación de los resultados que éste arroje.

4.2.1.5 Algoritmos de clasificación y medición de incertidumbre

Como ya se mencionó en los apartados anteriores y se ilustró en la figura 2, dentro del ML hay tipos de aprendizaje que se basan en si hay supervisión, no la hay, o si se toman aspectos de los dos anteriores. Considerando el objetivo general del presente trabajo, se

hará énfasis en los algoritmos de clasificación debido a la necesidad de relacionar variables de entrada para predecir categorías de desempeño académico en competencias básicas. A saber, desempeños superiores, altos, básicos y bajos. Con base a lo anterior y de acuerdo con la evidencia [Sucar y Tonantzintla (2006), Correa et. al. (2008), Méndez y López (2019); Bausili (2021); Jones- Ortiz y Guzmán – Seraquive (2022); Joaquín (2020)] a continuación se mencionan los algoritmos de clasificación más usados a considerar dentro de la investigación.

Árboles de decisión: O Decision tree, por su nombre en inglés, es una herramienta de clasificación que se fundamenta en la entropía de Shannon, entendida ésta –dentro del ámbito informático– como una medida promedio de la incertidumbre, de acuerdo a probabilidades de ocurrencia de uno u otro evento, de ahí que los árboles de decisión usen el concepto o el índice de Gini, que fundamentalmente –dentro del contexto informático– es un valor de pureza en los nodos que relaciona el algoritmo (que también se podría entender como: qué tan a menudo se etiqueta incorrectamente un elemento). En última instancia, la herramienta permite estructurar y mostrar gráficamente cómo se subdivide y relaciona la información de acuerdo a las categorías de las variables de entrada y cómo se llega a las respectivas salidas. Éstos se componen de Nodos, para el valor de un determinado atributo, Ramas, orientación dependiente entre nodos o de nodo a hoja y, esta última, justo muestra la clasificación de acuerdo a las categorías de la variable de salida. Podrían entenderse las hojas como el nodo de predicción o etiqueta final. La figura 8 ilustra el uso de un árbol de decisión desde la aplicación del algoritmo J48⁴ en el software Weka. Los datos de entrada están relacionados con los estudiantes objeto de estudio de la investigación y se tomaron a modo de ejercicio, para empezar a discernir sobre las relaciones de variables socio demográficas con -para éste ejercicio-, el haber repetido años anteriores.

⁴ Que, a su vez, es la implementación en WEKA, del algoritmo C4.5, desarrollado por Quinlan.

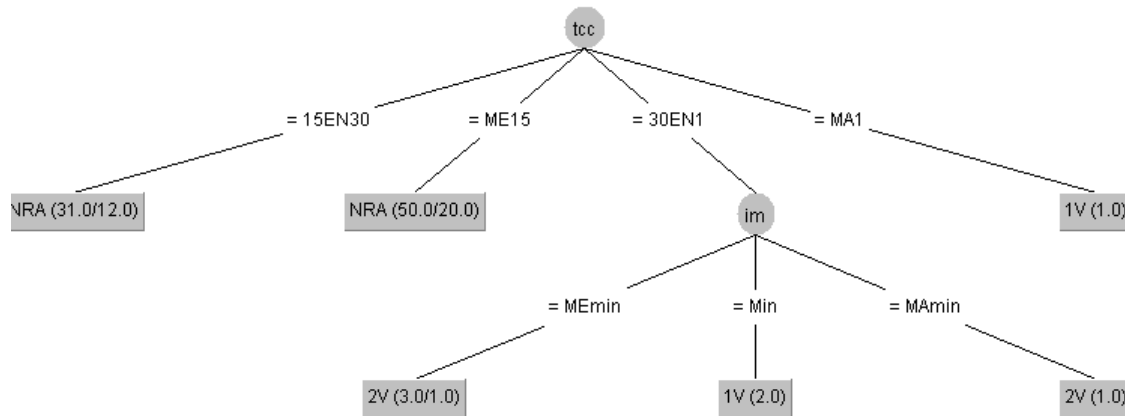


Figura 8. *Árbol de decisión usando el algoritmo J48 en Weka*
 Fuente: Elaboración propia

Regresión Logística: se trata de un algoritmo simple de clasificación en el que se tienen salidas binarias que se predicen con base a entradas cuantitativas. Se convierte un algoritmo ideal para asociar relación entre atributos de un objeto y el grupo del mismo dentro del cual clasificarlo. Debe recalarse que, este, opera con base a una función que calcula el logaritmo de la probabilidad y es dicha función la que permite entender que su operatividad, aunque pueda ser sencilla, puede permitir vislumbrar si ciertas características (atributos) tienen o no relación con las clases objeto (a predecir). De ahí que se use en el presente trabajo como primera herramienta a valorar para considerar relaciones entre aspectos sociodemográficos y desempeños.

Redes Bayesianas: Son representaciones gráficas en las que, a partir de grafos acíclicos, se muestran relaciones de interdependencia condicional. Así, los nodos muestran variables aleatorias y los arcos dan muestra de las dependencias entre éstas. Dentro de estas redes se resalta que simplifican razonamientos e interpretaciones de interrelación de variables al interior de grandes volúmenes de datos, a partir de condiciones probabilísticas expresadas matemáticamente, a partir variaciones del teorema de Bayes, es decir, de variaciones del mismo, conforme a las variables y parámetros ingresados. En la figura 9, se muestra un ejemplo de aplicación en el que se relacionan claramente la actividad física (ACTFIS), género (GE), la edad (ED), peso, preferencia de deporte

(DEP), el consumo de frutas y verduras (FRUTYVERD) y los lugares (LUGAR) cercanos para hacer ejercicio.

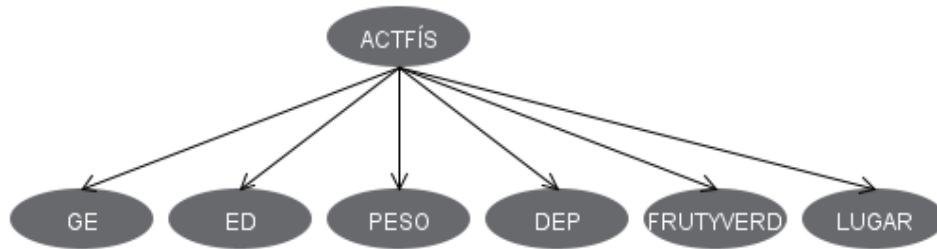


Figura 9. Red bayesiana usando Weka
 Fuente: Elaboración propia

Redes neuronales: Así como las neuronas del cerebro se interconectan y se comparten información por medio de la sinapsis, las redes neuronales obedecen a este esquema biológico. Solo que las interconexiones, para estas redes artificiales, permiten que fluya la información de manera interna, para generar una salida que se considera el final de la red. Con base a este principio, las redes se disponen a través de capas que van de las entradas, pasan por unas intermedias u ocultas y concluyen en una capa final o de salida. En la figura 10, se muestra el funcionamiento general de las redes neuronales.

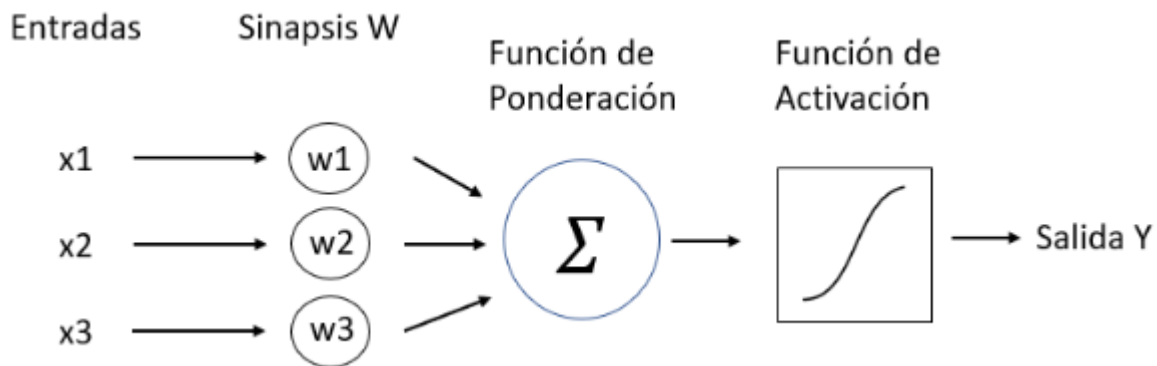


Figura 10. Representación del funcionamiento de una red neuronal.
 Fuente: Tomado de Castro (2020)

Random Forest: Es un clasificador dentro del aprendizaje supervisado que se fundamenta en la generación de árboles de decisión y que permite inferir, clasificar, predecir o mitigar relaciones dentro de un amplio conjunto de datos. Dentro de las bondades a considerar de ésta herramienta, se encuentra el que sean adecuados para

trabajar grandes volúmenes de datos sin excluir variables introducidas y dando, además, una estimación de las que podrían ser relevantes en la clasificación efectuada. En la figura 11, se ilustra la interpretación del mencionado clasificador.

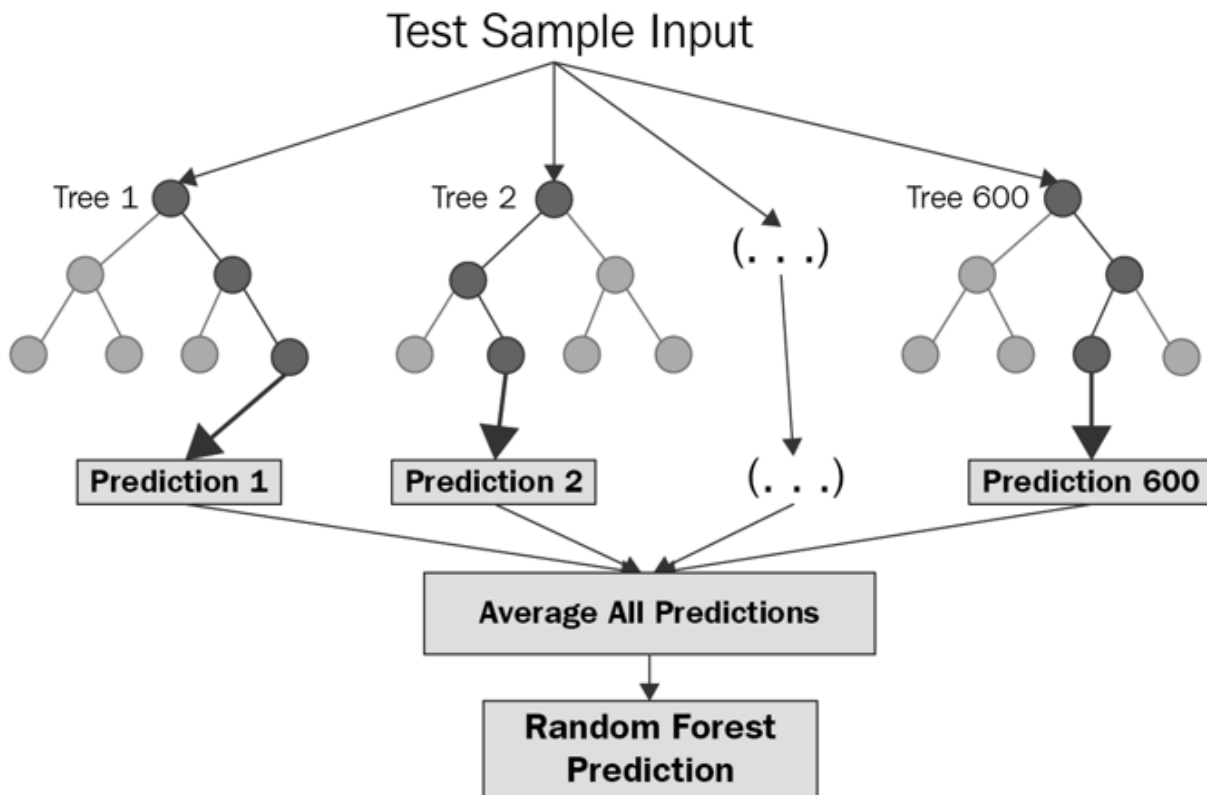


Figura 11. Interpretación del *RandomForest*.

Fuente: <https://corporatefinanceinstitute.com/resources/knowledge/other/random-forest/>

Máquinas vectoriales de soporte: también denotadas por su sigla SVM (por su nombre en inglés *support vector machines*), en sentido general, son un conjunto de algoritmos propios del aprendizaje supervisado que permiten, a través de vectores n-dimensionales, clasificar un “punto” nuevo (o clase) en una u otra categoría, de acuerdo al módulo vectorial respectivo. Lo anterior, mediante la generación de hiperplanos que, separan las clases previamente alimentadas y así, cuando un elemento nuevo –por ejemplo, de una nueva muestra– se introduce, éste se etiqueta o clasifica con base al vector soporte que marca la distancia más cercana a una u otra. En la figura 12, intuitivamente se puede representar el funcionamiento de este clasificador para una situación binaria.

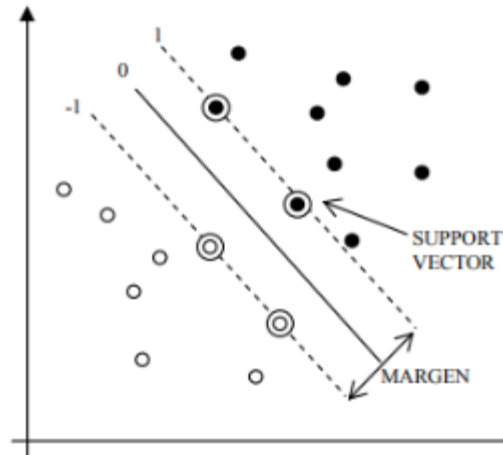


Figura 12. Representación intuitiva de un SVM.
 Fuente: Tomado de Castro (2020)

K Nearest Neighbors: o en español, k vecinos más cercanos, es otro algoritmo de clasificación dentro del aprendizaje supervisado que funciona en sentido general, de manera sencilla. Consiste en clasificar valores con base a similitudes con clases o atributos introducidos previamente. Esto último, mediante alguna función que se considere pertinente en el trabajo de entrenamiento. En ese sentido, de lo que trata el algoritmo, de manera intuitiva, es, en etiquetar nuevos datos con base a similitudes con las clases más cercanas. Así, se puede considerar una clasificación dentro de una categoría con base por ejemplo una función de distancia que considere las variables más similares –o cercanas– al nuevo elemento suministrado. A pesar de que se torna útil, tiene falencias en cuanto el volumen de datos tenga muchas variables y suficientes que puedan ser irrelevantes, puesto que, se podrían generar cercanías con aspectos intrascendentes para los propósitos del trabajo investigativo. En la figura 13, se representa el funcionamiento general del clasificador. Allí se pone como ejemplo, ¿cómo clasificar el elemento rojo con base a los 10 vecinos más cercanos? Pues bien, en sentido amplio, el modelo establece relaciones y clasifica nuevos elementos con base a similitudes de cercanía con otros ya etiquetados.

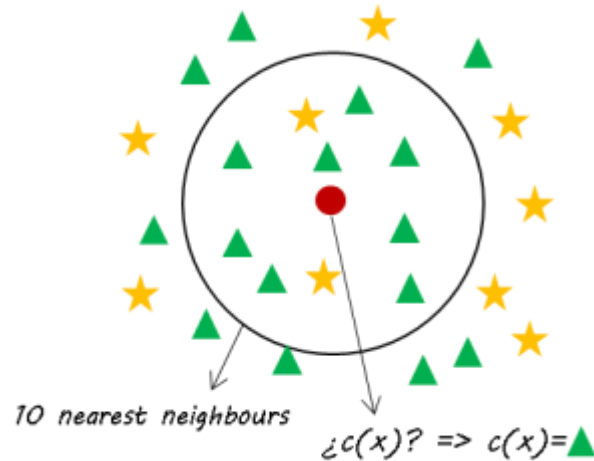


Figura 13. Representación de funcionamiento del K Nearest Neighbors.
 Fuente: <https://quandare.com/vecinos-serie-temporal/>

Gradient boosting tree: Es una herramienta que dentro del ML permite, a través de árboles de decisión, dispuestos y entrenados secuencialmente, mejorar aspectos errados de los predecesores para generar clasificaciones -principalmente- mucho más certeras. En ese sentido, si se quiere clasificar un elemento nuevo, el modelo toma como fundamento los tratamientos dados por todos los árboles que lo conforman. Es de destacar de ésta herramienta, que evidencia una interacción compleja entre diferentes árboles (predictores) para brindar clasificaciones en las que se podrían apreciar relaciones entre las variables que conforman la base de datos. No obstante, presenta como posible desventaja, justo algo contraproducente al uso de varios árboles, esto es, dificultad para interpretar el modelo debido a las relaciones complejas inmersas. En la figura 14, se muestra la estructura general e intuitiva del funcionamiento de este clasificador.

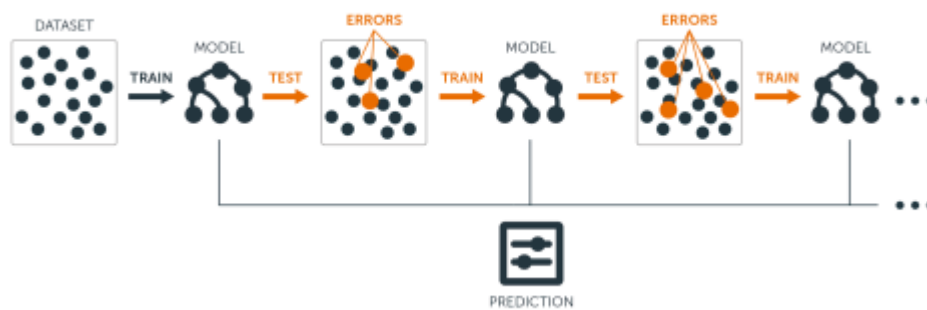


Figura 14. Representación del funcionamiento del Gradient boosting tree.
 Fuente: <https://machine-learning.paperspace.com/wiki/gradient-boosting>

Ahora, conforme se elige objetivamente uno de los anteriores algoritmos, se debe, consecuentemente, justificar su elección con base a métricas de incertidumbre o “testeo”, éstas, aunque son menos abundantes que la cantidad de algoritmos, con base a las investigaciones [Castro (2020); Murcia (s.f.); Jones- Ortiz y Guzmán – Seraquive (2022); Bausili (2021); Méndez y López (2019); Gil-Vera y Seguro-Gallego (2022); Gamarra (2020); Borja et. al. (2020); Ortíz (2021); Gatica (2020); Fuentes (2022); Florez y Pulido (2020); Ariza-López et. al. (2018); Martínez y Palencia (2021); De Ullibarri y Pita (1999)] se hará un comparativo (Ver capítulo de resultados) del rendimiento de los algoritmos de acuerdo a las siguientes:

Matriz de confusión: cuando de clasificadores se trata, las matrices de confusión son usualmente consideradas para las métricas de desempeño de los algoritmos o herramientas ML aplicadas. Éstas consisten en lo que en apariencia se ve como una tabla de doble entrada, pero que para ser más precisos realmente se trata de una representación matricial en la que se evalúan los aciertos del clasificador. En ese sentido, se estipulan cuatro posibilidades, dos que muestran clasificaciones correctas y dos que evidencian lo contrario. A continuación, en la tabla 4, se explicita la estructura de una matriz de confusión.

Tabla 4. *Matriz de confusión*

Observaciones	Predicciones	
	Positivos	Negativos
Positivos	Verdaderos Positivos	Falsos Negativos
Negativos	Falsos Positivos	Verdaderos Negativos

Modificado al español, de Ariza-López et. al., (2018). Pp. 21

Ahora, las relaciones dadas en la matriz, manifiestan:

- Verdaderos positivos (VP), dan muestra de las clasificaciones positivas correctas.
- Verdaderos negativos (VN), evidencian las clasificaciones negativas realizadas correctamente.
- Falsos positivos (FP), obedecen a clasificaciones positivas erradas por el algoritmo usado.
- Falsos negativos (FN), enmarcan las clasificaciones negativas incorrectamente predichas por el algoritmo.

Con base a lo anterior, una matriz de confusión que, para determinado algoritmo, muestre valores mucho más altos en la diagonal secundaria (FP y FN), que, en la primaria (VP y VN), deja de manifiesto que, es poco acertado en las predicciones que ha generado.

Accuracy: Conforme a las relaciones mostradas en la matriz de confusión, surgen otras métricas que dan muestra del rendimiento del algoritmo de clasificación. Así, esta métrica relaciona las clasificaciones correctas (VP y VN) respecto al total de instancias (VP, VN, FP y FN). La figura 15, muestra la relación que, oscila entre 0 y 1 y que, entre más cerca de 1 esté, mejor será el rendimiento del clasificador.

$$Acc = \frac{VP + VN}{VP + VN + FP + FN}$$

Figura 15. *Expresión para la métrica Accuracy*
Fuente: Castro, 2020

Es importante señalar que, la métrica expuesta, resulta útil, pero puede dar orientaciones confusas cuando por ejemplo los datos no están equilibrados y por ejemplo se toman altos volúmenes de éstos en torno una variable principal y muy pocos de variables secundarias. Por ello, de acuerdo a la base de datos trabajada, es importante considerar ese aspecto.

Precision: esta métrica, permite cuantificar, en términos de porcentaje, las predicciones positivas que el algoritmo generó correctamente. La figura 16, enmarca la expresión para esta métrica en una situación de clasificación binaria.

$$Pre = \frac{VP}{VP + FP}$$

Figura 16. *Expresión para la métrica precisión*
Fuente: Castro, 2020

Tanto esta métrica, como las dos que se enuncian a continuación, resultan favorables para tratar de evaluar mejor el desempeño de los algoritmos ML, especialmente cuando los dos datos se tornan desequilibrados.

Recall: también conocida como exhaustividad, es una métrica que permite hallar el porcentaje de casos positivos clasificados correctamente por el algoritmo ML y que en la figura 17, se enuncia como el cociente entre los VP y VP + FN.

$$Re = \frac{VP}{VP + FN}$$

Figura 17. *Expresión para la métrica recall*
Fuente: Castro, 2020

F1-Score: Como métrica que permite una valoración más adecuada respecto a los trabajos en los que los datos no están equilibrados, el F1-Score, da muestra del porcentaje de clasificaciones correctas enfatizando en las clases positivas. Así, esta métrica toma como referencia las dos anteriores y por medio de una media armónica (ver figura 18) genera un valor que puede ser más significativo, de acuerdo con las características propias de la base de datos usada en el trabajo y tratada con el algoritmo ML respectivo.

$$F1 - S = \frac{2 * VP}{2 * VP + FP + FN}$$

Figura 18. *Expresión para la métrica F1-Score*
Fuente: Castro, 2020

Índice Kappa: Es imprescindible mencionar que, éste índice (K) fue propuesto inicialmente por Cohen a inicios de los 60 y es usado en la clasificación de variables categóricas con no menos de dos categorías de salida. Su uso consiste en mostrar la concordancia de herramientas ML en la clasificación de este tipo variables haciendo uso de la métrica Accuracy ya antes mencionada y en esencia, es un valor que permite definir el porcentaje o proporción de acuerdos observados que trascienden de la aleatoriedad propia implícita en las herramientas aplicadas. En ese sentido, se convierte en un parámetro interesante para considerar el asertividad de un algoritmo o simplemente para cuestionar el rendimiento o efectividad en las clasificaciones del mismo. La figura 19, denota la expresión con la que se obtiene dicho índice y la tabla 5, los rangos de valoración de la fuerza de concordancia.

$$k = \frac{P_0 - P_e}{1 - P_e} \quad \left| \begin{array}{l} P_0: \text{proporción de acuerdos observados o accuracy} \\ P_e: \text{proporción de acuerdos dados por azar} \end{array} \right.$$

Figura 19. Expresión para hallar el índice Kappa

Fuente: Tomado de Jones- Ortiz y Guzmán – Seraquive (2022)

Tabla 5. Valoración de concordancia por el índice *k*

Valor de <i>k</i>	Fuerza de concordancia
<0.20	Pobre
0.21 – 0.40	Débil
0.41 – 0.60	Moderada
0.61 – 0.80	Buena
0.81 – 1.00	Muy buena

Tomado de, De Ullibarri y Pita (1999)

AUC - ROC: El área bajo la curva ROC (Receiver Operating Characteristic: Característica operativa del receptor) o AUC (por sus siglas en inglés: Area Under the ROC Curve), es una métrica interesante que permite comparar rendimientos de clasificadores ML de acuerdo a su potencial para diferenciar clases positivas y negativas desde la amplitud de los umbrales posibles. Es una métrica que se remonta a la segunda guerra mundial y fue introducida esencialmente para detecta en el radar, puntos que pudieran ser enemigos. No obstante, la métrica ha servido a distintos campos de aplicación entre los que la medicina ha sido una de las principales beneficiarias en tanto se ha considerado el uso de este método, para la valoración de diagnósticos. Dentro del campo informático actual también tiene cabida, en tanto se ha expandido el aspecto predictivo no solo a las enfermedades sino también a las actitudes comportamientos, o en general a la toma de decisiones y categorizaciones.

Así, como se muestra en la figura 20, el ejemplo de uso del AUC, permite, desde el análisis de especificidad y sensibilidad de un algoritmo, comparar rendimientos en los datos de entrenamiento y de prueba.

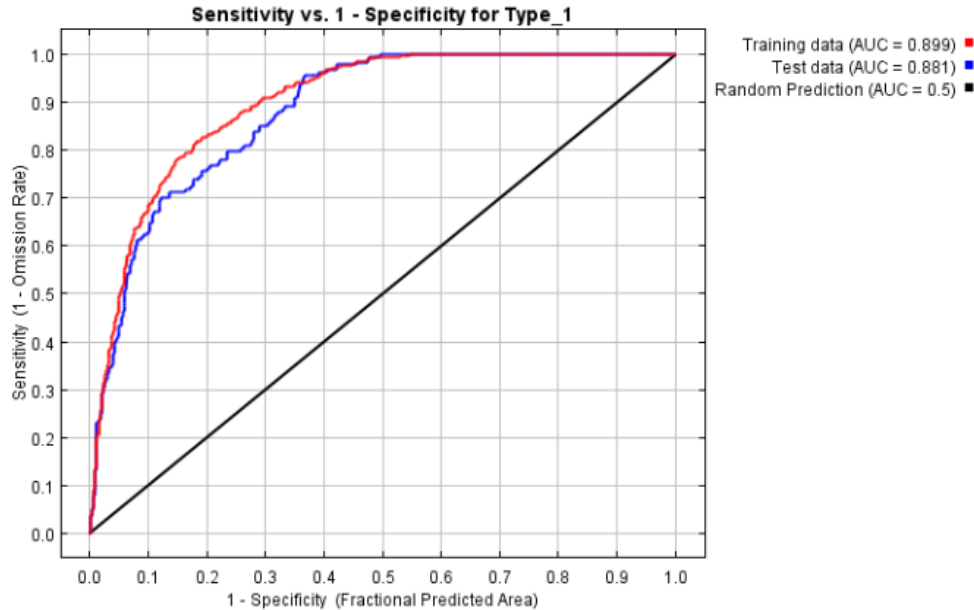


Figura 20. AUC – ROC
Fuente: Tomado de Murcia Ramírez, J. (s.f)

Ahora, para comprender los valores suministrados por la métrica, se han establecido rangos de valoración que permiten orientar las conclusiones al respecto del rendimiento de las herramientas. La tabla 6, enmarca dichos rangos como ayuda a la interpretación inherente a su uso.

Tabla 6. Rangos de valoración para el AUC – ROC

Valor	Consideración del Test
≤ 0.50	Deficiente
0.51 – 0.60	Malo
0.61 – 0.75	Regular
0.75 – 0.90	Bueno
0.91 – 0.97	Muy bueno
0.97 – 1.00	Excelente

Modificado de Concejero (2004)

Por último, de acuerdo a la relación entre especificidad y sensibilidad sobre las que se fundamente la métrica AUC – ROC, a continuación, se enuncia cada una de estas.

Sensibilidad: en términos de la matriz de confusión -previamente enmarcada-, la sensibilidad se puede entender como la proporción de verdaderos positivos clasificados así, de entre todos los que realmente lo son. En la figura 21 se explicita la expresión.

$$Sen = \frac{VP}{VP + FN}$$

Figura 21. *Expresión para la métrica sensibilidad*
Fuente: Tomado de Concejero (2004)

Especificidad: Por su parte, la especificidad representa la proporción de VN clasificados respecto a todos los negativos efectivamente dados. Se puede entender de este modo, una reciprocidad aparente entre las métricas pero que, de acuerdo a lo señalado anteriormente, de manera compleja, juntas, permiten generar una métrica interesante como es el caso AUC – ROC. La figura 22 explicita la relación enunciada.

$$Esp = \frac{VN}{VN + FP}$$

Figura 22. *Expresión para la métrica especificidad*
Fuente: Tomado de Concejero (2004)

4.2.2 Importancia y dificultades en la enseñanza-aprendizaje por competencias. Conforme avanza el tiempo, la humanidad evoluciona en términos de globalización y mejoras tecnológicas. La educación, por su parte, de acuerdo con la experiencia mostrada en diferentes lugares del mundo, no ha evolucionado de igual manera. No obstante, se han empleado diferentes modelos dentro del sistema educativo que promuevan la formación y la transmisión de costumbres culturales socialmente aceptadas. Enseñar no garantiza el aprendizaje y en torno a la necesidad de por lo menos lograr formación de personas competitivas que aporten a la sociedad, desde hace ya un par de décadas, se ha adoptado el aprendizaje por competencias el cual busca un desarrollo autosostenible de los grupos sociales (Murga – Menoyo, 2015). En Colombia, tratando de avanzar en la mejora de la educación, se ha recurrido al mismo enfoque y para efectos de considerar aspectos positivos y posibles dificultades al respecto, en los siguientes apartados se trata lo relacionado a cierta taxonomía de las competencias en el campo educativo y lo inherente a la importancia y dificultades del enfoque llevado a la práctica.

4.2.2.1 Competencias básicas, genéricas y específicas

Previo a presentar las diferentes competencias que se pueden categorizar de acuerdo a su cobertura, dimensiones o campos de acción. Se debe considerar de modo amplio, qué atañe al término competencia.

Así, es importante recalcar los siguientes aportes:

Retomando autores de un par de décadas atrás (años de auge del enfoque), es importante mencionar que, para algunos de ellos (Bunk, 1994; Bogoya, 2000; Oullet, 2000), las competencias en el campo de la educación se entienden como el uso de conocimientos para resolver problemas particulares, es una manera de referirse al saber hacer en contexto de acuerdo a la acepción de haber interiorizado saberes que promueven la creatividad, operatividad y productividad en el entorno.

Para Tobón (2003), a pesar de valorar las posibles buenas consideraciones anteriores, tal concepción debe ir más allá del aspecto mercantil y economicista con el que se ha adoptado. En ese sentido, critica con argumentos válidos, la carencia de incorporación dentro de las definiciones, la importancia de aspectos como la convivencia y la solidaridad. Con base a ello, aporta un referente desde el pensamiento complejo relacionado con el término competencia desde una perspectiva más integradora y menos reduccionista. Al respecto menciona que se debe "...conceptualizar las competencias como procesos complejos que las personas ponen en acción-actuación-creación, para realizar actividades sistémicas y resolver problemas laborales y de la vida cotidiana, con el fin de avanzar en la autorrealización personal, vivir auténticamente la vida y contribuir al bienestar humano" (pp. 10).

Por otro lado, valorando el aporte anterior y el ánimo bien intencionado de United Nations Educational, Scientific and Cultural Organization (UNESCO), es importante considerar que para el 2015 la UNESCO ha promovido una mejora de la calidad educativa a nivel mundial y para ello, dentro de distintas estrategias, han establecido concepciones respecto al proceso de enseñanza – aprendizaje por competencias. Así, en concordancia con los principales aspectos expuestos en el documento *Repensar la educación ¿hacia un bien común mundial?* (UNESCO, 2015), se recalca desde los propósitos del presente

trabajo, la necesidad de reestructurar políticas educativas que, entre otras cosas, den relevancia al mundo complejo en el que se relacionan los diferentes agentes que intervienen en los procesos educativos.

En ese sentido, es imperativo que, la visión mercantilista que desde PISA y la OCDE (organismo al que Colombia se vinculó desde el 2020) se promueven en términos formativos, sea reinventada desde una orientación más consciente respecto a la desconexión entre las enseñanzas formales y las necesidades que demandan los empleos, además, si de competencias se habla, consecuentemente es pertinente formar personas que compitan consigo mismos, que se autorregulen, que se valoren y sean emocionalmente inteligentes, es decir, las nuevas políticas deberían hacer énfasis en aspectos competitivos más allá de la dimensión cognoscitiva.

Con base a lo anterior, se considera importante resaltar cierta distinción que va más allá de la idea de competencia como el solo hecho de saber hacer en contexto, se trata también de saber ser, saber convivir y en general de saber orientarse para tomar decisiones que conlleven plenitud a favor del desarrollo personal y social.

De acuerdo con Méndez (2007), se puede diferenciar entre competencias básicas, disciplinares, interdisciplinares, transversales, generales, entre otras. No obstante, se retoman las básicas, genéricas y específicas (también llamadas disciplinares por diferentes autores), conforme al sistema educativo colombiano. Al respecto, se destaca desde la autora, la importancia de las *competencias básicas* y se definen éstas como las estrictamente necesarias para asumir nuevos aprendizajes, es decir, se trata de competencias fundamentales para el desarrollo de habilidades y destrezas posteriores.

Sobre las *específicas*, se considera que son los saberes puestos en práctica, necesarios para resolver situaciones de un campo disciplinar en particular. Por su parte, las *genéricas* las enmarcan como elementos a movilizar para la puesta en práctica de saberes dentro de la resolución de situaciones, a saber, se destacan: 1. Analizar y sintetizar; 2. Aprender; 3. Resolver problemas; 4. Uso de conocimientos en la práctica; 5. Adaptarse a diferentes situaciones; 6. Preocuparse por la calidad; 7. Practicar las técnicas de la información y la comunicación; 8. Trabajar de manera autónoma; 9.

Trabajar en equipo; 10. Organizarse y planificar. Las anteriores se toman como síntesis, puesto que en trabajos como el de Solanes et al. (2008), se consideran varias más distribuidas entre instrumentales, interpersonales y sistémicas, pero que concurren en las 10 ya mencionadas.

Con base a lo anterior, considerando la particularidad de la educación en Colombia, se debe destacar que hay un enfoque por estándares de competencias enmarcado en Schmidt (2006), como una guía clara que busca orientar los procesos educativos desde el desarrollo de competencias básicas y ciudadanas que promueven la academia y el desarrollo social pertinente de los estudiantes de educación básica y media. Así, el presente trabajo se centra en considerar la posibilidad de predicción de las dificultades que podrían presentar los alumnos en este tipo de competencias, de acuerdo a consideraciones sociodemográficas que yacen en la singularidad y complejidad relacional de cada uno.

4.2.2.2 Importancia de la formación por competencias básicas

Es interesante partir de una de las contribuciones del enfoque competencial en la educación y traer a colación el proyecto Tunning cuyo desarrollo ha permitido en el marco de la educación a nivel internacional –principalmente europeo–, concurrir en la comparación para el reconocimiento de titulaciones académicas no solo en diferentes universidades, sino también en diferentes países, de acuerdo a lo que denominaron “Sintonizar las estructuras educativas en Europa” (Bezanilla, 2003), de ese modo se crea el European Credit Transfer System (ECTS) o Sistema Europeo de Transferencia de Créditos y surgen los créditos educativos en las instituciones de educación superior, en pro de la mejora educativa y por tanto del nivel competencial inherente a los egresados de todas las facultades.

Con base a este referente de inicios del presente siglo, en Colombia, se adopta el enfoque en pro de una reforma educativa que, oriente la formación de personas en etapa escolar desde un saber hacer flexible en el que los contextos son variados por lo que la interiorización de lo aprendido, ha de generar las capacidades para resolver situaciones fuera de las aulas de clases (MEN, 2006).

Se resalta así dos elementos: 1. La necesidad de la mejora educativa de modo que esta se centre en el aprendizaje y 2. Que el aprendizaje trascienda de la tradicional memorización para la repetición de contenidos. De este modo, se señalan a continuación, aspectos de relevancia dentro del enfoque y desde el propósito de desarrollar competencias básicas, enmarcados en Perroun (2008), Rodríguez et. al. (2007) y MEN (2006).

- Implican que al final de un proceso formativo, la persona pueda desenvolverse en los diferentes ámbitos de la vida (personal, intelectual, social, laboral, ciudadana) con base al desarrollo de conocimientos, habilidades y actitudes fomentadas en las instituciones.
- Promoción de personas con sentido humano y moral para que actúen de manera constructiva tanto colectiva como individualmente y que aporten a la sana y pacífica convivencia.
- Conlleva considerar el aprendizaje más allá de solo conocimientos y vislumbrar en su lugar, operaciones mentales complejas entre las que destacan las relaciones, interpretaciones, inferencias, asociaciones, capacidad inventiva y creativa para la solución de situaciones y toma de decisiones.
- Suscita cuestionarse desde las instituciones, la docencia y las prácticas pedagógicas, sobre la utilidad real de las disciplinas escolares en términos de trascendencia en la cotidianidad.
- Fomenta la integración de conocimientos en tanto el estudiante se enfrente a situaciones complejas propias de su vida o las recreadas en el aula de clases.
- El tomar las Ciencias Sociales, las Ciencias Naturales, las Matemáticas, el Lenguaje y las competencias ciudadanas desde los desempeños, desde el hacer, desde la utilidad, los valores, el trabajo en equipo y la trascendencia de los procesos, entornos, contextos y la vida, antes que los saberes aislados –tipo enciclopedia–, genera formación integral en los que deberían ser, agentes activos de cambio para la sociedad.

Lo anterior no solo manifiesta consideraciones de relevancia desde el enfoque competencial, también deja latentes las dificultades inherentes a la aplicación efectiva

del enfoque. Por ello, a continuación, se explicitan las dificultades que subyacen en el anhelo de una mejor formación para los estudiantes en los niveles de educación media y secundaria.

4.2.2.3 Dificultades en la formación por competencias básicas

Dentro de un nuevo enfoque educativo que en su momento constituyó una reforma educativa, hoy por hoy prevalece la necesidad de cambiar el tradicionalismo y la hegemonía disciplinaria que ha fundamentado el sistema educativo y por tanto los currículos de las instituciones educativas en todos los niveles. En palabras de Perrenoud (1999): “El enfoque por competencias sólo se opone a la cultura general si se le da a esta última una orientación enciclopédica. La recuperación del valor instrumental de los contenidos constituye una de las tareas centrales del diseño y el desarrollo curriculares”

Con base a lo anterior y la evidencia investigativa [Velásquez y Hernández (2016); Perrenoud (2008); Fernández et. al. (2016)] se consideran algunas dificultades que subyacen en el enfoque competencial y que, como fundamento para la presente investigación, es importante resaltar.

- Considerando el hecho que las competencias se fomentan a partir del hacer, la práctica y sobre todo el contexto, se puede considerar dificultoso el que, se encuentren las transposiciones didácticas pertinentes a los diferentes campos cognoscitivos -junto a los tópicos respectivos- para que los alumnos, por medio de su propia activación y reconstrucción, los interiorice.
- Desde la docencia, el que se deban elegir situaciones concretas dentro de contextos singulares, respectivamente, para propiciar los espacios que generen la formación de competencias, puede hacer que, dentro de las prácticas pedagógicas, implícitamente, se transmita una visión cultural, social y personal, propia del docente. Limitando que el estudiante cree a través de su libre reflexión, sus propias perspectivas de vida.
- Cuando se presente resistencia al cambio, o impedimentos personales para llevar al aula situaciones concisas y pertinentes para el fomento de las competencias, puede retomarse el tradicionalismo y recurrir a las prácticas que por mucho tiempo prevalecen en las instituciones. Tal es el caso de la transmisión de contenidos,

fórmulas o métodos, antes que espacios para por ejemplo el desarrollo del pensamiento crítico.

- Se promueve, desde el enfoque, un aprendizaje –y por tanto una enseñanza– más interdisciplinaria y transdisciplinaria que específica y aislada dentro de los campos básicos de conocimientos (Matemáticas, Lenguaje, Ciencias, Ciudadanía). Esto conlleva interacción, planeación e innovación por parte de los docentes para romper las barreras de las propias disciplinas.

Con el propósito de considerar las singularidades propias de cada estudiante, las anteriores son solo las dificultades más concurrentes encontradas y que, se tornan coherentes con los retos enmarcados dentro del enfoque. Con base a ello, los apartados siguientes son un punto de reflexión acerca de cómo garantizar la formación básica en competencias si el sistema busca homogeneidad en un aula inevitablemente variado y más aún cuando las mismas herramientas de medición –no de valoración íntegra– favorecen la operatividad en términos mercantilistas antes que la inteligencia emocional y por ejemplo la dimensión axiológica con la que egresan de las escuelas los estudiantes.

4.2.2.4 Evaluaciones estandarizadas como sistema de medición de competencias

El campo docente no solo conlleva ejercicio y constante formación para actualizar prácticas conforme cambia y evoluciona el mundo, se debe, además, tener claridad sobre la importancia del “abordaje de contenidos programáticos de las asignaturas, lo cual privilegia la estandarización de los procesos de aprendizaje” (Hernández – Mosqueda, 2016, pp. 359) y del trabajo por competencias. Así, cabe recalcar de acuerdo con Chica (2011), que éstas contribuyen al desarrollo personal y social, siendo necesarias en los estudiantes para hacer frente a los desafíos que se imponen por diferentes y complejas situaciones de la vida.

De manera consecuente a lo anterior, desde la evaluación como proceso continuo para reflexionar sobre desempeños de los estudiantes en el propósito de desarrollar competencias y sobre las prácticas pedagógicas de los docentes, para tal fin, se señala que, un aspecto más de importancia –pero también de dificultad–, que estriba en el enfoque competencial, es la calificación y categorización de instituciones y países de

acuerdo a resultados en pruebas estandarizadas que se establecen como sistema de evaluación de desempeños en competencias básicas para la medición cuantitativa de lo que podría entenderse como calidad educativa.

Pues bien, en ese afán de mejorar la educación y promover la productividad de las naciones, Colombia, se une a la Organización para la Cooperación y el Desarrollo Económico (OCDE) y dentro de la misma, somete una muestra de sus estudiantes de 15 años seleccionados de manera sistemática, a la denominada prueba PISA (Programa de Evaluación Internacional de los Alumnos) cuyos resultados en sentido general, compara, mediante indicadores de aprendizaje, la calidad y progreso académico/educativo de los países miembros del organismo, en torno a competencias básicas de matemáticas, lenguaje y ciencias naturales; desconociéndose, dentro de este concepto de calidad, las dimensiones formativas de la evaluación que sí toman en consideración singularidades y valores humanos propios de la etapa escolar y de la interacción social, escolar y familiar de cada uno de los estudiantes (Patiño, 2017).

Por otro lado, como un elemento más de crítica y reflexión, que está dentro de los fines generales del presente trabajo -en tanto se busca valorar la singularidad y particularidad inevitablemente diferenciada de las instituciones y estudiantes-, se debe mencionar que, de acuerdo con Castaño (2015), los informes PISA, difícilmente trascienden de su forma instrumental para desde los estados, controlar y reglamentar aspectos al interior de las instituciones promovidos especialmente por políticas externas que mitigan las realidades de las escuelas en torno a por ejemplo, características económicas, democráticas, culturales, históricas, procedimentales y demás. Se favorece entonces, la burocracia educativa como una perspectiva equivocada de calidad en la que, se penalizan a las instituciones por no mostrar los resultados esperados en sus estudiantes.

A nivel nacional, en sentido medianamente diferenciado a las pruebas PISA, el Ministerio de Educación, desde el ICFES (Instituto Colombiano para el Fomento y evaluación de la Educación Superior), institucionaliza la presentación al examen saber 11°, para valorar no solo las aptitudes de los estudiantes que terminan su educación media, también para establecer un ranking de Instituciones respecto a los desempeños en competencias

básicas desde cinco áreas fundamentales del conocimiento: Matemáticas, Ciencias Naturales, Ciencias Sociales, Lectura crítica e Inglés.

Con base a lo anterior y lo ya mencionado referente a PISA, el examen saber 11 tiene gran utilidad instrumental para el estado y la posibilidad de ingreso de los estudiantes a las instituciones de educación superior, también se convierte en un parámetro para que las instituciones de educación primaria, secundaria y media, reflexionen sobre los niveles competenciales promovidos en cada una de las promociones evaluadas, desde los resultados diferenciados para cada uno de los 5 campos del conocimiento ya antes mencionados. No obstante, la valoración integral sigue siendo un aspecto de trascendencia que se desconoce. Retomando a Fernández et. al. (2017) a continuación se referencian de forma sintetizada puntos que se tornan desfavorables en torno al uso actual de este tipo de pruebas y que enmarcan la necesidad que fundamenta este trabajo: La valoración de la singularidad biológica y relacional que constituyen a los inextricablemente distintos estudiantes que llegan a las instituciones y por tanto a las aulas.

- Los análisis que emergen de los resultados de las pruebas son lineales y descriptivos, ocultan aspectos significativos que podrían generar una mejor intervención para mejora de las instituciones y sobre todo de la formación de los estudiantes.
- El ruido en la información, generado por las publicaciones de los resultados o los mismos informes ejecutivos, deja de lado aspectos procedimentales en los que tiene sentido el intercambio educativo –prácticas pedagógicas–, directamente relacionadas con la observación del quehacer educativo.
- Confusión del concepto de evaluación como sistema metacognitivo, para seguir siendo entendido como “regla” de medición.
- Tratamiento de los resultados de las pruebas con fines políticos antes que educativos. Es decir, se disponen o distribuyen recursos de acuerdo a los resultados que tales pruebas emiten.
- Inhibición o desmonte de la función social de las instituciones. Puesto que, se remiten sus prioridades al desarrollo de un currículo que compita con otros

establecimientos en la mal llamada mejora de la calidad educativa. Esto conlleva, además, de manera subyacente, una estigmatización de la educación pública y un amplio favorecimiento al sector privado. Se relega implícitamente, la población menos favorecida -económicamente hablando- porque son los lugares en los que se presentan los desempeños más bajos.

- Fomento de las prácticas educativas por y para los buenos desempeños en este tipo de pruebas. Es decir, se ejerce presión sobre las instituciones y por tanto sobre la labor docente para que los fines del aula traigan como producto la superación de los niveles mínimos o básicos en las pruebas.
- Centralización en las disciplinas evaluadas. Privilegiando, por ejemplo, para el caso de Colombia, áreas como matemáticas, lenguaje, ciencias naturales, sociales e inglés y discriminando en cierta o gran medida –en algunos casos– campos del conocimiento y la formación como la ética, el arte y los deportes.

Por último, los anteriores apartados no solo justifican la importancia de las competencias básicas en los sistemas educativos, también se pone en un espacio reflexivo, la necesidad de valorar todas las dimensiones formativas de los estudiantes. Así, dentro de la investigación, se busca dar uso a las singularidades sociodemográficas de los estudiantes, para prever una sugerencia de la priorización en las practicas pedagógicas que permitan promover el desarrollo de competencias en los estudiantes que posiblemente más lo necesiten. Esto propende una prospectiva de educación diferenciada, pero que, con base a evidencia, podría sentar un referente de reflexión acerca del enfoque de homogeneidad que se promueve actualmente.

4.2.3 Aspectos socio-demográficos y de desempeño que caracterizan a los estudiantes

Dentro de las instituciones educativas se promueve la uniformidad, la homogeneidad y busca propiciarse la invariancia en términos de rendimiento académico aun sabiendo que cada persona no puede prescindir de su singularidad, no solo biológica sino también contextual y personal. En este marco se señalan a continuación tres apartados que dejan en evidencia las bondades de la singularidad de los estudiantes para establecer

relaciones que podrían, en el caso particular de esta investigación, generar predicciones en los desarrollos de competencias básicas.

4.2.3.1 Entorno social, familiar, educativo y su influencia en los estudiantes

Trabajos variados que han enriquecido la teoría para mejorar prácticas educativas en el campo de la psicopedagogía, como el realizado por Piaget (1981), manifiestan la manera en la que pueden aprender las personas según su desarrollo biológico. En ese sentido se sugiere que, de acuerdo a la edad de desarrollo físico, los estudiantes podrán realizar ciertas operaciones en términos cognitivos. Por su parte, en sus trabajos Vygotsky –sin desconocer los aportes de Piaget– considera que más allá del desarrollo biológico ineludible de los individuos, es el ambiente o entorno un factor incontrovertiblemente significativo para que las personas aprendan. Así, señala que, parte de las nociones aprendidas por los estudiantes, están íntimamente relacionadas con la manera y con quiénes se ha desarrollado tanto en sociedad, familia y escuela (Álvarez y Del Río, 1990). Se constituye de esa manera un marco de desarrollo formativo a partir de las interacciones, de las relaciones y en general del contacto con el ambiente.

Trabajos como el de Moral y Ovejero (2013); Fernández (2013); Zambrano-Villaba y Almeida-Monge (2017), Uribe y Márquez (2022), definen relaciones significativas entre entornos sociales, familiares, educativos y variables como desempeños académicos y distintas conductas de los estudiantes asociados a por ejemplo acoso escolar, riesgo en adolescentes e incluso tendencia a la violencia. Se referencia, en este sentido, no solo la influencia del entorno social y familiar en los procesos de aprendizaje de los estudiantes, además, se dejan antecedentes que recuerdan la pluralidad de las singularidades de los estudiantes que acuden a las instituciones educativas y las posibilidades de prever ciertos comportamientos.

Conforme a los anteriores referentes, se recalca que, es la familia el primer sistema social en el que se desenvuelven las personas, es básicamente el núcleo de formación, pero no todas las familias se conforman ni de la misma manera ni tienen un número de integrantes similar, surgiendo así multiplicidad de situaciones que vale la pena valorar como características de los estudiantes. En segundo lugar, la escuela se convierte en el espacio social de formación con pares, donde se nutren relaciones de amistad y

competencia, se destacan aspectos de tipo axiológico y de reconocimiento. Adicionalmente y en sentido consecuente, es importante considerar que las diferentes agrupaciones sociales tienen su propia estructura y su subcultura propiciada por las interacciones en pequeñas sub-agrupaciones dentro de espacios sociales más generales. Se habla así de costumbres y economías diferentes al interior de un mismo departamento e incluso dentro de un mismo municipio (Uribe y Márquez, 2022). Esto particulariza cada estudiante y da espacio a la reflexión de si tales particularidades permiten hacer análisis más allá de la caracterización o descripción de lo ya mencionado. Es decir, ¿podrían –por ejemplo– tener alguna relación subyacente en el nivel económico, el tipo de familia o la lejanía a la escuela, los distintos subgrupos de estudiantes que conforman una institución educativa? Y de ser así, ¿no se debería, en términos educativos, usar tal información para mejorar prácticas educativas?

Conforme a lo anterior, se destaca en el siguiente apartado, los aspectos socio demográficos que, con base a la evidencia, se deben tomar en cuenta para llegar a un posible análisis predictivo en estudiantes de un grado o rango de edad determinada.

4.2.3.2 Aspectos específicos del entorno que podrían influir en los desempeños

De acuerdo con el apartado anterior y los trabajos de Monge-López et al. (2017); Castra-Sánchez et al. (2019); Álvarez et al. (2007); Carvajal et al. (2018); Brito-Jimenez y Palacio-Sañudo (2016); De Tejada (2012); Álvarez-Ramírez y Cáceres-Hernández (2010); Loli et al. (2010); Moral y Ovejero (2013); Fernández (2013); Zambrano-Villaba y Almeida-Monge (2017), Uribe y Márquez (2022), a continuación, en la tabla 7, se sintetizan los aspectos más importantes a considerar, que podrían influir en los desempeños o conductas de los estudiantes en edad adolescente.

Tabla 7. Aspectos de acuerdo al entorno, que pueden influir en el aprendizaje y conductas de los estudiantes

Entorno	Aspecto a considerar
Social	- Área poblacional en la que reside (Rural o Urbana)
	- Recursos con los que se dispone en el barrio o vereda (Servicios)
	- Seguridad de la zona de residencia
	- Lejanía a la institución (tiempo de recorrido)
	- Modo de desplazamiento a la institución
	- Clase social (estrato)

	- Grupo social especial (afrodescendiente, desplazado, víctima de conflicto, indígena)
	- Actividades complementarias ofrecidas por los barrios o veredas
	- Riesgo a drogas o alcohol en la zona donde habita
Familiar	- Tipo de familia según Sánchez – Lorite (2012): Elemental (padre, madre e hijos); Consanguínea (abarca padres, abuelos, tíos, primos...); Monoparental (solo el padre o solo la madre); Padres separados (situaciones de divorcio); Homoparental (si se trata de padres homosexuales o madres lesbianas)
	- Número de integrantes de la familia (personas con quienes vive)
	- Estilo educativo de las familias de acuerdo con Baumrid (1991) y Torio et al. (2008): Autoritarios; Permisivos; democráticos e indiferentes.
	- Ocupación de los padres y nivel de ingresos
	- Escolaridad de los padres
	- Relación con los padres
	- Número de hermanos y lugar entre ellos
	- Tratamiento de los conflictos en familia (diálogo, discusiones, violencia...)
	- Espacios para la comunicación asertiva
Escolar	- Infraestructura y aulas adecuadas para el aprendizaje
	- Uso de TIC en las prácticas pedagógicas
	- Relación con sus compañeros y docentes
	- Materia preferida
	- Hábitos y tiempo de lectura
	- Apoyo institucional en su proceso formativo
	- Reprobación en años previos
	- Desempeños académicos
	- Edad
	- Psico-orientación
	- Jornada Escolar

Fuente: Elaboración propia

Con base a lo anterior, en la presente investigación se toma en consideración, además de los aspectos ya mencionados, el género biológico del estudiante y la afinidad para con la escuela, puesto que, en concordancia con Finn (1989) y Ros (2009), la relación que surge entre los estudiantes y las instituciones, conlleva un alto grado de compromiso, o por el contrario, un amplio sentimiento de rechazo que se verá reflejado en bajos desempeños, convirtiéndose éste último aspecto en un posible predictor de su avance en competencias básicas. A continuación, se muestran los aspectos a considerar y las diferentes convenciones que se darían para efectos de análisis en la presente investigación.

Tabla 8. Variables socio demográficas consideradas en la investigación

Entorno	Variable y categorías	Convención
Social	- Lugar de residencia: a) Rural b) Urbano	TP_RES: Ru, Ur
	- En su lugar de residencia cuenta con: Acueducto, alcantarillado, electricidad, gas domiciliario, computador, Internet, televisión, bicicleta, moto, automóvil.	SER: Ac, Al, El, Ga, Pc, In, Bi, Mo, Car
	- Seguridad de la zona de residencia: a) Alta b) Media c) Baja	SEG_ZO: Al, Me, Ba
	- Lejanía a la institución (tiempo de recorrido): a) menos de 15 min b) 15 a 30 min c) 30 a 60 min d) más de 1h	LEJA: <15, 15-30, 30-60, >60
	- Modo de desplazamiento a la institución: a) Transporte escolar b) Carro c) Moto d) Caminando	DESP: tr_es, car, mot, cam
	- Clase social (estrato): 0, 1, 2, 3	ESTR: 0, 1, 2, 3
	- Grupo social especial: a) afrodescendiente b) desplazado c) víctima de conflicto d) indígena	PO_ESP: afr, des, vic, ind
	- Actividades complementarias ofrecidas por los barrios o veredas y a las que acude: Danzas, Pintura, Teatro, Deportes	ACT_COMP: dan, pin, tea, dep
	- Riesgo a drogas o alcohol en la zona donde habita: a) Alta b) Media c) Baja	RIES_DR_ALC: Al, Me, Ba
Familiar	- Tipo de familia: Elemental (vive con padre, madre y hermanos), Consanguínea (vive con padres, abuelos, tíos, primos...), Monoparental (vive solo con el padre o solo la madre), Padres separados (vive con alguno de los padres por situaciones de divorcio), Homoparental (si se trata de padres homosexuales o madres lesbianas)	TP_FAM: ele, con, mon, p_sep, hom
	- Número de integrantes de la familia (personas con quienes vive): 2, 3, 4, 5, más de 5	NUM_INT: 2, 3, 4, 5, >5
	- Sus padres lo han educado de manera: a) Autoritaria, b) Permisiva, c) Democrática, e) indiferente	TP_EDU: aut, per, dem, ind
	- Ocupación de los padres (pregunta abierta) y nivel de ingresos: 1 salario mínimo, 1 a 2 salarios mínimos, más de dos salarios mínimos	OCU: ¡ABIERTA! INGR: 1sm, 1-2sm, >2sm
	- Escolaridad de los padres (madre y padre por separado): a) primaria b) básica secundaria c) bachillerato d) Técnico o tecnólogo e) Universitario	ESC_P; ESC_M: pri, sec, bac, tec, uni
	- Relación con los padres (madre y padre por separado): a) Excelente b) Buena c) Regular d) Mala	REL_P; REL_M: exc, bue, reg, mal
	- Número de hermanos: 1, 2, 3, 4, 5, más de 5	N_HER: 1, 2, 3, 4, 5, >5
	- Lugar entre ellos: 1, 2, 3, 4...	LUG_HER: 1, 2, 3, 4,...
	- Tratamiento de los conflictos en familia: diálogo, discusiones, violencia, otra, cuál.	CONFL: dia, dis, vio
	- Espacios para compartir en familia: a) Sí, b) No, c) Algunas veces	ESP_FAM: S, N, AV
Escolar / personal	- Considera que la Institución cuenta con buena infraestructura y aulas adecuadas para el aprendizaje: a) Excelente b) Buena c) Regular d) Mala	INFR: exc, bue, reg, mal
	- Los docentes hacen uso de TIC en las prácticas pedagógicas: a) Sí, b) No, c) Algunas veces	TIC: S, N, AV
	- Relación con sus compañeros y docentes (por separado): a) Excelente b) Buena c) Regular d) Mala	REL_D; REL_C: exc, bue, reg, mal
	- Materia preferida (pregunta abierta)	MAT_P: ¡ABIERTA!
	- Usualmente, cuanto tiempo al día dedica a leer por entretenimiento: a) No lee por entretenimiento b) 30 minutos o menos c) entre 30 y 60 minutos d) entre 1 y 2 horas e) más de 2 horas	LEE: N, <30, 30-60, 60-120, >120



Entorno	Variable y categorías	Convención
	- Se siente a gusto y considera que la institución brinda apoyo en su proceso formativo: a) muy a gusto b) a gusto c) regular d) muy poco	APOY: m_agu, agu, reg, m_poc
	- Reprobación en años previos: 1 vez, 2 veces, 3 veces, más de 3 veces	REPR: 1, 2, 3, >3
	- Desempeños académicos (tomados de plataforma educativa)	--
	- Edad en años: 14, 15, 16, 17, 18 o más	ED: 14, 15, 16, 17, >=18
	- Considera que la jornada Escolar es: a) Muy adecuada b) adecuada c) inadecuada d) Completamente inadecuada	JOR_INST: m_ade, ade, ina, c_ina
	- Género biológico: a) Masculino b) Femenino	GEN: M, F

Fuente: Elaboración propia

En el capítulo 6 de la investigación se retoman las variables y la manera en que se obtienen para su análisis posterior.

5. OBJETIVOS DE LA INVESTIGACIÓN

Como en toda investigación, los propósitos de la misma, deben orientar el qué, el cómo y el para qué del trabajo a realizar, de ese modo, a continuación, se establecen los objetivos que dan sentido y significado a la investigación.

5.1 Objetivo general

Proponer un modelo predictivo de desempeños en competencias básicas, usando herramientas de *machine learning* que interrelacionen variables de entorno (aspectos socio demográficos) y de desempeños (competencias básicas) en estudiantes de décimo grado, para justificar el uso de estrategias diferentes en las prácticas pedagógicas de las instituciones Ramón Alvarado Y La Merced.

5.2 Objetivos específicos

- Caracterizar desde aspectos socio demográficos los estudiantes de grado décimo de las instituciones Ramón Alvarado y La Merced.
- Interrelacionar variables sociodemográficas y de desempeño académico, implementando una herramienta machine learning que pueda llegar a predecir dificultades en el desarrollo de las competencias básicas de estudiantes de grado décimo a través de un entorno virtual.
- Evaluar la incertidumbre de la información en los algoritmos para considerar la fiabilidad de la herramienta machine learning en la predicción de dificultades en el desarrollo de competencias básicas de los estudiantes de grado décimo de las instituciones educativas Ramón Alvarado y La Merced.
- Proponer una estrategia que permita reorientar prácticas pedagógicas que puedan ayudar a mejorar falencias en el desarrollo de competencias básicas.

6. METODOLOGÍA

Dentro del anhelo de dar solución a las preguntas que sistematizan el problema (Ver capítulo 2), es imprescindible considerar la manera en la que se van a solucionar éstas y la posible ruta que puede orientar el trabajo investigativo en pro de alcanzar los objetivos. Así, en los siguientes apartados, se explicita tal sendero de acuerdo a la singularidad del problema a abordar que se relaciona con aspectos predictivos vinculando medios tecnológicos como el machine learning para interrelacionar variables socio demográficas con desempeños competenciales de estudiantes en dos focos geográficos distintivos: Agrado y Zona rural de Garzón, Huila.

6.1 Tipo y enfoque de la investigación

De acuerdo con Hernández-Sampieri (2018), en su aporte sobre metodología de la investigación, se considera el presente trabajo está enmarcado dentro de la modalidad mixta, debido al abordaje de variables cualitativas relacionadas con la singularidad socio demográfica de los estudiantes y cuantitativas referidas a desempeños en competencias básicas e índices de fiabilidad en las predicciones aplicadas por los algoritmos ML, relacionadas con tales competencias, de los estudiantes de las I. E. La Merced y Ramón Alvarado. Seguidamente, por su diseño, dentro de la modalidad cuantitativa, es de tipo no experimental-longitudinal, puesto que, las variables y datos recopilados no serán manipulados más allá de la selección de categorías pertinentes para la caracterización de los estudiantes y las posibilidades de evidenciar relaciones. Adicional, como se harán estudios de prueba de modelos ML, entonces habrá varios momentos de intervención operativa con el trabajo de selección de los -o el- algoritmos más efectivos. En ese sentido, se puede afirmar que, es de carácter longitudinal.

Se reconoce, además, que los aspectos cuantitativos permean una base de datos en la que buscan interrelacionarse los aspectos socio demográficos y de desempeños en competencias básicas. Así, tales relaciones deberían generar fundamentación teórica concerniente a la posibilidad o no de predictibilidad desde el modelo propuesto, contenido en el enfoque cualitativo.

Por otro lado, conforme al alcance, se promueve una investigación correlacional entre los aspectos académicos (desempeño en competencias básicas) y sociales

(características sociodemográficas) que por medio del uso de algoritmos machine learning permita evaluar si las interrelaciones predicen dificultades en competencias básicas.

6.2 Universo de estudio, población y muestra

El trabajo investigativo se plantea para ser implementado en una población estudiantil rural y urbana en los municipios de Garzón y el Agrado, Huila. En ese sentido, la población de estudio serán los estudiantes de grado décimo de ambas instituciones. Se toma el grado décimo porque como se estipula en el marco teórico, son estudiantes con edades próximas o en las que se aplican pruebas estandarizadas que se espera den muestra del desarrollo competencial inherente a su formación.

La tabla 9, muestra aspectos generales de la población tomada para la investigación y describe singularidades como el género y el origen residencial de los estudiantes.

Tabla 9. Características generales de la población

Género	Aspectos Generales							Total
	Zona Residencial		Edad			Reprobación previa		
	Rural	Urbana	14	15	> = 16	Sí	No	
Femenino	25	27	2	28	22	8	44	52 ⁵
Masculino	27	13	0	23	17	12	28	40 ⁶
Total	52	40	2	51	39	41	51	92

Fuente: Elaboración propia

En las figuras 23, 24 y 25, se ilustra de manera gráfica la información general de la población objeto de estudio.

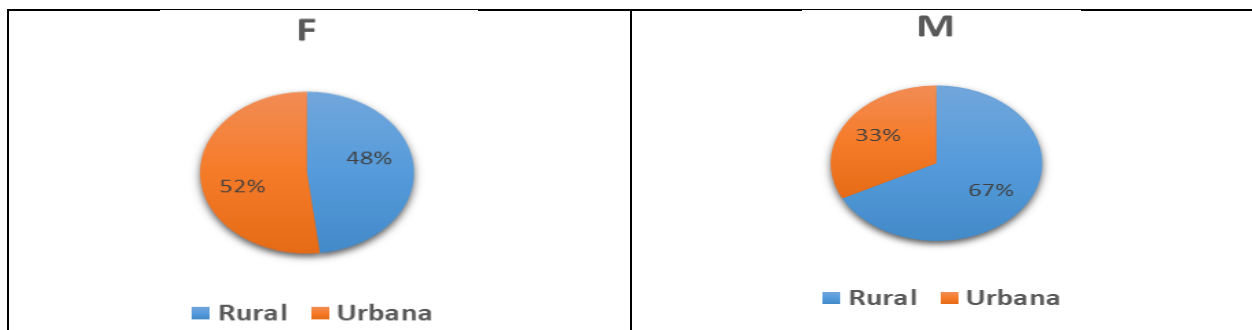


Figura 23. Zona Residencial de acuerdo al género Biológico

Fuente: Elaboración propia

⁵ Suma de mujeres, no de los subtotales por categoría

⁶ Suma de hombres, no de los subtotales por categoría

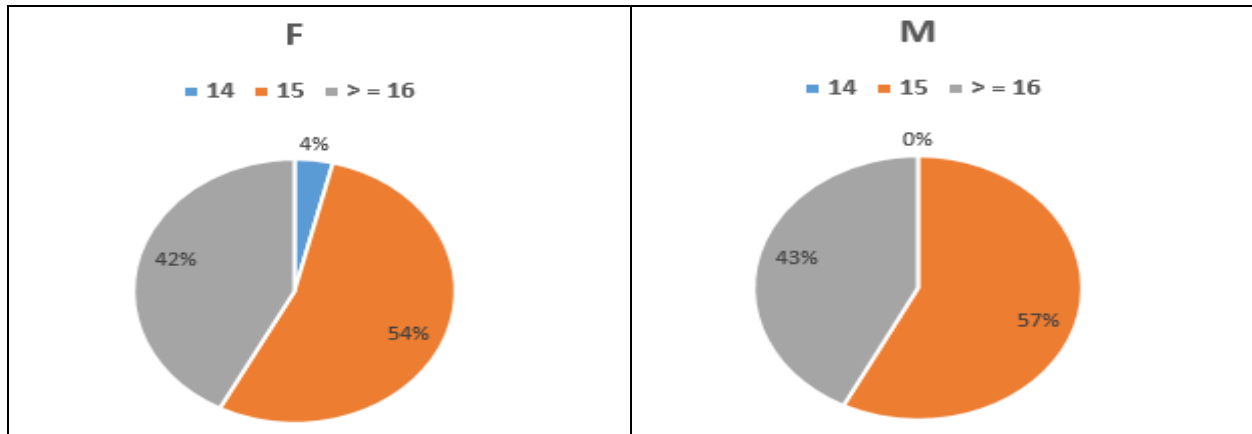


Figura 24. Edad de acuerdo al género biológico
Fuente: Elaboración propia

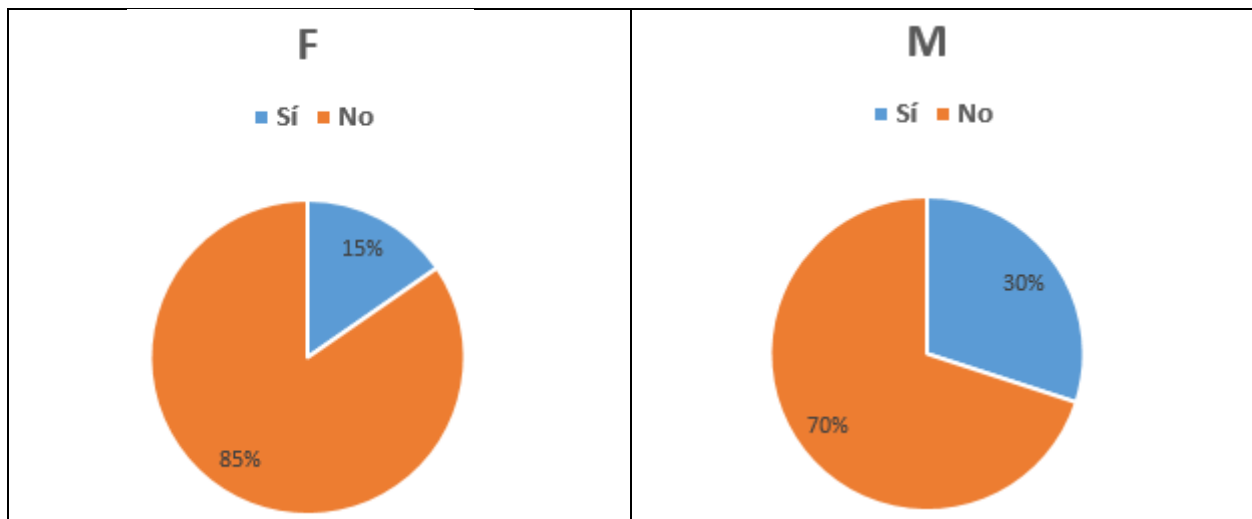


Figura 25. Reprobación de acuerdo al género biológico
Fuente: Elaboración propia

Con base a lo anterior, se toma una población de 92 individuos en la que es interesante ver que, aunque hay más mujeres, el % de hombres que han perdido años anteriores, es significativamente mayor. De igual manera, se podría hacer una relación de tal característica, con por ejemplo la edad, puesto que, mientras la mayoría de las mujeres tienen 14 o 15 años, los hombres en más del 40% tienen edades superiores a los 15. Ahora, en cuanto a la zona residencial, es notable que la mayor parte de los estudiantes del género femenino, se ubican en zona urbana y no hay interrelación aparente de esta característica con los demás aspectos mostrados. Sin embargo, es difícil que la tabla pueda dar pie para inferir alguna relación trascendente más allá de conteos. De ahí la

necesidad de estudiar la complejidad intrínseca con métodos no lineales o convencionales.

De manera consecuente, se trabajará un algoritmo de predicción, se usará un grupo de prueba para verificar los algoritmos dentro del campo del machine learning. Para ello se hará una selección aleatoria (tomando datos del ámbito rural y urbano) del 20% del total de la población y, en dichos datos, se empezará un proceso de tratamiento de los mismos para buscar la manera de reducir en mayor medida la incertidumbre de la información.

6.3 Estrategias Metodológicas

Es importante mencionar que hay una diferencia entre método y metodología, así, de acuerdo con Chagoya (2008), es pertinente considerar que, desde la etimología de ambas palabras, una se define con base a la otra. Por consiguiente, el método se entiende como una orientación, camino, sendero o ruta adecuada para llegar a un punto determinado. Mientras que la metodología se convierte en la descripción, análisis, valoración crítica, reflexión y/o discernimiento acerca de los métodos investigativos a usar. En ese sentido, desde una perspectiva compleja, las estrategias metodológicas planteadas se basan en la correlación de métodos para concurrir en el alcance de los objetivos propuestos.

Conforme el enfoque metodológico de la investigación, y lo mencionado en el anterior párrafo, se establecen algunas fases de acción para la consolidación del trabajo y, por ende, del alcance de los objetivos planteados. A continuación, en concordancia con las características propias del trabajo y las fases expuestas por Pérez y Gonzalez (2007) en el marco teórico, relacionadas con el uso de ML, en la tabla 10, se establecen las fases que permitirán desarrollar el trabajo de Tesis.

Tabla 10. Fases para el desarrollo del trabajo investigativo

Objetivo	Fase	Descripción
Caracterizar desde aspectos socio demográficos los estudiantes de grado décimo de las instituciones Ramón Alvarado y La Merced	1. Caracterización	De manera inicial, se caracterizará a población objeto de estudio, para ello se usará un cuestionario Google (Ver anexo 1) que servirá como medio de recolección de datos, luego se hará un estudio de las singularidades sociodemográficas de cada estudiante.
Interrelacionar variables sociodemográficas y de	2. Comprensión de los datos	Previo a la recolección de datos que ya se han estipulado en la caracterización socio



desempeño académico, implementando una herramienta machine learning que pueda llegar a predecir dificultades en el desarrollo de las competencias básicas de estudiantes de grado décimo a través de un entorno virtual		demográfica, es necesario tener claras las categorías de las variables y convenciones que pueden orientar la solución al problema.
	3. Limpieza de los datos	Para esta fase, como ya se han identificado las variables a estudiar en la población y la naturaleza de las mismas, debe eliminarse el ruido de la información. Es decir, depurar datos incompletos, consolidar una sola base de datos –en el caso que hayan varias fuentes de recolección-, establecer el formato a trabajar para que el algoritmo que se vaya a probar los lea sin problema, entre otras cosas que pueden emerger.
	4. Prueba del modelo ML	Posterior a la limpieza de los datos, se aplican los algoritmos, herramientas o modelos que permitan clasificar estudiantes en los diferentes niveles de desempeño en competencias básicas, de acuerdo a las entradas dadas que serán las características socio demográficas. Adicional, el modelo seleccionado, deberá poder vincularse a un aplicativo Web que se convierte en el entorno virtual de ayuda pedagógica a los docentes para en prospectiva, mejorar prácticas pedagógicas.
	5. Diseño del Entorno Virtual	Seguidamente se hará uso de un entorno virtual para vincular luego el modelo <i>machine learning</i> seleccionado y que se pueda autoalimentar para seguir generando respuestas automáticas con base a las relaciones encontradas en la prueba inicial.
Evaluar la incertidumbre de la información manejada en el entorno virtual para considerar la fiabilidad de la herramienta <i>machine learning</i> en la predicción de dificultades en competencias básicas de los estudiantes de grado décimo de las instituciones educativas Ramón Alvarado y La Merced	6. Evaluación	Dentro del uso del campo Big data y herramientas <i>machine learning</i> para la predicción de dificultades académicas, se debe considerar aspectos de exactitud, precisión y en general de fiabilidad del algoritmo -o los- que se vayan a utilizar. En ese sentido, se ha de valorar los niveles de incertidumbre en la información comparando los valores de las métricas accuracy, precision, recall y F-1 score ya expuestas en el marco teórico para así mismo definir lo apropiado de uno u otro algoritmo a usar. En esencia, es necesario tomar en consideración la entropía de la información para así tener un fundamento de pertinencia.
	7. Integración del modelo	Con base a las métricas de incertidumbre consideradas en los modelos en la fase anterior, luego de seleccionarse el apropiado, se debe integrar este al entorno virtual para que el modelo pueda comunicarse con otras partes del sistema y seguir generando respuestas automáticas que deben monitorearse para verificar que no se genere un aumento progresivo en los errores emergentes.
Presentar una estrategia que permita reorientar practicas pedagógicas que puedan	8. Estrategia	Por último, de acuerdo con los hallazgos encontrados, se presenta una estrategia que

ayudar a mejorar falencias en el desarrollo de competencias básicas	busca reorientar practicas pedagógicas, en pro del desarrollo de competencias básicas.
---	--

Fuente: Elaboración propia

Conforme se plasman las fases anteriores, se hace necesario planificar las acciones que confluyan en la realización efectiva de las mismas. A continuación, en la tabla 11, se relacionan las acciones correspondientes a cada fase/objetivo, dentro de los propósitos de la investigación.

Tabla 11. Acciones relacionadas con cada fase/objetivo

Fase/Objetivo	Acciones	Subproducto
1. Caracterización/Obj-1	- Caracterización de la población y muestreo para el estudio. Establecimiento de la relevancia de las variables sociodemográficas dentro de la singularidad de los estudiantes.	Análisis teórico sobre características socio demográficas.
2. Comprensión de los datos/Obj-2	- Identificación con fundamentos teóricos, de la importancia de la formación por competencias haciendo uso de investigaciones y reflexionando sobre el sistema educativo colombiano. - Comparación y diferenciación –con fundamentación teórica- entre competencias genéricas, básicas y específicas. - Identificación de las dificultades en el desarrollo de competencias básicas en los estudiantes de grado décimo de las instituciones Ramón Alvarado (Garzón, Rural) y La Merced (Agrado, Urbana). Mediante las calificaciones generadas periódicamente - Establecer de manera coherente, las variables sociodemográficas y de desempeños que se van a utilizar, tomando en consideración la naturaleza de las mismas y las respectivas categorías. - Estipular los datos que estarían en la población y el porcentaje considerado para muestras con los que se va a trabajar en los algoritmos.	Artículo: Sobre herramientas machine learning y su aplicación en la educación. Artículo: Sobre las relaciones entre variables socio-demográficas y de desempeño en estudiantes de grado 10°. Aplicativo Web: Análisis predictivo de desempeños en competencias básicas
3. Limpieza de los datos/Obj-2	- Establecer los datos a recopilar y la manera de hacerlo, el tiempo de recopilación y el formato en que se manejarán éstos. Por ejemplo, se puede usar un Google forms y tratar los datos con la extensión .xlsx. (un solo formato). - Recopilados los datos, depurar los que estén incompletos o que llegue a establecerse, no son acordes a los propósitos investigativos.	
4. Prueba del modelo ML/Obj-2	- Buscar herramientas machine learning que existan en el mercado y que puedan permitir analizar datos para hacer informes descriptivos y predictivos. - Realizar un filtro en la búsqueda anterior con indagación bibliográfica que dé fundamento a la prueba de algunos algoritmos de clasificación que puedan servir en la investigación de acuerdo a la naturaleza de los datos obtenidos (Ver tabla 2).	

	<ul style="list-style-type: none"> - Probar algoritmos ML de clasificación cuyas entradas a relacionar serán las características socio demográficas de los estudiantes y la salida sean los niveles de desempeños en competencias básicas. - Considerar el –o los– modelos que adicional a la muestra de interrelaciones entre las variables, se pueda integrar a un entorno virtual. 	
5. Diseño del Entorno Virtual/Obj-2	<ul style="list-style-type: none"> - Investigar sobre los diferentes códigos, lenguajes y métodos que utilizan las herramientas para hacer análisis predictivo. - Codificar un aplicativo Web usando alguno de los lenguajes de programación, para que funcione como entorno virtual educativo y que además pueda vincularse a algoritmo de predicción de desempeños académicos. 	
6. Evaluación/Obj-3	<ul style="list-style-type: none"> - Analizar las métricas de incertidumbre que arrojen las herramientas machine learning y valorar la fiabilidad en el propósito de predecir niveles de desempeños en competencias básicas de los estudiantes. - Evaluar la aceptación -en la comunidad educativa de las dos instituciones- del interfaz propuesto dentro del aplicativo Web a través de una encuesta de satisfacción sobre la utilidad del sistema. 	Artículo: Sobre la fiabilidad de herramientas predictivas en entornos virtuales-educativos.
7. Integración del modelo/Obj-3	<ul style="list-style-type: none"> - Realizar una prueba del entorno virtual tanto a nivel descriptivo como predictivo en la que pueda valorarse la integración efectiva en términos operativos del algoritmo ML. 	
8. Estrategia/Obj-4	<ul style="list-style-type: none"> - Construir una estrategia pedagógica que permita usar los hallazgos encontrados para posibilitar la mejora en el desarrollo de competencias básicas. 	

Fuente: Elaboración propia

Adicionalmente, para ilustrar la sinergia metodológica establecida en la anterior tabla y que sea más entendible, en la figura 23, se muestra el flujo metódico usado en la investigación.

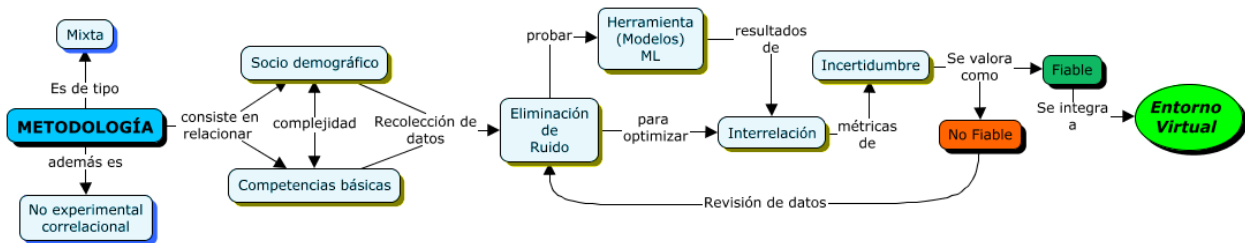


Figura 26. Relación metodológica

Fuente: Elaboración propia

6.4 Técnicas e instrumentos de Investigación

Se puede encontrar en la red, múltiples trabajos investigativos con un número igualmente variado de técnicas e incluso más, de instrumentos para realizar un trabajo investigativo. En ese sentido, es importante resaltar en este apartado que, dentro del trabajo científico, los métodos, técnicas e instrumentos usados, son tan importantes como la

fundamentación teórica sobre la que estriba el tema de investigación. En concordancia con Maya (2014), se considera importante posibilitar la obtención de conocimiento sistemático que, de forma coherente oriente el proceso que se realiza. Conforme a lo anterior, en la tabla 12, se presentan las técnicas e instrumentos usados para el trabajo investigativo relacionado con la caracterización de estudiantes desde sus singularidades sociodemográficas y sus desempeños académicos. Desempeños que se busca predecir desde la aplicación de herramientas *machine learning* y que se describen también.

Tabla 12. Técnicas e instrumentos usados en la investigación

Técnica(TC) /Instrumento(IN)	Fase	Descripción
TC: Encuesta	1. Caracterización	Uso de un formulario Google para conocer las variables sociodemográficas solicitadas en el Proyecto Educativo de Orientación Escolar (POEG) que caracterizan a los estudiantes.
TC: Observación directa	2. Comprensión de los datos	Uso de los desempeños académicos de los estudiantes de grado décimo, obtenidos a base de valoraciones que se basan en la observación directa del desarrollo de sus competencias.
IN: Google Forms	3. Limpieza de los datos	Los cuestionarios de google tienen amplias ventajas para recopilar información. Entre estas, podemos hablar de la amabilidad de la interfaz gráfica, de las múltiples y sencillas herramientas para crear encuestas y además que solo requiere uso de internet para su uso porque es gratuito y no conlleva un alto nivel de proceso, de ahí que se pueda usar incluso desde un celular.
IN: Microsoft Excel		Usaremos Microsoft Excel como hoja de cálculo para recopilar, tratar y limpiar los datos que darán muestra de la información particular de cada estudiante en términos sociodemográficos y de desempeño. Es decir, el Google Forms se usará como el medio para obtener una base de datos sociodemográficos con base a las variables que se explicitan en el cuestionario.
IN: Python	4. Prueba del modelo ML	Uso del lenguaje de programación Python, para el uso de herramientas machine learning que se probarán y serán validadas posteriormente, de acuerdo a las métricas de incertidumbre. Dentro de los algoritmos a considerar están los siguientes clasificadores: Regresión logística (Logistic Regression:LR) Árbol de decisión (Decision Tree:DT); K vecinos más cercanos (K nearest neighbors: KNN); Red neuronal (Neural net: NN) y máquinas de vectores soporte (Support Vector Machine: SVM)
IN: Anaconda		Uso de la plataforma libre anaconda, en la que, para efectos de efectividad en la praxis, no solo se cuenta con interfaces para el trabajo con Python, también está R y, en general, incluye todo lo necesario en ciencia de datos para el tratamiento de información.
IN: Base de datos Mysql	5. Diseño del Entorno Virtual	Para almacenar la información recolectada usaremos el gestor de base de datos Mysql, se utilizará de tal forma

		que nos permita relacionar todos los datos obtenidos de acuerdo a las necesidades investigativas.
IN: Lenguaje de programación PHP		La aplicación se desarrolla en base al lenguaje de programación PHP, lo cual permitirá una conexión a la base de datos y consultar la información requerida y así poder trabajar con ella dentro de la aplicación, el lenguaje php aportará la parte dinámica de la aplicación permitiendo realizar trabajos por roles de usuario, además de insertar, consultar, modificar, eliminar y mostrar los datos necesarios para que el sistema funcione de acuerdo a criterios particulares de configuración, adicionalmente, permitirá incluir los módulos inherentes al trabajo con los algoritmos de Machine Learning (probados y validados previamente con Python) y así poder realizar predicciones conforme a los datos.
IN: Servidor Web		Se utilizara el servidor Web apache, que tendrá alojada la aplicación, este servidor estará instalado en un equipo de cómputo con sistema operativo Linux Ubuntu 20.04LTS
IN: Python	6. Evaluación	Esta fase, paralelo con la fase 4, se hará mediante la codificación en Python, puesto que, las métricas de incertidumbre (Accuracy, Precision, Recall y F1-Score) se reportan dentro del código de los algoritmos, usando la importación de módulos que generan la matriz de confusión.
IN: Microsoft Excel		Conforme se generen las métricas de incertidumbre, éstas se van a resumir en una hoja de cálculo (Excel) para comparar y visualizar de manera sencilla los diferentes rendimientos de los algoritmos en torno a las variables predictoras.
IN: Lenguaje de programación PHP	7. Integración del modelo	Haciendo uso del lenguaje php, se integrará de la manera más pedagógica posible, el modelo predictivo validado en la fase anterior.
IN: Aplicativo Web	8. Estrategia	Con base a los resultados de la investigación, se fomenta una propuesta para la mejora de prácticas pedagógicas que se explicita en la tabla 19.

Fuente: Elaboración propia

7. ANÁLISIS DE RESULTADOS

Como se menciona en el capítulo 6, se trabajó inicialmente con estudiantes de los grados décimo de las instituciones educativas Ramón Alvarado y La Merced, de los municipios de Garzón y Agrado, respectivamente. A continuación, de acuerdo a las acciones planteadas para el desarrollo de cada fase (ver tabla 13), y de la relación metodológica (Ver figura 26), se describen los procesos desarrollados para el alcance de los objetivos planteados en la investigación.

Fase 1. Caracterización

En la tabla 9, se resumen las generalidades de la población tomada para esta fase preliminar de la investigación. Adicionalmente, de acuerdo con lo explicitado en la tabla 8, respecto a las variables sociodemográficas consideradas, en el anexo 2, se adjunta la encuesta aplicada de manera virtual usando googleforms como herramienta ideal para cuestionarios, debido a su fácil acceso y practicidad para descargar bases de datos en diferentes formatos. Particularmente, en .xlsx como extensión ideal para limpiar y pre procesar los datos usando Microsoft Excel. A continuación, en la tabla 13, se relacionan las convenciones que, para efectos del uso de los diferentes algoritmos machine learning, permitirán la lectura de los datos sacados del cuestionario.

Tabla 13. Variables de entrada, convenciones y numeración de sub categorías

No.	Variable	Convención	Sub Categoría	Equivalencia numérica
1	Tipo de residencia	TP_RES	Rural	0
			Urbano	1
2	Tiene acceso a internet	INT	Sí	1
			No	0
3	seguridad de la zona	SEG_ZO	Baja	0
			Media	1
			Alta	2
4	Lejanía a la institución	LEJA	Menos de 15 min	0
			15 a 30 min	1
			30 a 60 min	2
			Más de 1h	3
5	Actividades complementarias a las que acude	ACT_COMP	Ninguna	0
			1 actividad	1
			2 actividades	2
			3 actividades	3
6		RIES_DR_ALC	Baja	0



No.	Variable	Convención	Sub Categoría	Equivalencia numérica
	Riesgo a drogas o alcohol en la zona		Media	1
			Alta	2
7	Tipo de familia	TP_FAM	elemental	0
			consanguínea	1
			monoparental	2
			padres separados	3
			vive con otro familiar	4
			homoparental	5
8	Número de integrantes de la familia	NU_INT	entrada numérica	0,1,2,3,...
9	Usted ha sido educado de manera	TP_EDU	autoritaria	0
			democrática	1
			permisiva	2
			indiferente	3
10	Nivel de ingresos	INGR	1 salario mínimo	1
			1 y 2 salarios	2
			más de 2 salarios	3
11	Escolaridad de la madre	ESC_M	Primaria	0
			Básica secundaria	1
			Bachillerato	2
			Técnico o tecnólogo	3
			Universitario	4
			Ns/Nr	5
12	Escolaridad del padre	ESC_P	Primaria	0
			Básica secundaria	1
			Bachillerato	2
			Técnico o tecnólogo	3
			Universitario	4
			Ns/Nr	5
13	Relación con la madre	REL_M	Mala	0
			Regular	1
			Buena	2
			excelente	3
			Ns/Nr	4
14	Relación con el padre	REL_P	Mala	0
			Regular	1
			Buena	2
			excelente	3
			Ns/Nr	4
15	Tratamiento de los conflictos	CONFL	Diálogo	0
			discusiones	1
16	Considera que la Institución cuenta con buena infraestructura y aulas adecuadas para el aprendizaje	INFR	Mala	0
			Regular	1
			Buena	2
			excelente	3
17		REL_C	Mala	0

No.	Variable	Convención	Sub Categoría	Equivalencia numérica
	Relación con sus compañeros		Regular	1
			Buena	2
			excelente	3
18	Relación con sus docentes	REL_D	Mala	0
			Regular	1
			Buena	2
			excelente	3
19	Usualmente, cuánto tiempo al día dedica a leer por entretenimiento	LEE	no lee por entretenimiento	0
			30 minutos o menos	1
			entre 30 y 60	2
			entre 1 y 2	3
			más de 2 horas	4
20	Se siente a gusto y considera que la institución brinda apoyo en su proceso formativo	APOY	muy poco	0
			Regular	1
			a gusto	2
			muy a gusto	3
21	Reprobación en años anteriores	REPR	entrada numérica	0,1,2,3,...
22	Edad	ED	entrada numérica	0,1,2,3,...
23	Género Biológico	GEN_B	Femenino	0
			Masculino	1

Fuente: Elaboración propia

De manera continua, se recopilaron los desempeños de los estudiantes encuestados en las áreas relacionadas con las competencias básicas, a saber: Matemáticas, Ciencias Naturales, Ciencias Sociales, Inglés y Lectura crítica; puesto que, serán estas las variables de salida o a predecir. En ese sentido, la base de datos se complementa con tales variables y sus respectivas categorías. La tabla 14, muestra las convenciones asignadas.

Tabla 14. Variables de salida, convenciones y numeración de sub categorías

No.	Variable	Convención	Sub Categoría	Equivalencia numérica
24	Matemáticas	MAT	Reprueba	0
			Aprueba	1
25	Ciencias Naturales	CN	Reprueba	1
			Aprueba	0
26	Ciencias Sociales	CS	Reprueba	0
			Aprueba	1
27	Inglés	ING	Reprueba	0
			Aprueba	1
28	Lectura Crítica	LEC	Reprueba	0
			Aprueba	1

Fuente: Elaboración propia

Se debe destacar que, los valores de desempeños para cada una de las asignaturas relacionadas con las cinco competencias básicas valoradas en el sistema educativo colombiano, tiene cuatro sub categorías en los colegios: Bajo, Básico, Alto y Superior. No obstante, con el objetivo de mejorar el posible rendimiento de los algoritmos, se han tomado solo dos categorías: Reprobado (0) y Aprobado (1), así, se tendrá una salida binomial que, dentro de las necesidades interpretativas de la investigación y el propósito general de predecir dificultades, es suficiente.

Fase 2. Comprensión de los datos.

Como se presentó en la tabla 13, los datos en su mayoría son de tipo categóricos, por ello, se hizo a partir de la fundamentación teórica relacionada con la trascendencia e influencia de variables sociodemográficas (Apartado 4.2.3), un filtro de selección que inicia en la tabla 8 y se consolida en la tabla 13. Adicionalmente, en la tabla 14 se presentan las variables objetivo o de salida, las cuales se sustentan como asignaturas que promueven las competencias básicas (Apartado 4.2.2) y que deben sintetizarse en salidas binomiales, para considerar efectividad en los resultados de los algoritmos a aplicar, y a su vez generar lo que será un insumo para establecer relaciones y predicciones.

Conforme a lo anterior, se definen las variables que, dentro de sus relaciones implícitas, aparentemente inextricables, siendo inherentes a la singularidad de los estudiantes población de estudio, pueden considerarse como los atributos que por medio del machine learning, podrían orientar prácticas educativas que ayuden a mejorar el fomento de competencias básicas. Consecuentemente, serían los desempeños académicos en torno a las mencionadas competencias, la variable dependiente a predecir y que, en ambas instituciones, puede evidenciarse que hay dificultades. Puesto que, como se observa en las figuras 27 y 28 los niveles de desempeño en las pruebas saber, durante los últimos 3 años, no superan en promedio anual, más de los 52 puntos para el caso de La Merced y los 51 para el caso de Ramón Alvarado. Se puede inferir de esta manera, la prevalencia de desempeños bajos dentro de cada una de las cinco asignaturas y la relevancia de generar estrategias de mejora.

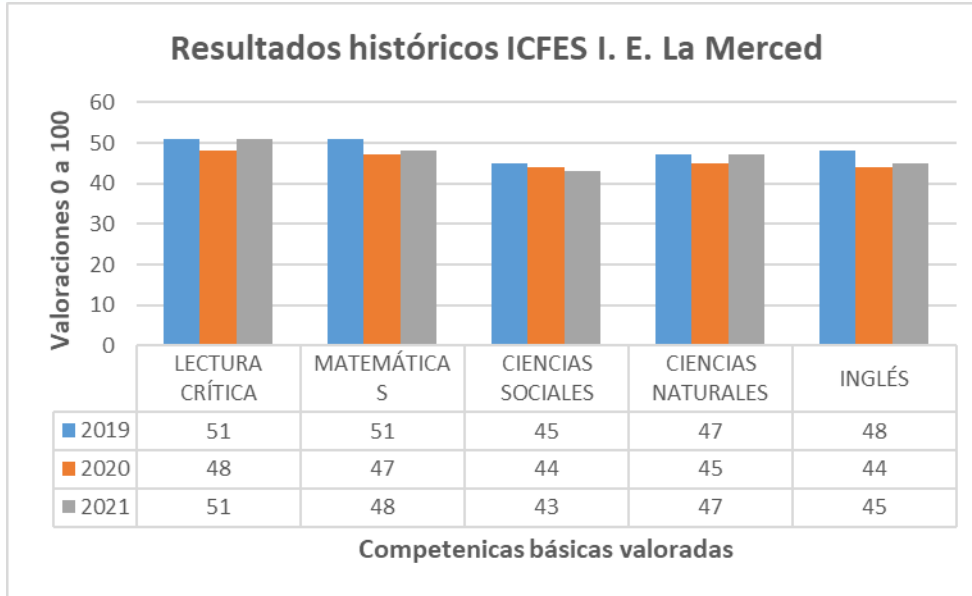


Figura 27. *Desempeños en competencias básicas I. E. La Merced*
 Fuente: Elaboración propia

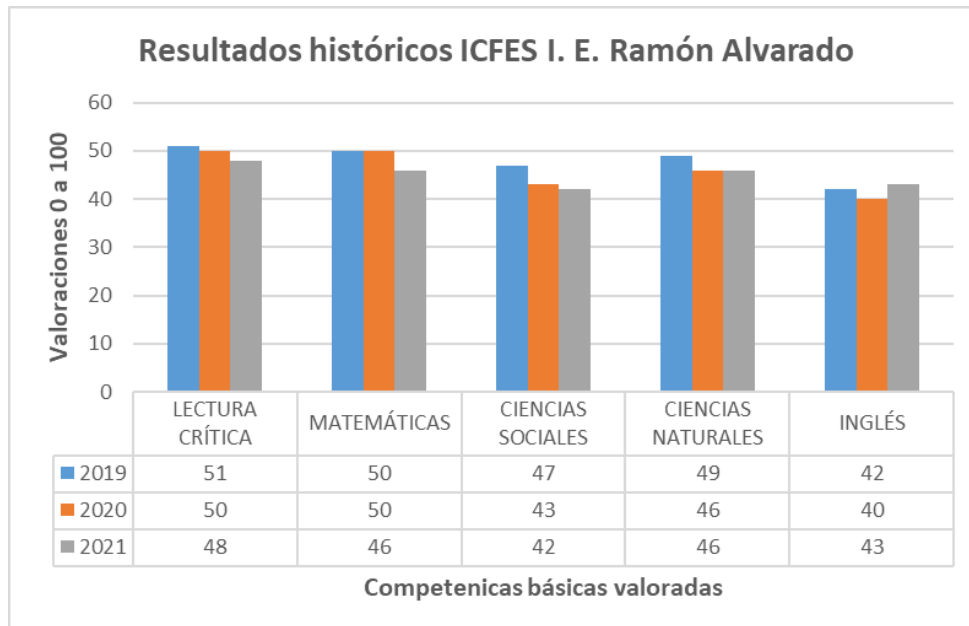


Figura 28. *Desempeños en competencias básicas I. E. Ramón Alvarado*
 Fuente: Elaboración propia

En este orden de ideas, se tomaron 92 muestras con sus respectivos atributos correspondientes a los estudiantes de grado décimo (población) de ambas instituciones y se obtiene la base de datos que se muestra en la figura 29.

Nombre	Número de muestra	Lugar de residencia	En su lugar de residencia	Seguridad de la zona	Lejanía a la institución	Modo de desplazamiento	Clase social	Grupo social especial: Actividades complementarias	Riesgo a drogas
sanchez	1	123456789 Urbano	Acueducto, Alcantarill; Media		Menos de 15 min	Caminando		2 No pertenezco a ningun	Ninguna Alta
herrera palma bustos	2	12279381 Rural	Acueducto, Alcantarill; Media		Menos de 15 min	Moto		1 No pertenezco a ningun	Deportes Media
Canvajal Morales	3	1.023E+09 Urbano	Electricidad, Gas dom; Media		Menos de 15 min	Moto		1 No pertenezco a ningun	Deportes Baja
orjuela mancera	4	1.079E+09 Urbano	Acueducto, Electricida; Media		15 a 30 min	Moto		1 Víctima de conflicto	Ninguna Baja
sandoval cruz	5	1.079E+09 Rural	Electricidad, Televisor; Media		30 a 60 min	Transporte escolar		1 No pertenezco a ningun	Ninguna Baja
navía	6	1.074E+10 Rural	Electricidad, internet, ; Media		15 a 30 min	Moto		2 Víctima de conflicto	Ninguna Baja
sofía leal rojas	7	1.06E+09 Urbano	Electricidad, Computa; Baja		15 a 30 min	Caminando		1 Desplazado	Deportes Baja
Brinyth	8	1.079E+09 Urbano	Acueducto, Electricida; Alta		Menos de 15 min	Caminando		1 No pertenezco a ningun	Pintura, Deportes, (Baja
herrera cortes	9	1.079E+09 Urbano	Acueducto, Alcantarill; Media		15 a 30 min	Caminando		1 No pertenezco a ningun	Canto y banda Baja
herrera	10	1.082E+09 Urbano	Moto	Baja	15 a 30 min	Moto		1 No pertenezco a ningun	galda Baja
Camila dussan	11	1.082E+09 Urbano	Acueducto, Alcantarill; Media		15 a 30 min	Moto		1 No pertenezco a ningun	Danzas, Pintura, D; Media
almario losada	12	1.082E+09 Urbano	Acueducto, Alcantarill; Media		Menos de 15 min	Moto		1 No pertenezco a ningun	Danzas, Teatro, De; Media
perdomo fierro	13	1.078E+09 Urbano	Acueducto, Alcantarill; Media		15 a 30 min	Caminando		1 No pertenezco a ningun	Ninguna Media
Naranjo	14	1.082E+09 Urbano	Acueducto, Alcantarill; Media		15 a 30 min	Caminando		1 No pertenezco a ningun	Ninguna Media
perdigon ruiz	15	26451127 Urbano	Acueducto, Alcantarill; Media		15 a 30 min	Moto		1 No pertenezco a ningun	Danzas Media
Nicolle Leiva	16	1.107E+09 Urbano	Acueducto, Alcantarill; Baja		Menos de 15 min	Caminando		1 No pertenezco a ningun	Deportes Baja
Jurado Chavarro	17	1.078E+09 Urbano	Acueducto, Alcantarill; Media		Menos de 15 min	Caminando		1 No pertenezco a ningun	Ninguna Baja
Quintero joven	18	1.079E+09 Urbano	Acueducto, Alcantarill; Media		Menos de 15 min	Moto		1 Afrodescendiente	Danzas, Deportes Media
Hoyos Almarío	19	1.082E+09 Urbano	Acueducto, Alcantarill; Alta		15 a 30 min	Moto		1 No pertenezco a ningun	Ninguna Baja
Parra Vargas	20	1.079E+09 Urbano	Acueducto, Alcantarill; Media		Menos de 15 min	Moto		1 No pertenezco a ningun	Deportes Alta
trujillo Escalante	21	1.079E+09 Urbano	Acueducto, Alcantarill; Media		15 a 30 min	Caminando		1 No pertenezco a ningun	Ninguna Media
Ramírez	22	1.079E+09 Rural	Acueducto	Media	15 a 30 min	Caminando		0 No pertenezco a ningun	Danzas Baja
Adames Olaya	23	1.075E+09 Urbano	Electricidad, Computa; Media		Menos de 15 min	Moto		1 No pertenezco a ningun	Danzas Baja
	24	1.082E+09 Rural	Electricidad, internet, ; Baja		30 a 60 min	Moto		0 No pertenezco a ningun	Ninguna Baja

Figura 29. Algunos atributos considerados en la base de datos

Fuente: Elaboración propia

Se eligieron estudiantes de grado décimo, porque se espera apoyar el fomento de las competencias básicas y posibilitar la mejora en los desempeños en el año posterior. Se recalca, que las 92 muestras fueron tomadas para el tratamiento de las características de las mismas en cada algoritmo a probar, y que, tal y como se evidencia al final de la tabla 14, hay 28 variables⁷ (columnas) a considerar, dentro de las cuales 5 son de salida y 23 de entrada.

Fase 3. Limpieza de los datos.

Conforme se aclaró qué características debían tomarse de la población de estudio, se procede a recopilar los datos mediante un cuestionario en el que se relacionan todos los aspectos sociodemográficos y al que se le adicionan, posterior a tener un archivo .xlsx los desempeños de las cinco competencias básicas que serán objeto de predicción. La figura 29 muestra parte de los datos obtenidos y en la figura 30 se resaltan celdas que

⁷ En el ámbito de la programación, las variables se mencionan indistintamente de términos como características, atributos o dimensiones

se deben empezar a depurar por ser datos errados, incompletos o irrelevantes para efectos de análisis.

Nombre	Número de identi	3. Lugar de residencia	En su lugar de residen	Seguridad de la zona	Lejanía a la institución	Modo de desplazamier	Clase social (estrato):	Grupo social especial: Actividades comple
David Felipe	1	123456789 Urbano	Acueducto, Alcantarill	Media	Menos de 15 min	Caminando	2	No pertenezco a ningun
Duvan	2	12279381 Rural	Acueducto, Alcantarill	Media	Menos de 15 min	Moto	1	No pertenezco a ningun
ordóñez	3	1023372415 Urbano	Electricidad, Gas dom	Media	Menos de 15 min	Moto	1	No pertenezco a ningun
Canajal	4	1079388122 Urbano	Acueducto, Electricida	Media	15 a 30 min	Moto	1	Victima de conflicto Ninguna
yulian	5	1079388671 Rural	Electricidad, Televisor	Media	30 a 60 min	Transporte escolar	1	No pertenezco a ningun
andres	6	10738538837 Rural	Electricidad, internet,	Media	15 a 30 min	Moto	2	Victima de conflicto Ninguna
danilo	7	105906738 Urbano	Electricidad, Computa	Baja	15 a 30 min	Caminando	1	Desplazado Deportes
rojas	8	1078748735 Urbano	Acueducto, Electricida	Alta	Menos de 15 min	Caminando	1	No pertenezco a ningun
Brinyth	9	1079388998 Urbano	Acueducto, Alcantarill	Media	15 a 30 min	Caminando	1	No pertenezco a ningun
alexandra	10	1082154084 Urbano	Moto	Baja	15 a 30 min	Moto	1	No pertenezco a ningun
lorena	11	1082155504 Urbano	Acueducto, Alcantarill	Media	15 a 30 min	Moto	1	No pertenezco a ningun
dussan	12	1082155503 Urbano	Acueducto, Alcantarill	Media	Menos de 15 min	Moto	1	No pertenezco a ningun
jasbleidy	13	1077851316 Urbano	Acueducto, Alcantarill	Media	15 a 30 min	Caminando	1	No pertenezco a ningun
yurany	14	1082154811 Urbano	Acueducto, Alcantarill	Media	15 a 30 min	Caminando	1	No pertenezco a ningun
Ciceri	15	26451127 Urbano	Acueducto, Alcantarill	Media	15 a 30 min	Moto	1	No pertenezco a ningun
victoria	16	1106515680 Urbano	Acueducto, Alcantarill	Baja	Menos de 15 min	Caminando	1	No pertenezco a ningun
Leiva	17	1077851711 Urbano	Acueducto, Alcantarill	Media	Menos de 15 min	Caminando	1	No pertenezco a ningun
Alexis	18	1079388979 Urbano	Acueducto, Alcantarill	Media	Menos de 15 min	Moto	1	Afrodiescendiente Danzas, Deportes
Stefany	19	1082155624 Urbano	Acueducto, Alcantarill	Alta	15 a 30 min	Moto	1	No pertenezco a ningun
Fernando	20	1079388887 Urbano	Acueducto, Alcantarill	Media	Menos de 15 min	Moto	1	No pertenezco a ningun
Javier	21	1079388889 Urbano	Acueducto, Alcantarill	Media	15 a 30 min	Caminando	1	No pertenezco a ningun
lorena	22	1079388871 Rural	Acueducto	Media	15 a 30 min	Caminando	0	No pertenezco a ningun
Alejandra	23	1074782041 Urbano	Electricidad, Computa	Media	Menos de 15 min	Moto	1	No pertenezco a ningun
Adames	24	1082155394 Rural	Electricidad, internet,	Baja	30 a 60 min	Moto	0	No pertenezco a ningun
andres	25	1079389073 Urbano	Acueducto, Alcantarill	Media	15 a 30 min	Caminando	1	No pertenezco a ningun

Figura 30. Muestra de la base de datos inicial

Fuente: Elaboración propia

Continuamente, para evitar la redundancia de dimensionalidad (variables de entrada o atributos) al momento de codificar y empezar a probar algoritmos de clasificación, se pasa de la tabla mostrada en la figura 29 a una con menos ruido y que permita efectividad en su tratamiento desde los lenguajes de programación a usar, para ello, se establecieron las convenciones explícitas en la tabla 13 y 14. La figura 31 muestra parte de la base de datos final obtenida y con la que se prueban los algoritmos dentro de la plataforma anaconda.

TP_RES	INT	SEG_ZO	LEJA	ACT_COMP	RIES_DR	ALCTP_FAM	UN_INT	TP_EDU	INGR	ESC_M	ESC_P	REL_M	REL_P	CONFL	INFR	
1	1	1	1	0	0	2	0	5	1	1	2	2	2	2	0	1
0	0	1	0	1	1	0	0	5	0	2	1	2	4	3	0	2
1	0	1	0	1	0	0	0	4	2	1	4	1	3	3	0	2
0	0	1	2	0	0	3	2	0	1	2	0	3	0	0	0	1
0	1	1	1	0	0	0	0	4	2	1	0	1	3	3	0	2
1	0	1	1	2	0	2	3	1	1	0	0	2	3	0	0	1
1	0	0	1	1	0	0	0	4	1	1	2	0	2	2	0	1
1	1	1	1	3	1	4	3	0	1	2	0	1	2	0	0	1
1	1	1	0	3	1	0	3	0	1	2	4	3	3	0	0	1
1	0	1	1	0	1	4	4	1	2	1	1	3	3	0	0	1
1	1	1	1	1	1	0	5	1	2	3	1	2	3	0	0	2
1	1	0	0	1	0	0	0	5	1	1	0	2	3	3	0	1
1	1	1	0	0	0	2	2	2	2	1	0	0	2	2	0	2
1	1	1	0	2	1	1	3	1	1	4	4	2	4	0	0	1
1	1	2	1	0	0	0	0	4	0	1	4	4	2	2	0	2
1	1	1	0	1	2	2	2	2	1	1	3	2	3	1	0	1
0	0	1	1	1	0	0	4	0	3	1	1	3	3	0	0	3
1	1	1	0	1	0	0	0	3	1	1	2	3	2	2	0	1
1	1	1	1	1	1	1	1	6	0	1	2	0	3	4	0	2
1	0	1	1	1	1	2	6	1	1	1	1	1	1	2	0	2
1	1	1	0	0	2	0	6	1	1	4	5	3	3	0	0	3

Figura 31. Dataset preliminar: 92 muestras

Fuente: Elaboración propia

Superada la fase de limpieza de los datos, se tiene todo a disposición para empezar a hacer pruebas de algoritmos de predicción.

Fase 4. Prueba del modelo ML.

Dentro de esta fase, se requiere seguir algunos pasos que de manera lógica deben propiciar una buena práctica dentro del campo de la Inteligencia Artificial. De este modo, en primera instancia, se considera el tipo de aprendizaje en el que se enmarca el problema a resolver. Así, como se expresa de forma más precisa en el marco teórico del trabajo (Ver apartado 4.2.1.3), se podrían usar algoritmos para el aprendizaje supervisado como para el no supervisado. No obstante, de acuerdo a las fases anteriores, es necesario orientar el abordaje de acuerdo a la naturaleza del problema a tratar (Ver figura 7).

En sentido anterior, tomando en cuenta que, la salida buscada es categórica: aprueba o reprueba, en cada una de las asignaturas seleccionadas; y que, se busca entrenar alguno de los algoritmos para luego validar los rendimientos de predicción, entonces la investigación se ubicó dentro del aprendizaje supervisado y ya que, la salida no es numérica, los algoritmos de regresión no serían los indicados sino los de clasificación. Ahora, dentro del amplio compendio de estos, de acuerdo a los referentes bibliográficos expuestos en el apartado 4.2.1.4, se consideraron los que, más allá de sus características de funcionamiento en campos principalmente mercantilistas, podrían ser aplicados al campo educativo. De ahí que se expliciten por sus nombres en inglés: Logistic Regresion (LR), Decision Tree (DT), K nearest neighbors (KNN), Neural net (NN) y Support vector machine (SVM) como los algoritmos a probar para posteriormente validar de acuerdo a las métricas de rendimiento.

Tomando las 92 muestras (filas) y las 23 variables de entrada (columnas) para valorar las relaciones que puedan tener con la salida probada (columna 24: MAT, CN, CS, ING, LEC; Ver anexo 3), en la tabla 15, se resumen los rendimientos obtenidos en cada uno de los ya mencionados algoritmos. Cada uno ha sido codificado desde el uso de Python, a través de la plataforma Anaconda y el interfaz Spyder.

Se aclara que, el algoritmo probado debe correrse con cada salida en cuestión, de ahí que en la tabla se sintetizan las métricas inherentes a cada una de estas.

Tabla 15. Comparación de rendimientos de algoritmos machine learning

PRUEBA DE ALGORITMOS																					
		MÉTRICAS DE INCERTIDUMBRE PARA LAS SALIDAS																			
TIPO	ALGORITMO	MAT				CS				CN				ING				LEC			
		PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE
CLASIFICADOR	LR	0.94	0.94	1	0.96	1	1	1	1	0.94	0.94	1	0.96	0.87	0.82	0.93	0.9	1	1	1	1
	DT	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	KNN	0.83	0.91	0.91	0.87	0.83	0.91	0.91	0.87	0.76	0.87	0.87	0.81	0.48	0.7	0.7	0.57	1	1	1	1
	SVM	1	1	1	1	0.8	0.89	0.89	0.85	1	1	1	1	0.5	0.53	0.53	0.51	1	1	1	1
	NN	1	1	1	1	0.85	0.74	0.74	0.78	1	0.89	0.89	0.94	0.62	0.63	0.63	0.62	err	err	err	err

Fuente: Elaboración propia

Fase 5. Diseño del Entorno Virtual.

Conforme se empezó la prueba de algoritmos, de manera paralela, se fue construyendo lo que será un entorno virtual educativo para posibilitar el diagnóstico de estudiantes en torno a las dificultades académicas que, en medio de la incertidumbre inherente a la singularidad per se, pueden llegar a presentar. En este sentido, a continuación, se enmarcan las características consideradas para hacer uso tanto de PHP como de Python para codificar tal aplicativo.

De acuerdo con Heuter (2016), Ortiz y Rojas (2018), el uso del lenguaje PHP para el desarrollo de la aplicación se debe a que es ampliamente conocido en el ámbito del desarrollo web, esto, ya que se encuentra entre los lenguajes que se ejecutan en el lado del servidor (backend) cumpliendo ciertas características que lo hacen óptimo para el trabajo en los servidores de aplicación.

Entre otras propiedades del lenguaje se puede destacar que es de código abierto, lo cual permite que gran cantidad de personas y comunidades tengan libertad de uso, lo que fomenta una constante evolución del código y variedad de funcionalidades, consecuentemente, ha resultado mucha documentación acerca del lenguaje en cuestión, además de las orientaciones para vincularse con otros lenguajes como Python, el cual también es de código abierto pero tiene como ventaja –además de la facilidad para

aprenderse—, su potencia en el manejo de grandes cantidades de datos, lo que conlleva a su elección para utilizarse en el ámbito del machine learning; al ser código abierto, existe también una amplia comunidad detrás del desarrollo de este lenguaje, permitiendo que exista una gran cantidad de módulos de desarrollo enfocados a la aplicación de la ciencia de datos (Delgado, 2022 y Gonzalo, 2019), de ahí que pueda integrarse con relativa facilidad en lenguajes versátiles como PHP.

Del modo anterior, se establece la importancia de Python para el trabajo con bases de datos en investigaciones relacionadas con la inteligencia artificial y cómo importa a la presente investigación vincular aspectos inherentes a éste, dentro de la codificación del aplicativo web que se da en PHP. Se destaca que tal vinculación es posible gracias a lo ya mencionado de ambos lenguajes y a la simplicidad de sintaxis, adicionalmente, se reconoce la relevancia del lenguaje PHP como código base para el website, porque es soportado por distintas plataformas y servidores web que evitan configuraciones adicionales ya que viene integrado por defecto en estos sistemas.

Una última característica que justifica el uso de PHP, es que puede interactuar con diferentes componentes como es el lenguaje de marcado de hipertexto HTML que es un elemento primordial para visualizar una página web y no es necesario la instalación de módulos adicionales para lograr dicha relación, también es compatible con múltiples gestores de bases de datos importantes para almacenamiento de información y entre estas características estriban los factores de su elección en el presente trabajo.

A manera de síntesis, en las figuras 32 y 33 se visualiza la codificación en PHP y el producto generado: Screen del aplicativo Web.

```

1 <?php
2 /* este archivo va a contener la clase que va a hacer referencia a la tabla personal "Cursos" */
3
4 //require_once('Model.php');/*este quiere decir que vamos a requerir una sola vez el archivo "Model.php" que para este
5 // caso se encuentra en la misma carpeta. la sentencia require_once verificará si el archivo
6 // ya ha sido incluido y si es así. no se incluye (require) de nuevo*/
7
8 class CursosModel extends Model{ /*se crea la clase PersonalModel que hereda de la clase abstracta Model. Una clase que hereda de una
9 // clase abstracta debe de ejecutar todos los metodos abstracto de la clase que heredo*/
10
11 //Como en este caso vamos a utilizar la tabla cursos entonces esta tabla tiene 2 campos que son "id cursos", "curso", entonces se utiliza 2
12 // atributos de esta clase*/
13 public $id_cursos; //estos son los 7 atributos de la clase para los campos de la tabla Personal
14 public $curso;
15
16
17 /*Ahora se crea el constructor de la clase*/
18 public function __construct(){
19     $this->db_name='bd_rms'; /*este atributo se heredo de la clase Model y por lo tant lo puede utilizar esta clase que heredo*/
20 }
21
22
23 /*si esta clase hereda de una clase abstracta por obligacion en esta clase hay que definir los metodos abstractos de la clase
24 padre porque si no se hace aparece un error*/
25 public function set($personal_data=array()){ /*La variable $personal_data son los datos de una nueva persona en la tabla personal
26 // definimos que el tipo de datos que va a recibir este parametro es array() "$personal_data=array()"
27 // Esta funcion por medio de la variable $personal_data debe de recibir los datos que va a ingresar
28 // a la base de datos*/
29 foreach($personal_data as $key=>$value){ /*Este comando sirve para recorrer arrays, el primer comando es para indicar el arreglo a recorrer
30 // y los segundos son las asignaciones de la llave(campo) y el valor(valor del campo) del array */
31 //La $$key de abajo es una variable de variable se le dice así porque tiene doble signo de "$$".
32 // $$key=$value; /*aca se iguala el valor de la llave al valor, esto se hace ya que se esta utilizando el concepto de "variables de una variable",
33 // osea que el valor del indice que contiene la llave $key puede guardar el valor al que apunta le indice por ser de tipo
34 // "variables de una variable", esto se hace para limpiar el query para insertar a la base de datos, en otras palabras
35 // se hace para que el query sea correcto y no genere errores en la base de datos.
36 }
37 }
38 }
39 }
40 }
41 }
42 }
43 }
44 }
45 }
46 }
47 }
48 }
49 }
50 }
51 }
52 }
53 }
54 }
55 }
56 }
57 }
58 }
59 }
60 }
61 }
62 }
63 }
64 }
65 }
66 }
67 }
68 }
69 }
70 }
71 }
72 }
73 }
74 }
75 }
76 }
77 }
78 }
79 }
80 }
81 }
82 }
83 }
84 }
85 }
86 }
87 }
88 }
89 }
90 }
91 }
92 }
93 }
94 }
95 }
96 }
97 }
98 }
99 }
100 }

```

Figura 32. Código del aplicativo web: PHP
Fuente: Elaboración propia

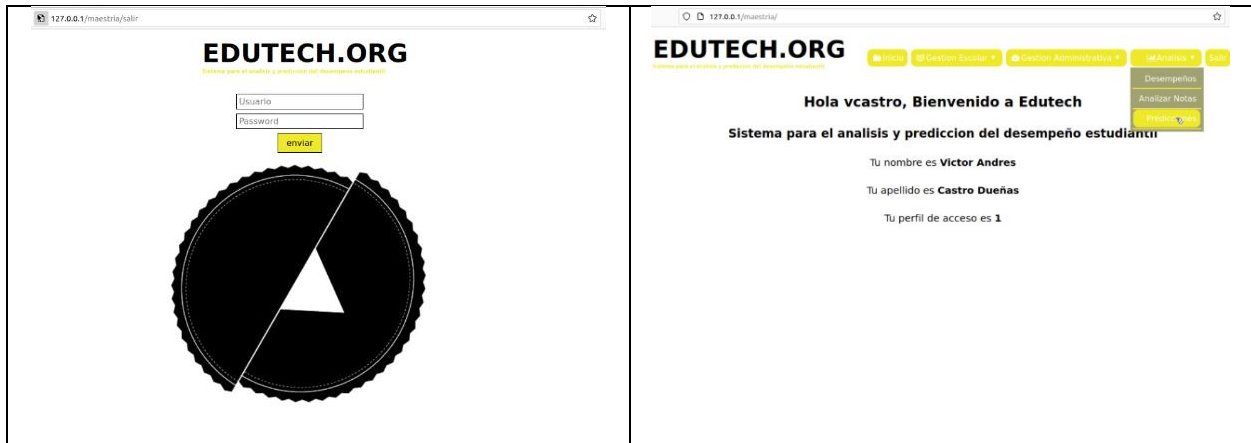


Figura 33. Screen del aplicativo Web generado⁸
Fuente: Elaboración propia

Por su parte, en la figura 34, se evidencia el uso de Python dentro del interfaz Spyder en la plataforma Anaconda, para la prueba de algoritmos machine learning que, posteriormente se integran al código en PHP.

⁸ En los anexos 5 y 6 se explicitan los códigos del aplicativo de manera más amplia y la dirección web del aplicativo.

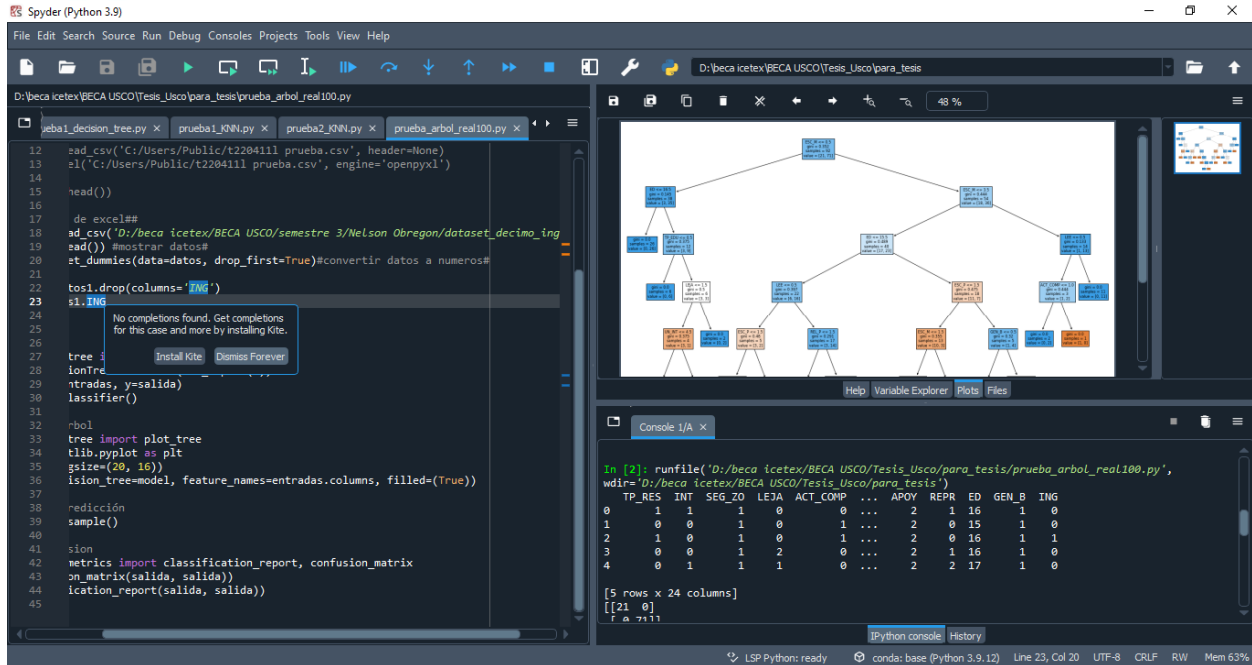


Figura 34. *DecisionTreeClassifier usando Python*
Fuente: Elaboración propia

Fase 6. Evaluación.

Con base a la prueba de algoritmos que se resume en la tabla 15, a continuación, se explicitan las gráficas radiales que permiten visualizar los rendimientos de las métricas en cada una de las salidas. Así, no solo se facilita la interpretación de las métricas, sino que, además, se puede tomar una mejor decisión en cuanto a la elección de la herramienta pertinente al trabajo investigativo.

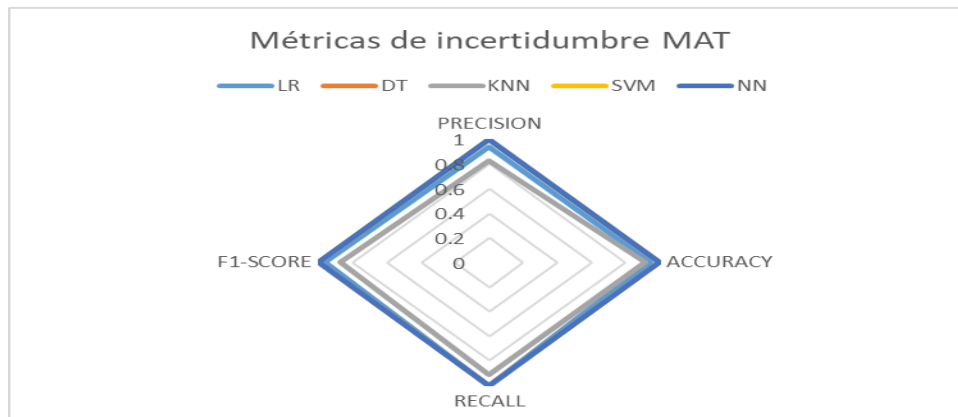


Figura 35. *Rendimiento de los algoritmos para la salida Matemáticas*

Fuente: Elaboración propia

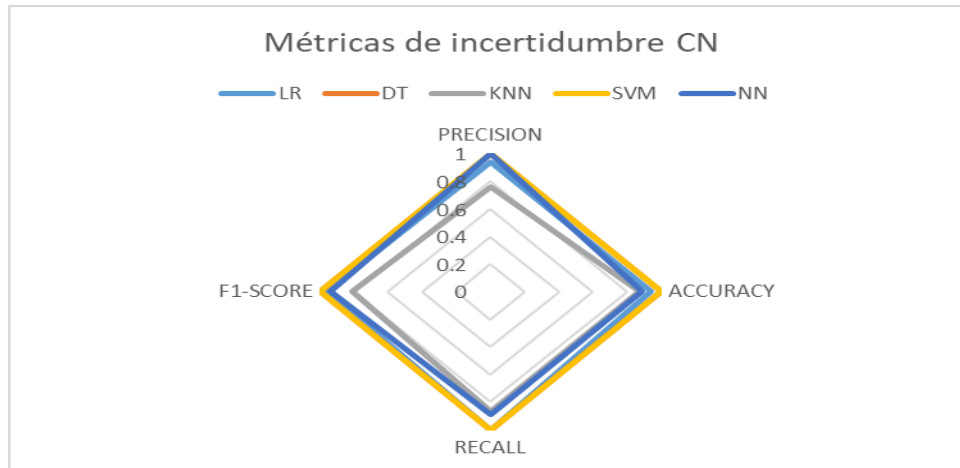


Figura 36. Rendimiento de los algoritmos para la salida Ciencias Naturales
Fuente: Elaboración propia

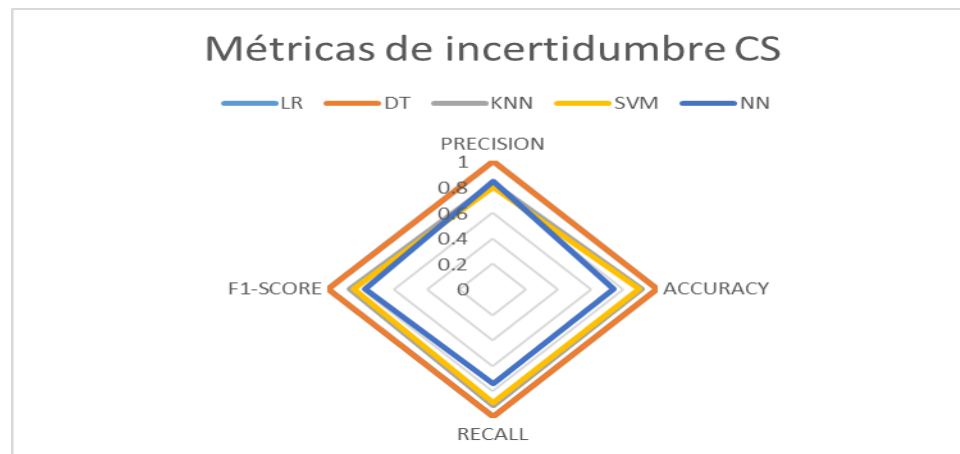


Figura 37. Rendimiento de los algoritmos para la salida Ciencias Sociales
Fuente: Elaboración propia

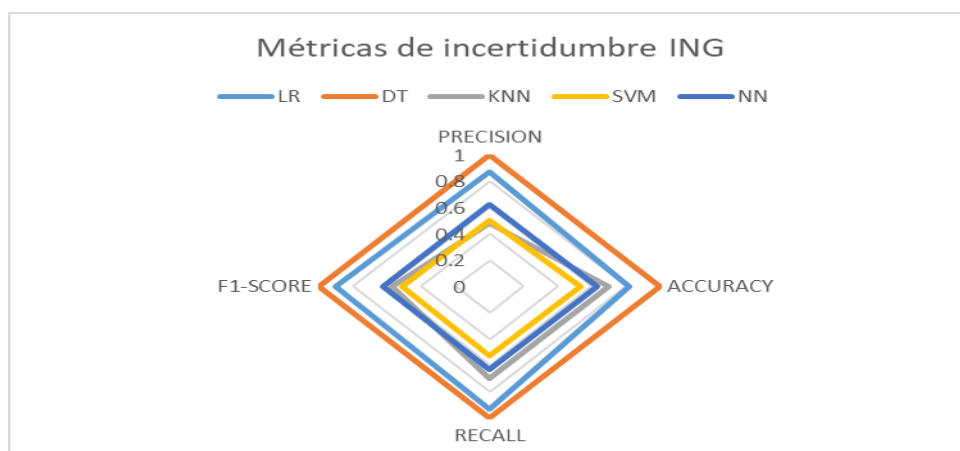


Figura 38. Rendimiento de los algoritmos para la salida Inglés

VIGILADA MINEDUCACIÓN

Fuente: Elaboración propia

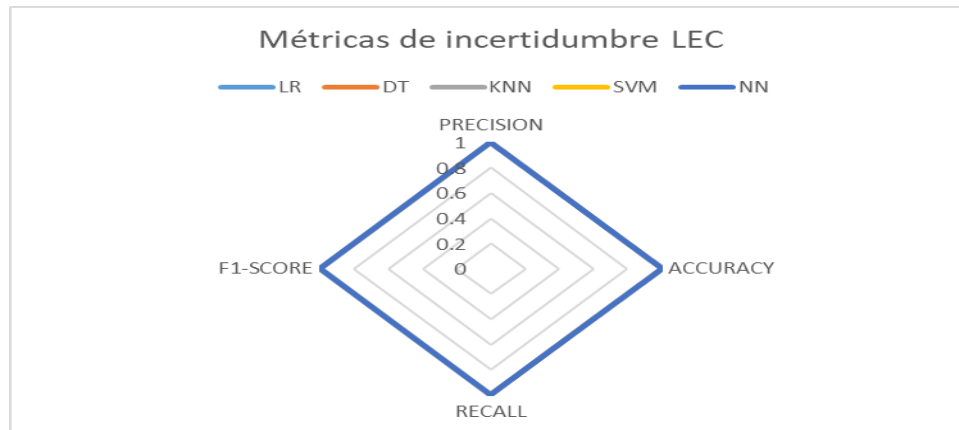


Figura 39. Rendimiento de los algoritmos para la salida Lectura
Fuente: Elaboración propia

Las anteriores gráficas dejan de manifiesto varios aspectos a considerar para garantizar la objetividad en el análisis de las métricas de incertidumbre, de cara a los propósitos de la investigación.

Inicialmente, en la tabla se puede apreciar que, para Lectura, hay por lo general, 100% de efectividad en las métricas de rendimiento de los algoritmos, excepto por la red neuronal (NN) que arroja error. Inspeccionando los datos, se encuentra que tal situación se debe a que no hay reprobación en dicha asignatura, por lo que los algoritmos se encuentran con simplicidad en la decisión: solo se puede decir que las entradas, en medio de su multiplicidad, generan aprobación (1) y que, en el caso de la red neuronal, no hay modo de estipular una predicción, cuando no hay nada qué predecir.

Se puede notar entonces, que principalmente en la salida inglés, se logró ver un rendimiento variable de los algoritmos y es que, en el dataset, justo inglés es el área con mayor balanceo de los datos. Así, puede considerarse que los árboles de decisión (DT), muestran el mejor rendimiento para cada una de las salidas, no obstante, es de anotar que, para Lectura, similar a lo que ocurrió con NN, este no suministró ningún gráfico de decisión.

En ese sentido se establece un evidente desbalanceo de los datos y tal situación se soluciona con dos procesos simples: 1. Reemplazar datos por otros que den peso a la

otra clase (0), o 2. Ampliar el rango de muestras para tomar más datos en los que haya en su mayoría, “samples” con la clase 0.

Conforme a lo anterior, se procede de acuerdo a la relación metodológica enmarcada en la figura 23 y fundamentada desde el apartado relacionado con herramientas *machine learning* aplicadas a la educación (4.2.1.4). Considerando como no fiable los resultados obtenidos.

Consecuentemente, se procedió a reemplazar valores y ampliar un poco la base de datos, considerando, entre otras cosas, corregir la información errada que se había suministrado inicialmente y la vinculación de algunos estudiantes que no se habían registrado en el inicio de aplicación del cuestionario sociodemográfico; para que la nueva prueba de algoritmos genere rendimientos más significativos que los ya mostrados.

En la tabla 16, se enmarcan los nuevos resultados de las métricas de incertidumbre que permitieron contrastar significado de los rendimientos en los algoritmos tras modificar la base de datos

Tabla 16. Nuevas métricas de incertidumbre

NUEVA PRUEBA DE ALGORITMOS																					
		MÉTRICAS DE INCERTIDUMBRE PARA LAS SALIDAS																			
TIPO	ALGORITMO	MAT				CS				CN				ING				LEC			
		PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE	PRECISION	ACCURACY	RECALL	F1-SCORE
CLASIFICADOR	LR	0.66	0.44	0.64	0.65	0.62	0.56	0.65	0.63	0.76	0.68	0.62	0.69	0.58	0.58	0.66	0.63	0.86	0.57	0.63	0.6
	DT	0.65	0.55	0.55	0.57	0.72	0.68	0.68	0.69	0.75	0.65	0.65	0.68	0.73	0.65	0.65	0.68	0.68	0.65	0.65	0.66
	KNN	0.65	0.68	0.68	0.64	0.87	0.84	0.84	0.82	0.75	0.68	0.68	0.7	0.71	0.72	0.72	0.71	0.41	0.48	0.48	0.44
	SVM	0.52	0.52	0.52	0.52	0.42	0.52	0.52	0.47	0.6	0.6	0.6	0.6	0.58	0.56	0.56	0.57	0.52	0.52	0.52	0.52
	NN	0.6	0.64	0.64	0.61	0.59	0.4	0.4	0.42	0.61	0.44	0.44	0.48	0.61	0.36	0.36	0.4	0.6	0.36	0.36	0.38

Fuente: Elaboración propia

Conforme se visualizaron los nuevos valores que arrojó cada algoritmo en las diferentes competencias evaluadas, se pudo esclarecer que, todos los algoritmos tuvieron una disminución significativa en sus rendimientos. Tales variaciones pueden compararse con los que se resumieron en la tabla 15 y, adicionalmente, en los gráficos radiales ilustrados en la figura 40, pueden interpretarse de mejor manera.

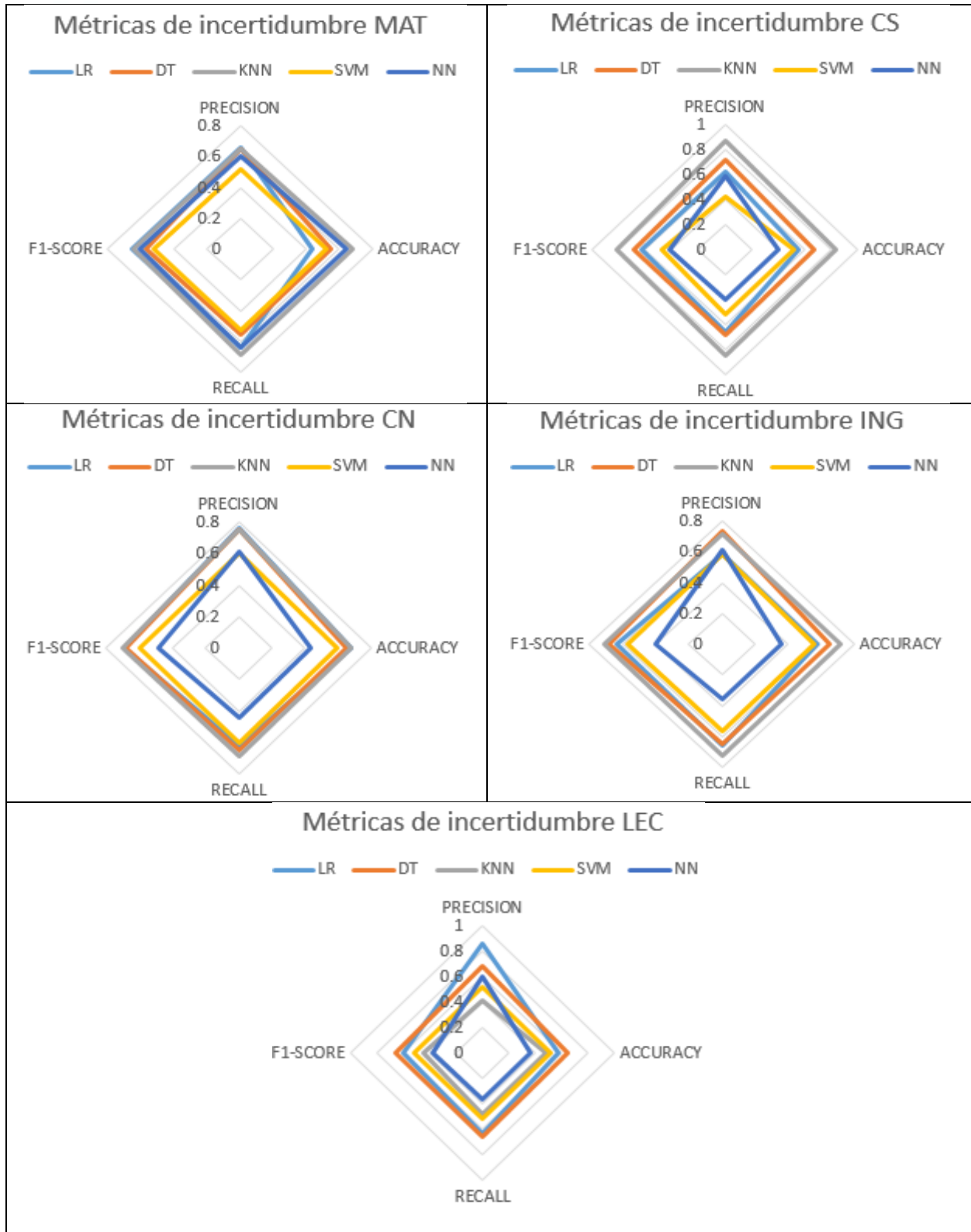


Figura 40. Nuevos rendimientos de los algoritmos ML
Fuente: Elaboración propia

VIGILADA MINEDUCACIÓN

De acuerdo a lo anterior, se destacó la mejora, no en los rendimientos de los algoritmos, puesto que, de hecho, desmejoraron las métricas, pero sí, en el significado de los nuevos valores de cara al tratamiento de un dataset más balanceado y evitando que se generaran errores en las salidas evaluadas.

Por otro lado, al ver que los árboles de decisión mantuvieron un rendimiento similar en cada una de las salidas, además de ser el algoritmo con mejores rendimientos después de KNN y que, para esta segunda prueba, el algoritmo mostró en cada una de las variables objetivo, relaciones interesantes, se estipuló su elección por ésta y las siguientes dos razones:

1. El algoritmo muestra, además, de manera explícita, la relevancia de las variables sociodemográficas que considera más influyentes en cada uno de los nodos.
2. Arroja una visualización gráfica fácilmente entendible para que, dentro de los centros educativos, pueda ser interpretado por docentes con o sin experiencia en el manejo de herramientas computacionales.

Ahora, dentro de la sencillez del mismo, se debe explicitar el funcionamiento interno del DecisionTreeClassifier. De ese modo, es de anotar que, algoritmos para árboles de decisión hay varios, pero el tomado como referente es el CART, por sus siglas en inglés: Classification And Regression Trees, que usa la librería Scikit-learn de Python.

Dicho algoritmo tiene un funcionamiento interno que lo hace, dentro de los fundamentos matemáticos, fácil de entender y de mostrar para que comunidades, sea que estén o no, familiarizadas con el ámbito computacional, logren vislumbrar la esencia de su construcción. En ese sentido, como se ve en la figura 41, las cajas con ramificaciones que se conocen como nodos y dependiendo de la posición, serán denominados: raíz, si se trata del primer nodo, ramas, si son nodos posteriores con otras ramificaciones y hojas, si son nodos terminales, es decir, sin ramificaciones posteriores. Adicionalmente, hay un hiperparámetro muy conocido en los árboles de decisión que es el nivel de profundidad y se ilustra en la ya mencionada figura 41.

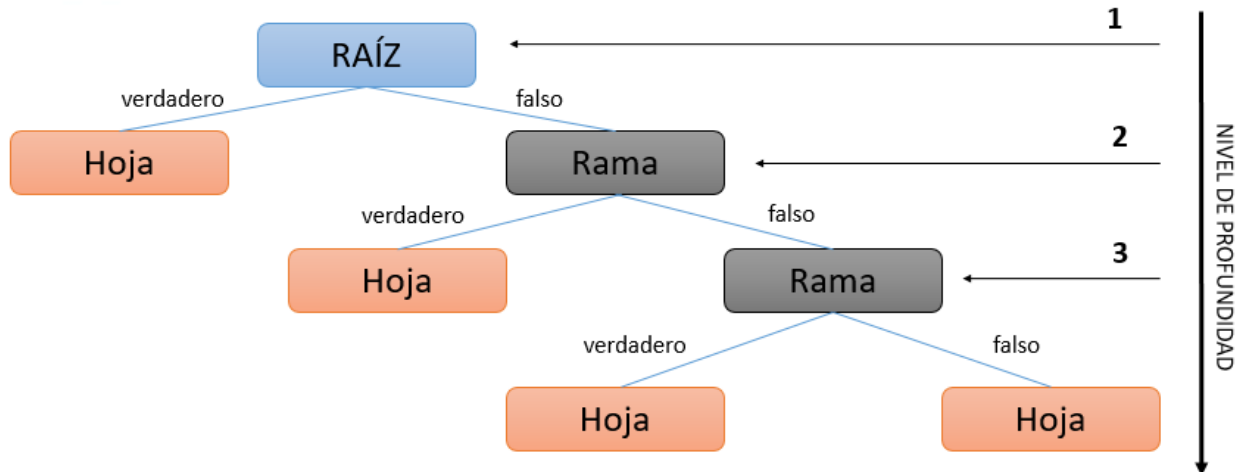


Figura 41. Esquema general de un árbol de decisión
Fuente: Elaboración propia

Conforme a lo anterior, el árbol muestra en sus nodos diferentes condiciones y si se cumplen (verdadero) o no (falso), se va por alguna de las ramas, para nuevamente enfrentar alguna condición que eventualmente culmina en una hoja. Dicho proceso estriba en las características matemáticas que se estipulan a continuación.

Los CART, funcionan mediante el índice de Gini, el cuál determina la pureza de los nodos tomando en cuenta la suma de los cuadrados de las probabilidades de las k clases. Por ejemplo: verdadero ($P(0)$) y falso ($P(1)$), mediante la expresión que se ilustra en la figura 42.

$$Gini = 1 - \sum_{k=1}^K (P_k)^2$$

Figura 42. Expresión para el índice Gini
Fuente: Elaboración propia

Conforme a lo anterior, es de resaltar que, nodos puros se darán cuando una de las clases tiene una probabilidad de ocurrencia de 1 y la otra de 0, puesto que, en esas circunstancias, el coeficiente Gini será igual a 0 y es deducible que, las muestras en dicho nodo corresponden a una sola de las clases, de ahí se genera ineludiblemente, una hoja. Por otro lado, se establece que el máximo valor del índice es 0.5 y correspondería al mayor nivel de impureza, cuya correspondencia, análoga a las

circunstancias anteriores, expondrá un número de muestras iguales para las dos clases en cuestión. Es decir, cuando éstas tengan igual probabilidad. En el anterior caso, o en el caso en el que las muestras para cada clase sean diferentes de cero, se genera un nodo hijo o una rama del árbol.

Ahora, para entender la manera en la que se elige el primer nodo, o nodo raíz, se debe pensar en que, independientemente del dataset, siempre deben existir 2 o más variables de entrada y una de estas será la raíz del árbol. Pues bien, para considerar el nodo inicial, el algoritmo estima el umbral completo de relaciones de cada variable de entrada respecto a la de salida y, en ese proceso, se encuentra una que, según el motor de la herramienta es preponderante.

En la figura 43 se muestra un ejemplo de una distribución en la que se deben clasificar distintas muestras (puntos azules y verdes) de acuerdo a dos variables (X_1 y X_2) de entrada. Como puede notarse, se podrían hacer múltiples subdivisiones para establecer un mínimo de regiones a usar para llegar en profundidad a la clasificación de los datos, tomando las dos variables de manera similar como posibles nodos de inicio. No obstante, la herramienta CART, estima la mejor partición dentro de cada una de las variables, estableciéndose la que se muestra a la derecha, por lo que, la variable que se tomará en la raíz del árbol es X_1 y no X_2 .

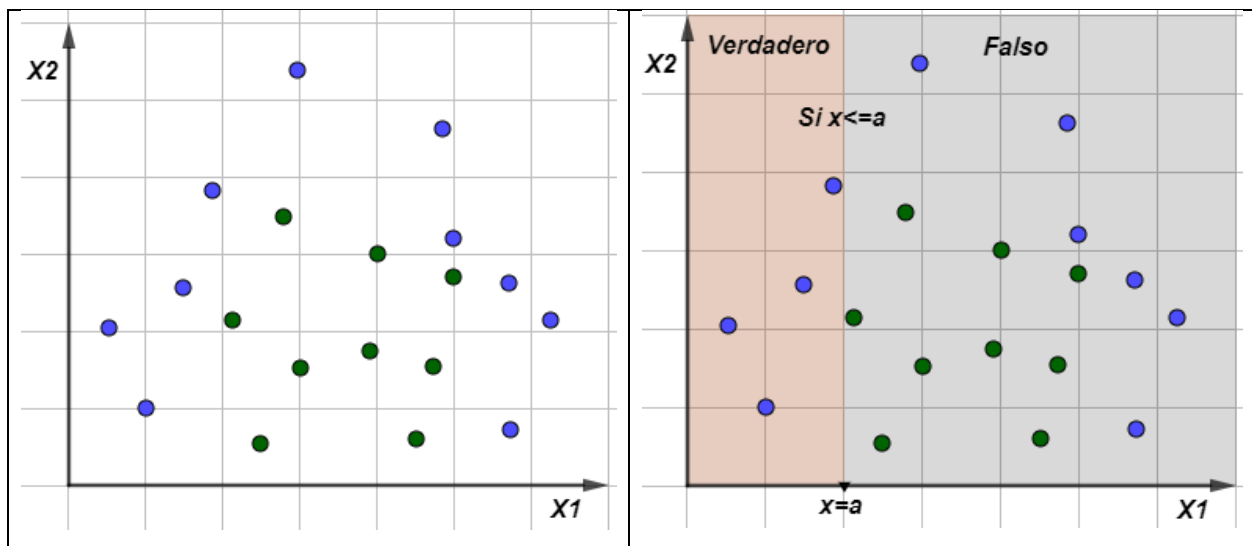


Figura 43. Idea intuitiva de elección de un nodo Raíz

Fuente: Elaboración propia

De manera intuitiva, se puede afirmar que, los nodos se eligen de acuerdo con las regiones que mitigan de manera óptima las muestras. En ese sentido, buscando generar regiones no solapantes (rectángulos de distintas dimensiones pero que no se sobrepongan), se llega a la construcción del árbol que, eventualmente, no solo clasifica, sino que, sirve como guía de diagnóstico en el presente trabajo.

En sentido más técnico y formal, el algoritmo calcula una función de costo que permite ir disminuyendo dicho valor para proporcionar un tratamiento óptimo de los datos. Así, se puede establecer -matemáticamente hablando-, la mejor partición para la creación de cada nodo. Dicha función de costo, se expresa en la figura 44 y representa básicamente un promedio ponderado de los índices Gini entre los nodos hijos.

$$F(C) = \sum_{k=1}^K G * \left(\frac{m_k}{M_T}\right)$$

Figura 44. *Función de costo en el algoritmo CART*
Fuente: Elaboración propia

En la función de costo, G es el índice Gini respectivo al nodo hijo en cuestión, m_k es la suma del número de muestras en cada uno de los nodos hijos y M_T representa el total de muestras antes de la partición. De ese modo, el algoritmo evalúa las particiones posibles y selecciona la que presente el menor valor en la función de costo, puesto que, de esa manera se garantiza que se elige la partición con el menor nivel de impureza y esto consecuentemente ha de generar una mejor clasificación.

Ahora, un valor agregado de este algoritmo, es que permite determinar cuantitativamente, la importancia de las variables o atributos, dentro de las clasificaciones. En otras palabras, se puede establecer el nivel de incidencia de las variables entrada, respecto a las de salida. Lo anterior, mediante la visualización gráfica puede entenderse, así, si en la figura 43, se generan más nodos relacionados con X_1 que con X_2 , se deduce fácilmente que el nivel de incidencia de la variable X_1 es mayor que X_2 . No obstante, para estructuras amplias, es necesario establecer dichas relevancias mediante una métrica que consiste en calcular la reducción total normalizada en el parámetro o criterio de división aplicado en el árbol, es decir, se trata de un ajuste

de los índices de Gini para las entradas, de modo que se establezca una jerarquía de incidencia entre todas las características que aparecen como nodos en el árbol de decisión.

Considerando todo lo relacionado al algoritmo seleccionado, se procedió a generar los distintos árboles de decisión para cada uno de los desempeños que buscan predecirse y en las figuras 45 a 49 se ilustran cada uno de estos esquemas.

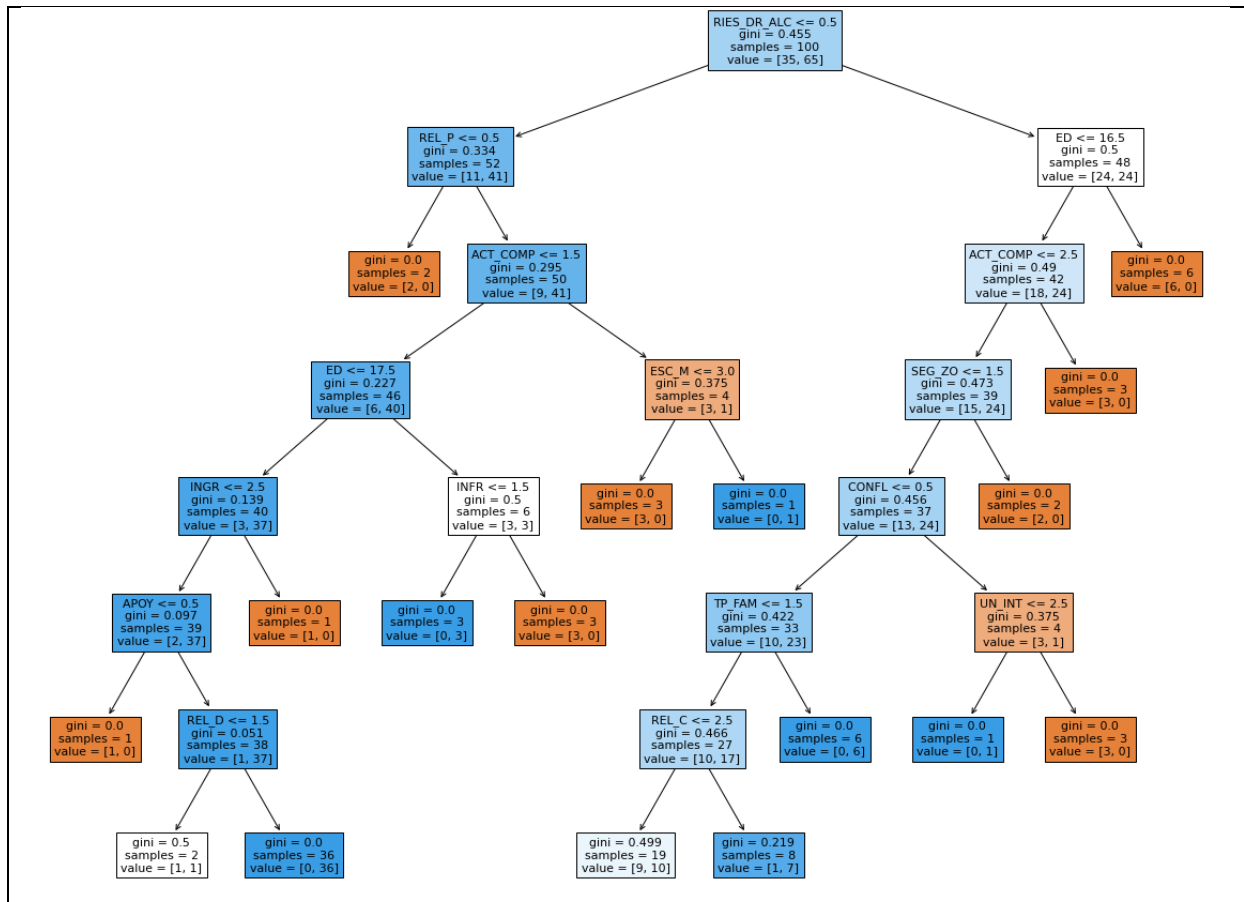


Figura 45. DT, predicción de desempeños en Matemáticas
 Fuente: Elaboración propia

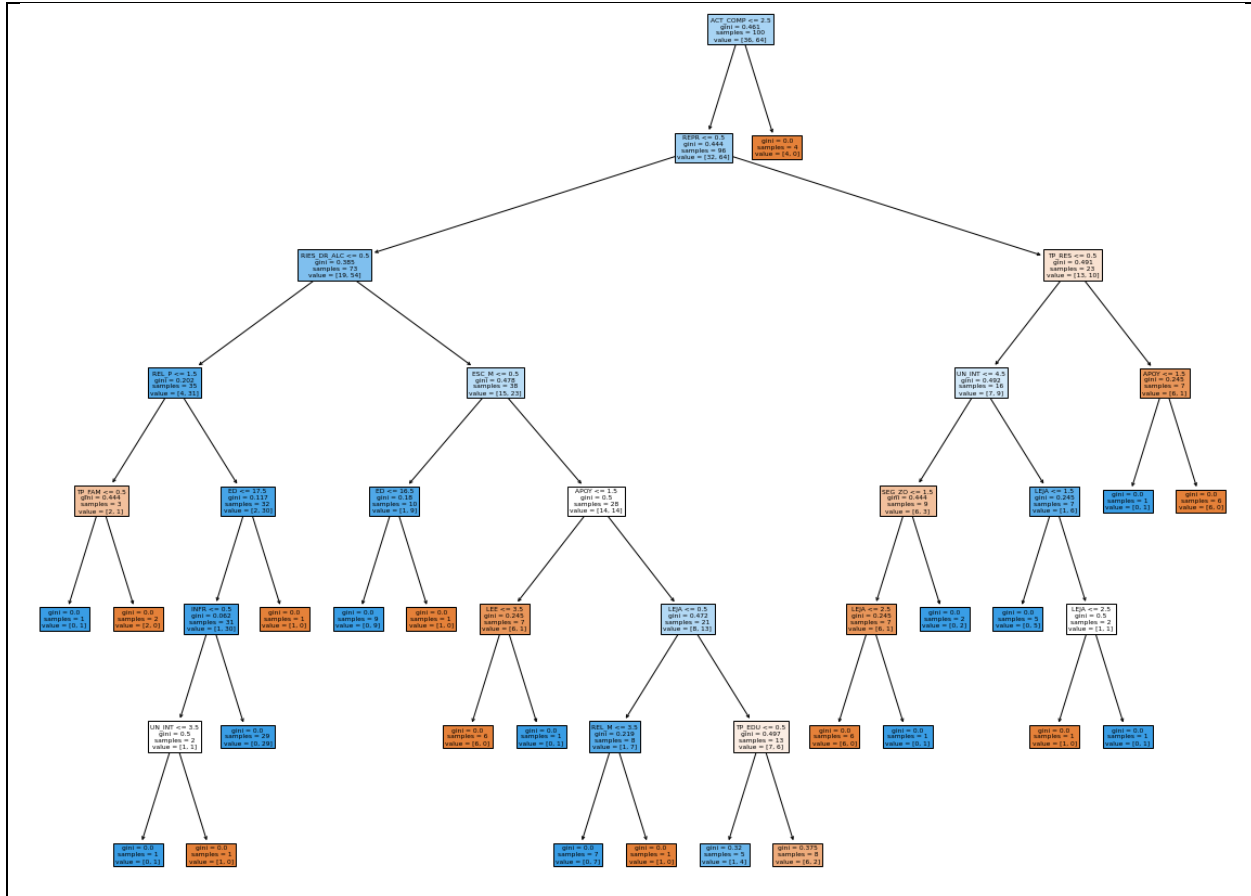


Figura 46. DT, predicción de desempeños en Ciencias Naturales
Fuente: Elaboración propia

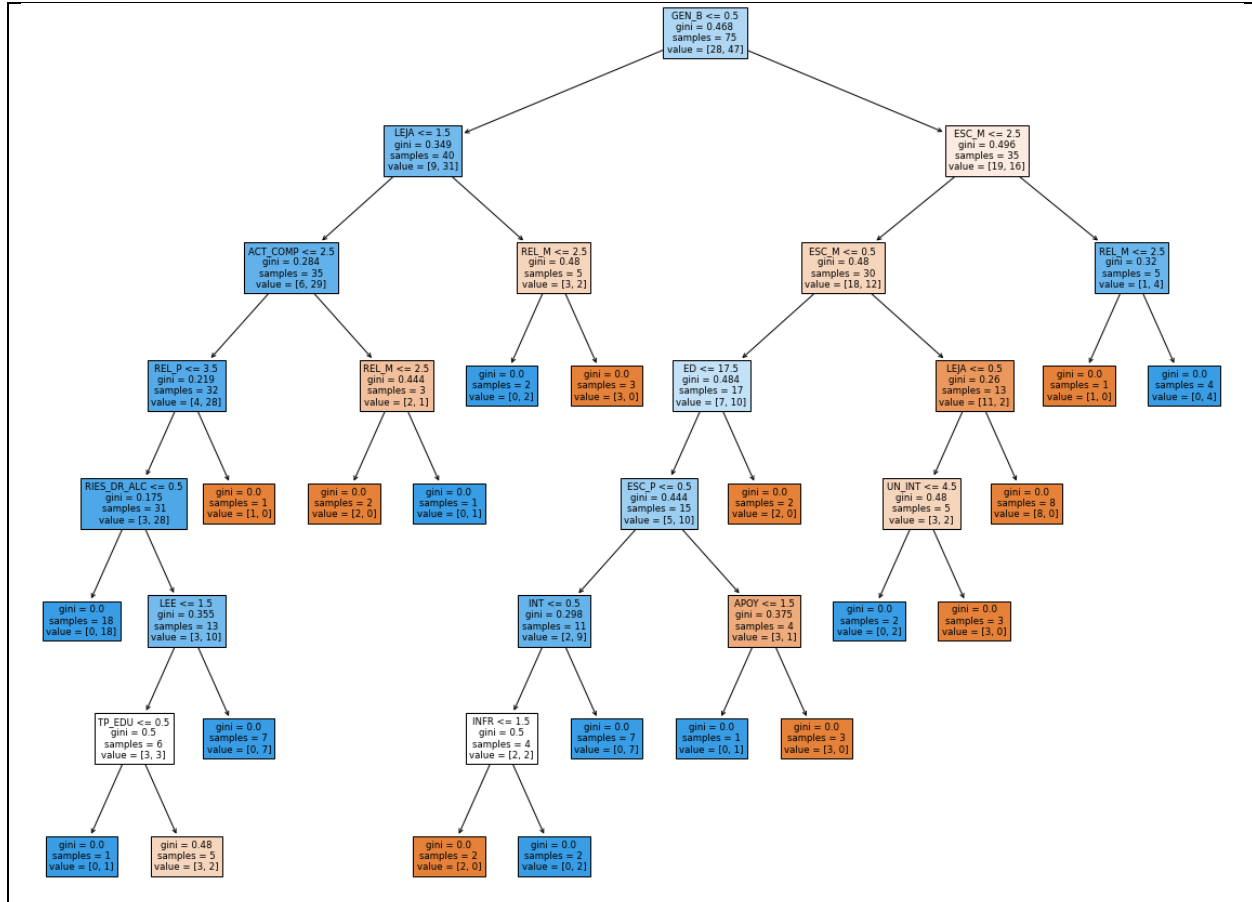


Figura 47. DT, predicción de desempeños en Ciencias Sociales
Fuente: Elaboración propia

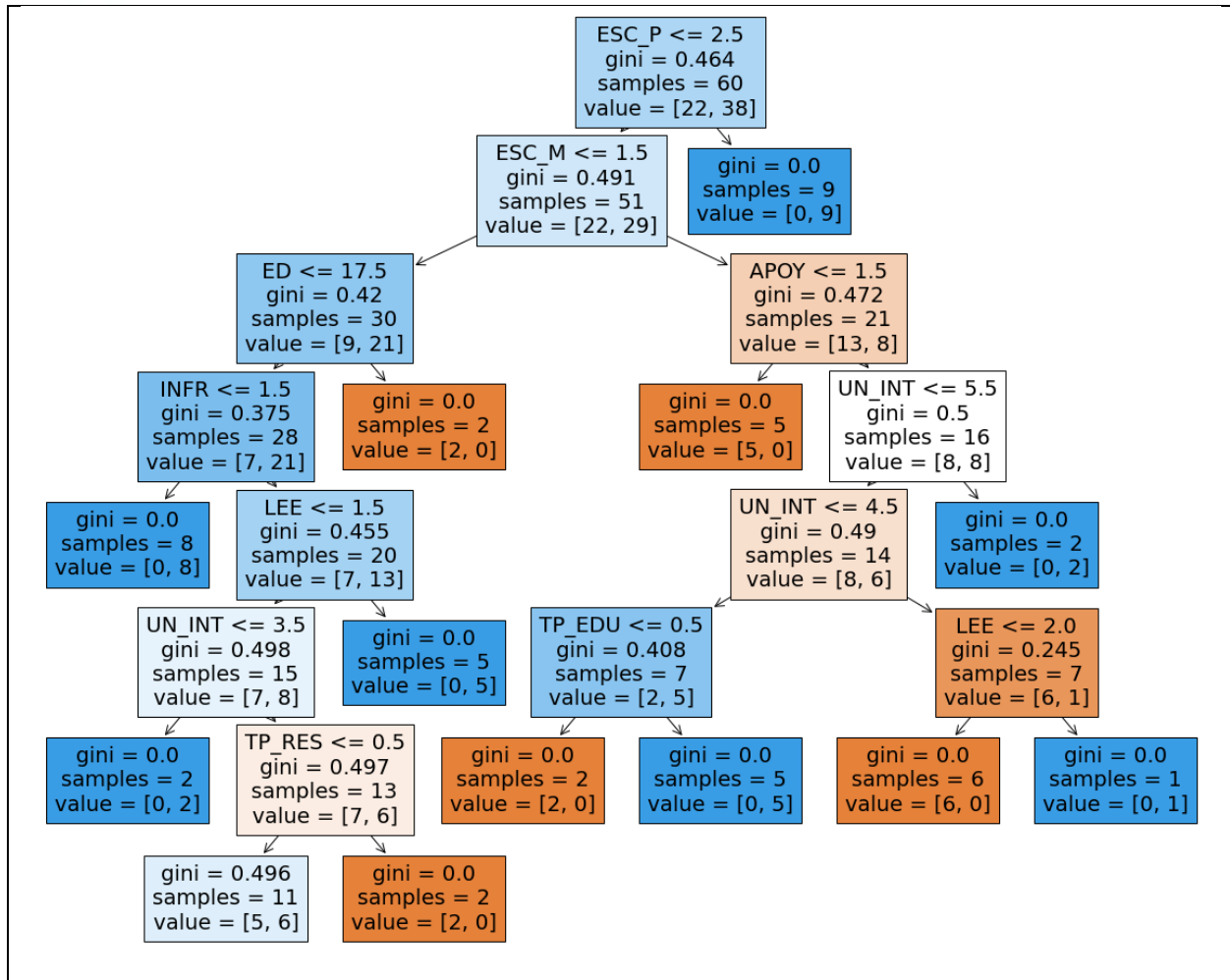


Figura 48. DT, predicción de desempeños en Lectura

Fuente: Elaboración propia

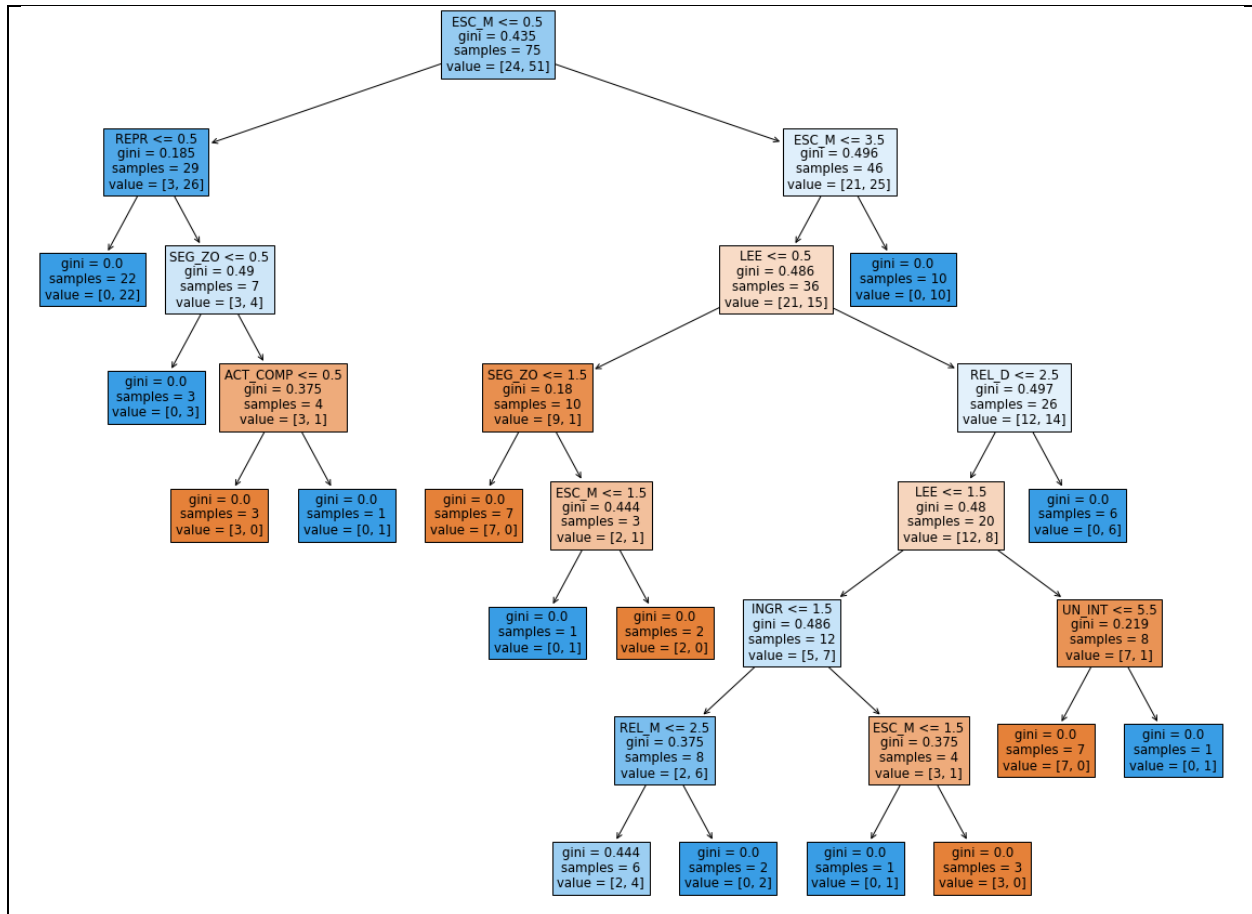


Figura 49. DT, predicción de desempeños en Inglés
Fuente: Elaboración propia

Ahora, en la tabla 17, se resume la jerarquía en la relevancia (import.) de todas las características sociodemográficas (predict.) de los estudiantes que operaron como predictores en cada uno de los árboles de decisión generados. Se recuerda que, dentro de la base de datos trabajados, se estipularon 23 predictores de los cuales en la tabla 17 se cuantifican los más relevantes en el marco del diagnóstico hecho por el algoritmo en cada una de las salidas.

Tabla 17. Importancia de las características sociodemográficas para cada predicción

Matemáticas		Lectura		Ciencias Nat.		Ciencias Soc.		Inglés	
Predic.	Import.	Predic.	Import.	Predic.	Import.	Predic.	Import.	Predic.	Import.
RIES_DR_ALC	0.203987	ESC_M	0.173797	ACT_COM_P	0.166066	REL_M	0.163132	ESC_M	0.420174
ED	0.180532	ESC_P	0.151308	REL_D	0.151774	ESC_M	0.126984	LEE	0.159928
REL_C	0.142265	ED	0.140505	INFR	0.145457	GEN_B	0.115372	REL_D	0.110868
APOY	0.09382	NU_INT	0.138236	REL_P	0.127392	LEJA	0.079275	SEG_ZO	0.079912
ESC_M	0.092956	APOY	0.108134	REL_M	0.122238	NU_INT	0.073409	REPR	0.065082
TP_EDU	0.081355	RIES_DR_ALC	0.080834	GEN_B	0.082966	INFR	0.061175	NU_INT	0.058385
REL_P	0.066726	INFR	0.078498	SEG_ZO	0.056487	ESC_P	0.05793	ACT_COM_P	0.050044
LEJA	0.056398	LEJA	0.070169	REL_C	0.051601	LEE	0.04941	INGR	0.044484
GEN_B	0.043403	ACT_COM_P	0.058518	REPR	0.042809	ACT_COM_P	0.049231	REL_M	0.011121
LEE	0.038558			LEJA	0.038048	REL_P	0.048348		
				RIES_DR_ALC	0.015162	ED	0.04798		
						APOY	0.045881		
						INT	0.038929		
						RIES_DR_ALC	0.024591		
						TP_EDU	0.018352		

Fuente: Elaboración propia

De manera posterior, conforme se estableció el modelo a usar y se establecieron los análisis anteriores; se empezó a trabajar en el tratamiento de los datos inherentes solo a las variables objetivo. Es decir, se consideró relevante no solo establecer relaciones entre aspectos sociodemográficos y los desempeños en competencias, sino que, además, se pudo señalar que entre las mismas variables objetivo (desempeños) hay interrelaciones. En ese sentido, a continuación, se resumen los hallazgos.

La figura 50 muestra las relaciones entre las cinco áreas cuando la variable objetivo es Matemáticas. Se evidencia que el nodo LEC, correspondiente a Lectura, es el más trascendente y eso se puede constatar en la tabla 18 que resume las jerarquías relacionales en las 5 salidas cuando las restantes son las entradas.

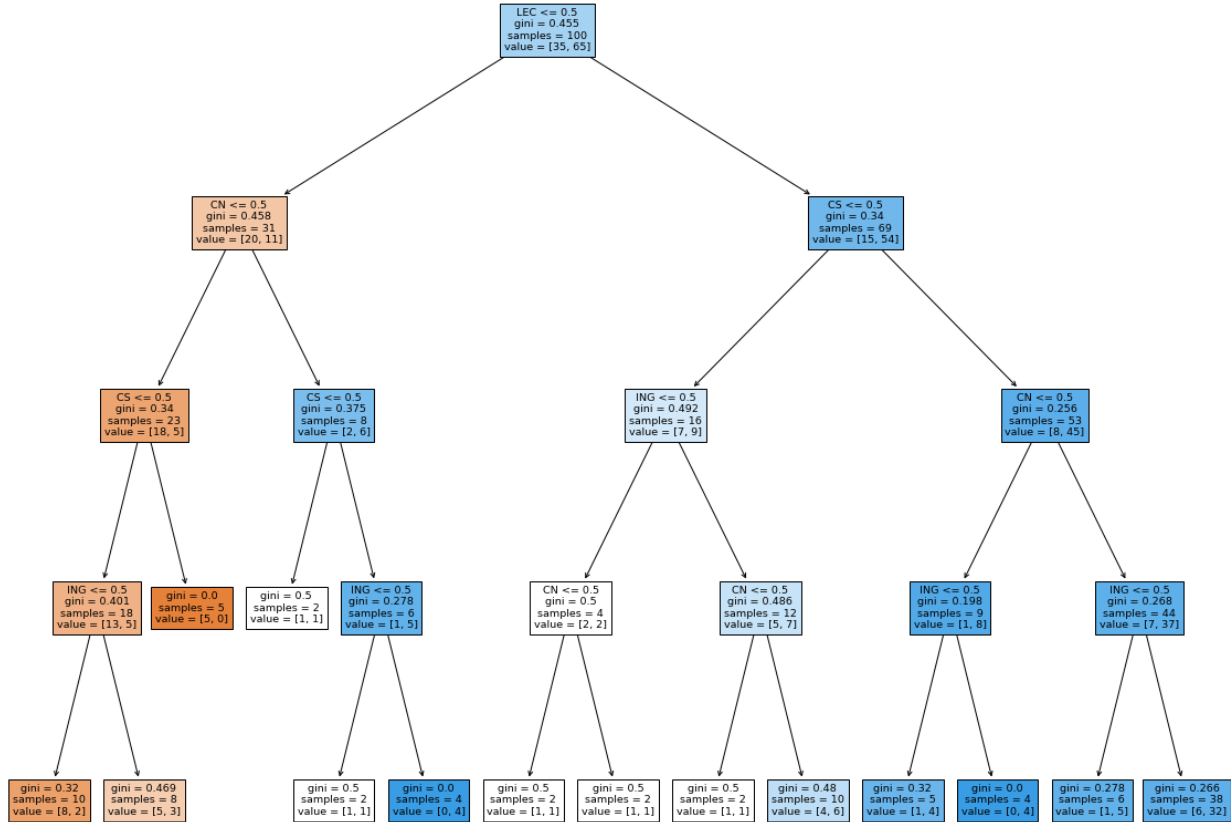


Figura 50. Relación entre desempeños-Matemáticas
Fuente: Elaboración propia

La tabla 18, no solo muestra el predictor (Predic.) más trascendente o importante (Import.) en relación con la salida respectiva. Adicionalmente, muestra en orden descendente, la jerarquía para la influencia de los otros cuatro predictores respecto a la variable objetivo que se toma.

Tabla 18. Importancia de los predictores para cada una de las salidas

Matemáticas		Lectura		Ciencias Nat.		Ciencias Soc.		Inglés	
Predic.	Import.	Predic.	Import.	Predic.	Import.	Predic.	Import.	Predic.	Import.
LEC	0.5090	CN	0.5579	LEC	0.5591	LEC	0.4651	CN	0.7406
CN	0.2239	MAT	0.2421	CS	0.1694	MAT	0.2408	MAT	0.1574
CS	0.1921	CS	0.1886	ING	0.1446	CN	0.2352	CS	0.0720
ING	0.0748	ING	0.0112	MAT	0.1267	ING	0.0587	LEC	0.0720

Fuente: Elaboración propia

Con la información de la tabla se pueden inferir relaciones interesantes que no serían posibles desde el uso de herramientas convencionales o desde métodos lineales. Se enmarca, en este sentido, la complejidad inherente a las interrelaciones dadas y la necesidad de usar nuevas estrategias –o por lo menos diferentes– para el abordaje de

problemas en los que, a simple vista es imposible destacar relaciones que podrían mejorar la comprensión de las situaciones estudiadas.

La figura 51 muestra las relaciones entre las cinco áreas cuándo la variable objetivo es Lectura. Se evidencia que el nodo CN, correspondiente a Ciencias Naturales, es el más trascendente y eso se puede constatar en la tabla 18 que resume las jerarquías relacionales en las 5 salidas cuando las restantes son las entradas.

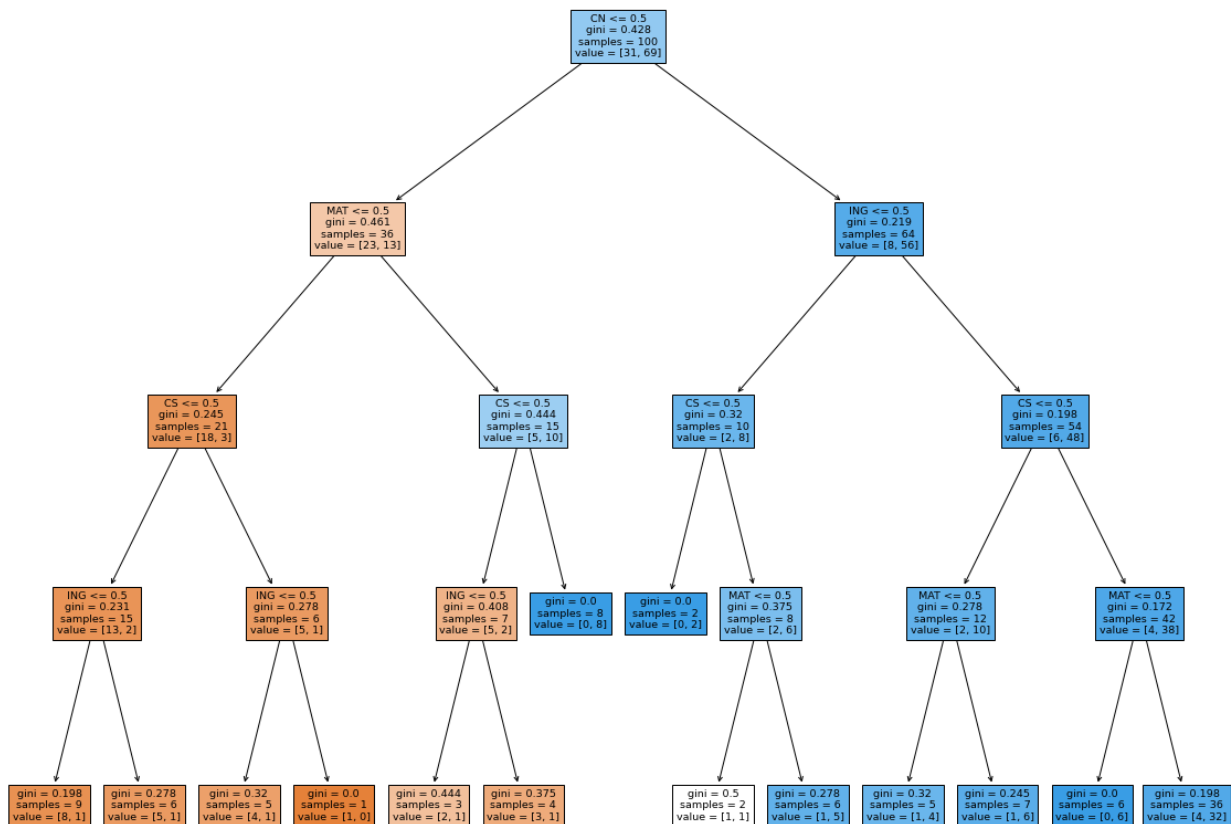


Figura 51. Relación entre desempeños-Lectura
Fuente: Elaboración propia

La figura 43 muestra las relaciones entre las cinco áreas cuándo la variable objetivo es Ciencias Naturales. Se evidencia que el nodo LEC, correspondiente a Lectura, es el más trascendente y eso se puede constatar en la tabla 17 que resume las jerarquías relacionales en las 5 salidas cuando las restantes son las entradas

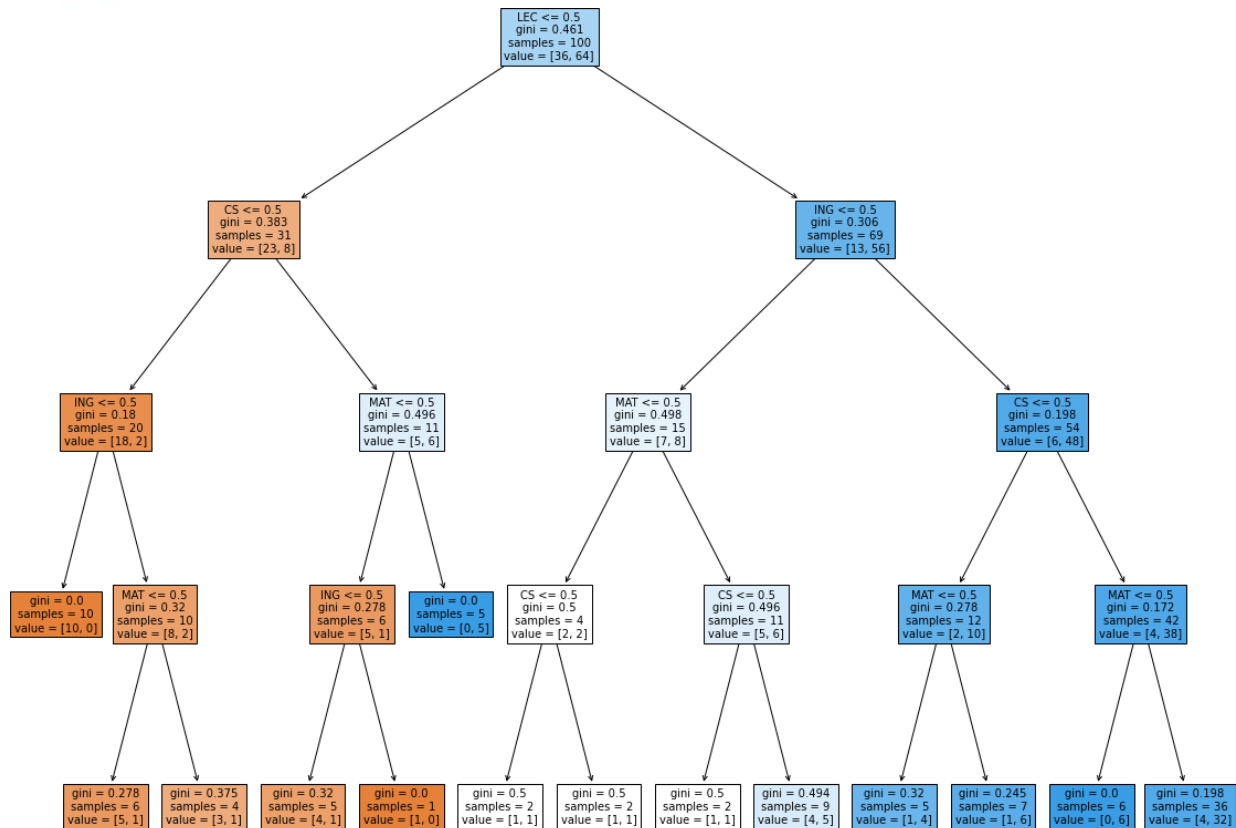


Figura 52. Relación entre desempeños-Ciencias Naturales
 Fuente: Elaboración propia

La figura 44 muestra las relaciones entre las cinco áreas cuándo la variable objetivo es Ciencias Sociales. Se evidencia que el nodo LEC, correspondiente a Lectura, es el más trascendente y eso se puede constatar en la tabla 17 que resume las jerarquías relacionales en las 5 salidas cuando las restantes son las entradas

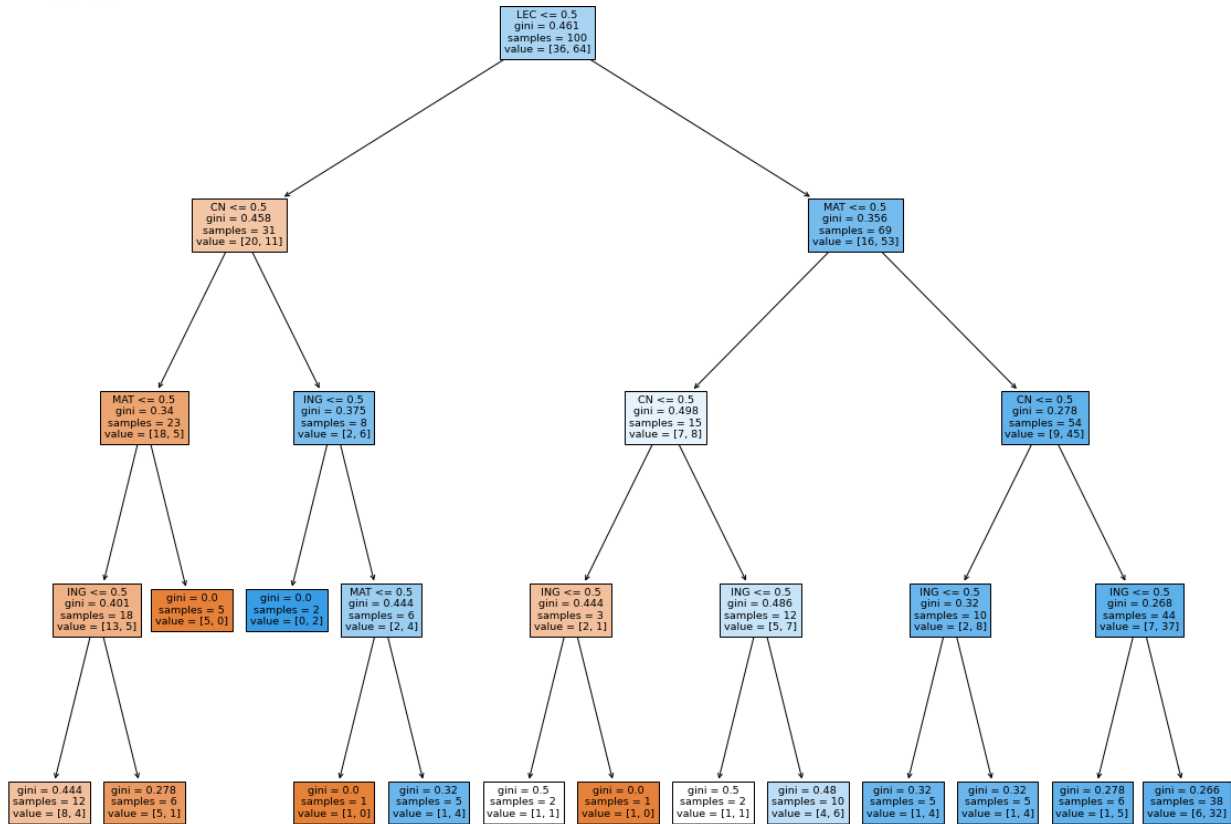


Figura 53. Relación entre desempeños-Ciencias Sociales
 Fuente: Elaboración propia

La figura 45 muestra las relaciones entre las cinco áreas cuándo la variable objetivo es Inglés. Se evidencia que el nodo CN, correspondiente a Ciencias Naturales, es el más trascendente y eso se puede constatar en la tabla 17 que resume las jerarquías relacionales en las 5 salidas cuando las restantes son las entradas

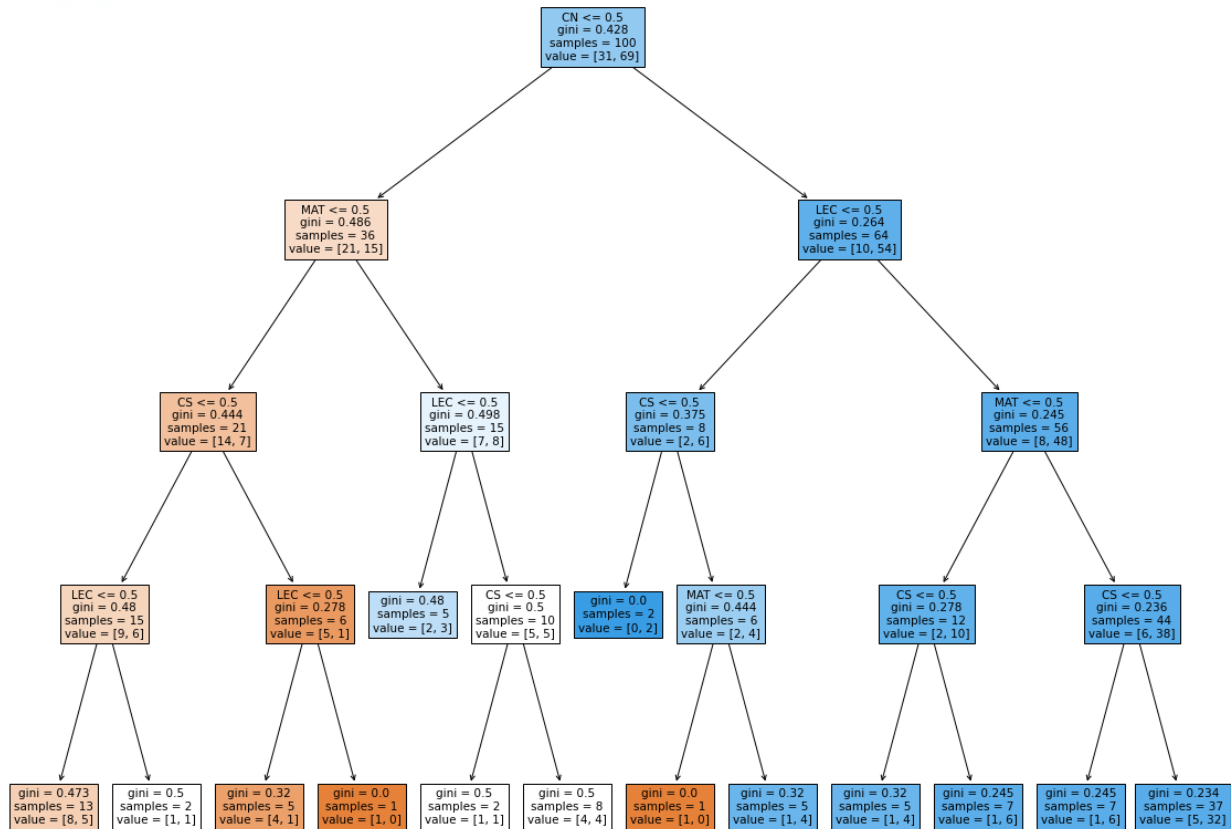


Figura 54. Relación entre desempeños-Inglés
Fuente: Elaboración propia

Fase 7. Integración del modelo.

La aplicación se desarrolló en base a la arquitectura de software llamada “modelo vista controlador”, la cual separa el código de desarrollo en tres fases que permiten diferenciar las capas de Front-End (lo que ve el usuario) y el Back-End (programación lógica de la aplicación, no interesa al usuario) de esta manera se facilita la implementación del código y además el poder administrar y hacer cambios sin afectar partes del código.

La figura 55 muestra la subdivisión interna por carpetas que constituyen el aplicativo en el que se vinculó el algoritmo que funcionó como herramienta de predicción.

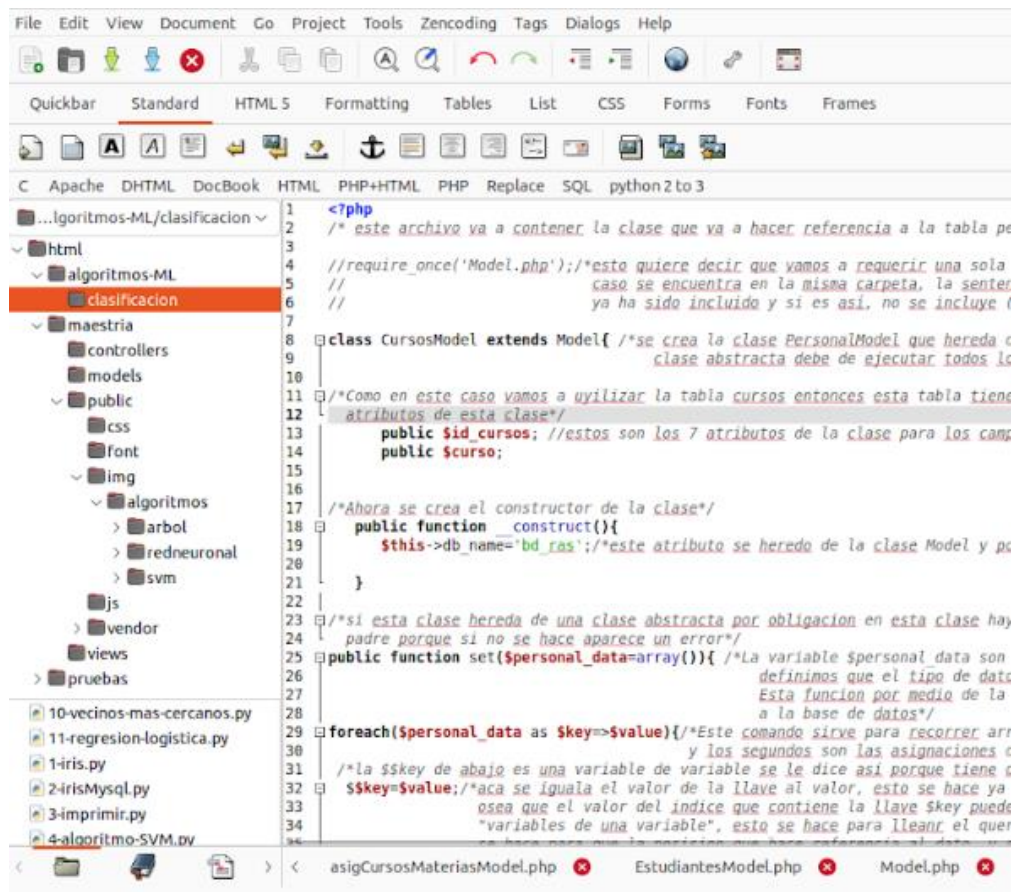


Figura 55. Subdivisión de carpetas del aplicativo
Fuente: Elaboración propia

- * La carpeta models (Modelo) contiene todos los archivos que va a interactuar con la base de datos y en ellos se desarrolla la lógica y las operaciones a trabajar con dicha información.
- * La carpeta views (Vista) contiene los archivos de interfaz con el usuario, en otras palabras, lo que se le va mostrar al usuario y con lo cual él va a interactuar.
- * La carpeta controllers (controlador) contiene los archivos que hacen de intermediarios entre los archivos de la vista y los archivos del modelo, ósea que gestiona las solicitudes de los usuarios y la respuesta del modelo al procesar y adaptar los datos a dichos requerimientos
- * La carpeta public (publico) contiene otros archivos necesarios para funcionamiento de la aplicación como son imágenes, archivos de código JavaScript para asignar dinamismo

a la página o archivos de hojas de estilos que se encarga de generar los formatos de diseño (tamaño de letra, tipo de letra, color, fondo, etc..) del sistema

Conforme se eligió el modelo DecisionTreeClassifier, se incorporó este al aplicativo para poder dar uso a las relaciones que se pudieron establecer entre aspectos sociodemográficos y los desempeños académicos. En la figura 56 se visualiza una ruta de diagnóstico realizado a dos estudiantes (líneas azul y roja) nuevos que puede ser fácilmente interpretado por cualquier ente académico de las instituciones en cuestión. En este caso, se precisaron dos cosas imprescindibles: 1. Mostrar un árbol de decisión menos técnico y más amigable para el seguimiento de las ramificaciones que conlleva un diagnóstico que debería poder hacer cualquier usuario; 2. Elegir una salida a diagnosticar, en este caso, se hizo el seguimiento del estudiante para considerar si podría o no, tener desempeños bajos en matemáticas.

Conforme a lo anterior, a continuación, en la figura 56 se relaciona de manera más sencilla la información de la figura 45 (predicción de desempeños en matemáticas) y los predictores enmarcados en la tabla 13 (variables de entrada, convenciones y sub categorías).

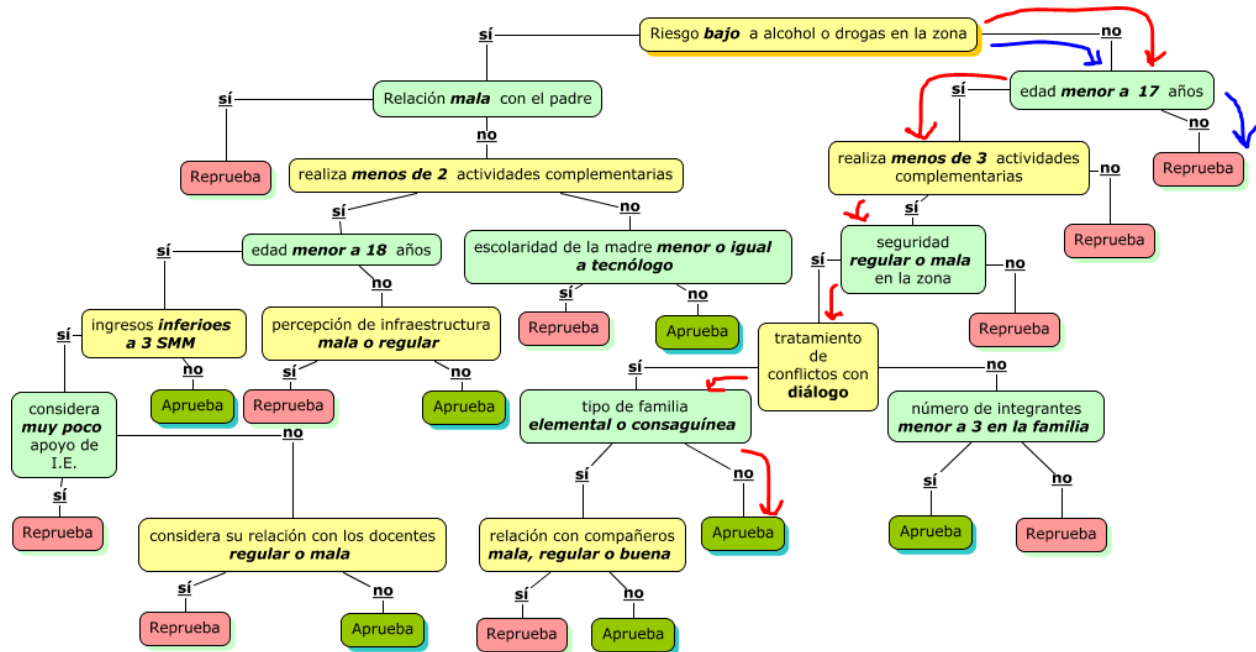


Figura 56. Ruta de predicción de dificultades de estudiantes en matemáticas

Fuente: Elaboración propia

En la anterior figura se establece una ruta de análisis, diagnóstico y de posible predicción para dos estudiantes, uno de color azul que debido a dos de los nodos del árbol se podría inferir que presentaría dificultades en el área de matemáticas y otro de color rojo que, de acuerdo a la ruta por seis nodos, podría considerarse como posible aprobador del área en cuestión. De manera similar, la ruta se puede estipular para cada una de las demás variables de salida que se han expuesto en las gráficas de la 45 a la 49. No obstante, es de aclarar que, el algoritmo es susceptible a fallos en las predicciones por razones ya señaladas y que vale la pena sintetizar en este apartado del trabajo. A saber, los fallos pueden darse debido a una mala elección o estructuración del algoritmo a aplicar; a la insuficiencia de datos o desbalanceo y ruido en los mismos; a la necesidad de actualización de la base de datos y a la presentación de datos atípicos o muy diferenciados respecto a los datos que se usan como entrenamiento del algoritmo.

Adicionalmente, se realizó un cuestionario sobre la percepción de los docentes y administrativos respecto del entorno mostrado, para considerar opciones de mejora y sobre todo para valorar su utilidad.

Fase 8. Estrategia.

En primera instancia, el propósito es brindar una aplicación que permita a los docentes diagnosticar posibles dificultades de los estudiantes dentro del desarrollo de competencias básicas a través de la interrelación entre variables sociodemográficas y de desempeños académicos. Con tal herramienta, no solo se trató de orientar prácticas docentes para fortalecer desempeños, sino que, además, se buscó sentar un referente para crear una cultura de cambio que favorezca en sentido global, la evolución del sistema educativo.

El aplicativo fue dispuesto para las instituciones objeto de estudio y posteriormente, se abre para su uso en el departamento, con el objetivo de favorecer la orientación diferenciada de la educación.

A continuación, en la tabla 19, se resalta un plan de acción en el que se establecen las metas y estrategias que se promovieron al interior de las instituciones educativas como resultado de las relaciones inherentes al tratamiento de los datos, bajo la aplicación de

los diferentes algoritmos probados y del uso del ya seleccionado: DecisionTreeClassifier (árbol de clasificación). El objetivo que se buscó fue presentar una estrategia que permitiera reorientar prácticas pedagógicas que pudieran ayudar a mejorar falencias en el desarrollo de competencias básicas.

Tabla 19. Estrategia de intervención en las instituciones educativas

Meta	Estrategia	Acciones	Indicador	Herramienta	Actores
Potenciar la comprensión de las relaciones entre competencias básicas y aspectos sociodemográficos por medio del uso de herramientas ML.	Socialización de las interrelaciones encontradas entre variables sociodemográficas y de desempeño. Muestra de la importancia de un aplicativo Web con funciones descriptivas y predictivas dentro del sistema educativo	<ul style="list-style-type: none"> * Resumir los hallazgos en tabla y diagrama de síntesis. * Presentación de la importancia de los predictores en cada una de las áreas * Ejemplo de análisis de información descriptivamente. * Ejemplo de análisis de información dentro de situaciones complejas como lo son las relaciones entre variables sociodemográficas * Muestra del uso de aplicativo web <p>Conclusiones</p>	Número de respuestas positivas / total de respuestas en el ítem 1 y 2 de la encuesta de pertinencia del trabajo	<ul style="list-style-type: none"> * Medios audiovisuales (proyector, sonido, diapositivas) Medios usados desde ML (Plataforma Anaconda, Spyder, Python) * Uso de Excel para análisis descriptivos y lineales. Uso de herramientas ML para el análisis de datos con múltiples variables. * Uso de servidor web. 	<ul style="list-style-type: none"> *Docentes Investigadores *Docentes de aula *Psico-orientadora *Directivos *Administrativos.
Fomentar la colaboración entre docentes para el desarrollo de competencias básicas en estudiantes	Muestra de las posibles interrelaciones que se proyectan desde el uso del algoritmo ML	<ul style="list-style-type: none"> * Generación de equipos de trabajo interdisciplinarios * Orientación de cómo trabajar el ABP desde las relaciones entre las competencias 	Número de respuestas positivas / total de respuestas en el ítem 4 de la encuesta de pertinencia del trabajo	<ul style="list-style-type: none"> * Debate. * Mesas de trabajo. 	<ul style="list-style-type: none"> *Docentes Investigadores *Docentes de aula *Psico-orientadora *Directivos *Administrativos.

Generar una estrategia pedagógica basada en la singularidad de las instituciones con el apoyo del aplicativo web	Implementación del algoritmo para el diagnóstico de desempeños	* Trabajo interdisciplinar entre docentes	Reflexión sobre el ítem 6 de la encuesta de pertinencia y número de guías generadas de 5 propuestas	* Guías de aprendizaje	*Docentes Investigadores *Docentes de aula
Validar el trabajo investigativo	Promover la valoración del trabajo presentado	* Realización de una encuesta de satisfacción sobre la propuesta socializada	Gráficas resumen sobre las respuestas positivas en la encuesta	Google Forms Aplicativo Web	*Docentes Investigadores *Docentes de aula

Fuente: Elaboración propia

Con base al plan de acción mostrado anteriormente, de manera inicial, se logró compartir los hallazgos de la investigación con cada una de las áreas de las dos instituciones que fueron objeto de estudio (Ver figura 57). En ese sentido, no solo se socializó lo que, en el marco del trabajo investigativo se determinó, como es el caso de las interrelaciones encontradas entre los aspectos sociodemográficos y los desempeños de los estudiantes de la muestra y de las posibles interrelaciones entre las cinco áreas correspondientes a las competencias básicas que desde el MEN se establecen.



VIGILADA MINECUCACIÓN



Figura 57. Socialización de la investigación con docentes de las Instituciones.
Fuente: Elaboración propia

También se mostró el uso general del aplicativo creado y, en una segunda fase del plan de acción, se generaron unas propuestas de problemas que, con base a las interrelaciones de desempeños compartidas, orientaron mejor la asimilación entre las comunidades académicas de ambos colegios para la cooperación mutua en pro de mejores prácticas educativas.

A continuación, se resumen algunas de las opiniones que se adjuntan a la par de las gráficas relacionadas con las demás evidencias que arrojó la socialización del trabajo investigativo.

Frente a la pregunta: ¿Qué estrategias cree que se podrían usar para mejorar las prácticas pedagógicas en la institución, de cara a lo presentado en la investigación? En la figura 58 se resumen algunas de las respuestas de los docentes.

Estrategias interdisciplinarias. ABP. Investigaciones multidisciplinarias. Proyectos transversales

Utilizar herramientas y recursos tecnológicos

Mejorar la evaluación

Estrategias que permita a cada estudiante desarrollar sus competencias a cuerdo a sus necesidades teniendo en cuenta su contexto social

Que sea socializado a todos los docentes

Trabajo interdisciplinar

Aplicar transversalidad con aprendizaje basado en problemas.

Realización de actividades que generen el gusto por el aprendizaje para una mejor comprensión de la asignatura o área

Figura 58. *Opiniones sobre mejorar las prácticas pedagógicas*

Fuente: Elaboración propia

Con relación a la pertinencia de la propuesta y los resultados socializados, a continuación, se resumen las respuestas en cada pregunta del cuestionario de validación que se aplicó.

En la pregunta: Considera que la idea central del proyecto es. En la figura 59 se resumen las respuestas

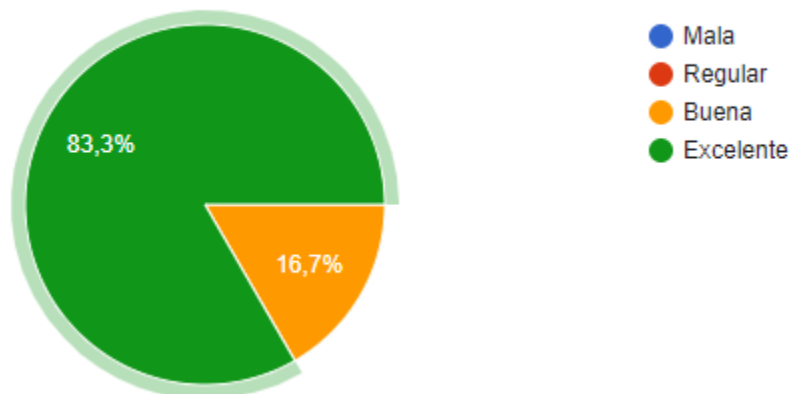


Figura 59. *Opinión, idea central del proyecto investigativo*

Fuente: Elaboración propia

En la pregunta: ¿Piensa que es confiable el estudio de la singularidad de estudiantes para prever desempeños académicos? la figura 60 resume las respuestas.

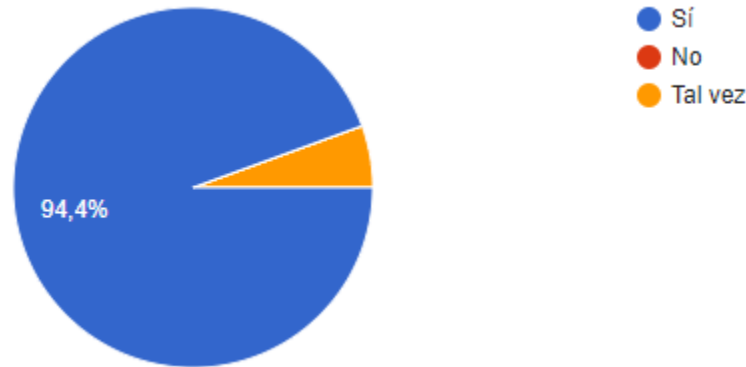


Figura 60. *Percepción de la fiabilidad del estudio realizado*

Fuente: Elaboración propia

En la pregunta: ¿Cree que los desempeños bajos en su área de enseñanza (o de formación) pueden tener relación con la formación en otras áreas? La figura 61 resume las respuestas.

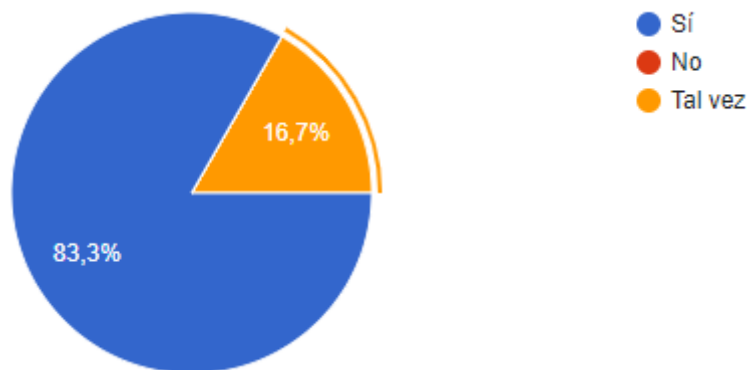


Figura 61. *Consideración sobre interrelación entre los desempeños de áreas*

Fuente: Elaboración propia

En la pregunta: ¿Estaría dispuesto (a) a reorientar sus prácticas pedagógicas (en pro de la mejora educativa), de acuerdo a los hallazgos en la investigación presentada? La figura 62 resume las respuestas

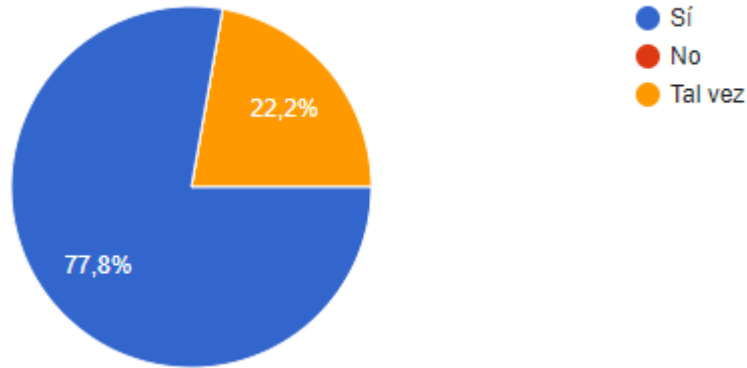


Figura 62. Disposición para reorientar prácticas pedagógicas
Fuente: Elaboración propia

Por último, en la pregunta: En una escala de 1 a 10 (donde 10 es la mejor valoración y 1 la peor), qué valor darías a la utilidad del aplicativo presentado. La figura 63 resume las respuestas.

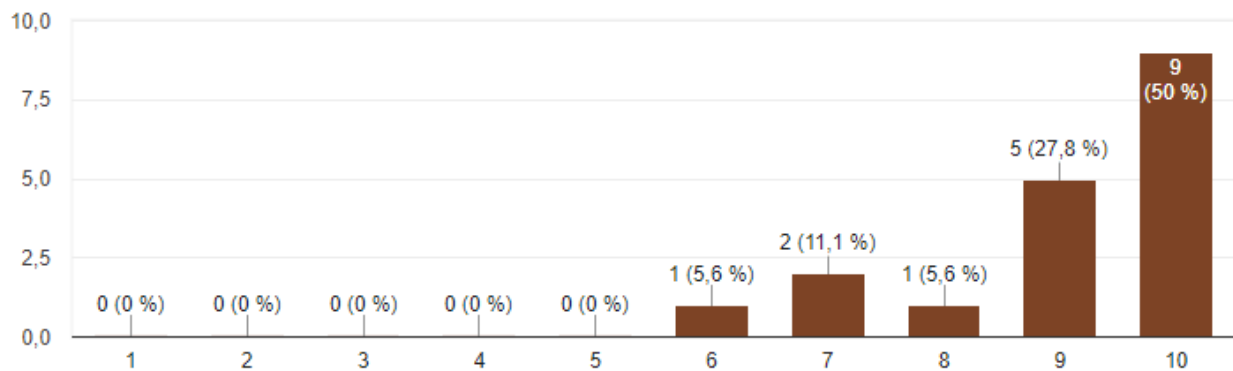


Figura 63. Valor de utilidad dado a la aplicación
Fuente: Elaboración propia

La figura 62, muestra entonces una amplia aceptación por parte de los docentes de las dos instituciones respecto a la utilidad del aplicativo presentado. Se evidencia que más de la mitad de los docentes dan calificaciones muy altas a dicha herramienta y que solo unos pocos consideran que la utilidad está en nivel medio.

Conforme a la observación de los resultados anteriores, se pueden resaltar algunos aspectos que favorecen la pertinencia de la investigación y que, posibilitan el impacto que pueden llegar a tener los hallazgos encontrados. En consecuencia, de acuerdo a los

anteriores resultados y retomando los indicadores que se disponen en el plan de acción (ver tabla 19) se destaca que:

* Se mostraron los beneficios de la implementación de herramientas computacionales – en este caso, algoritmos ML– para el tratamiento de datos que abundan en las diferentes instituciones educativas y que pudieron relacionarse con las competencias básicas de los estudiantes objeto de estudio.

* Los ítems 1 y 2 de acuerdo al indicador expuesto en la tabla 19, refleja el cumplimiento de la meta 1 al generarse una calificación de excelente por parte del 87,5% de los encuestados respecto a la idea central del proyecto y un 91,7% de respuestas positivas con relación a si se considera fiable o no las relaciones arrojadas entre aspectos sociodemográficos y de desempeños académicos.

* Por su parte, el que un 87,5% hayan respondido de manera positiva al ítem 3, muestra el espacio reflexivo generado alrededor de si consideran la existencia o no de interrelaciones entre áreas de formación por competencias. Con ello, no sólo se dio un primer paso a la posibilidad de cooperación académica, también se establece la introspección acerca de: ¿no sería benéfico re orientar mis propias prácticas pedagógicas?

* El cumplimiento de la segunda meta surge al notar en las respuestas del ítem 4, que el 83,3% de los docentes encuestados tienen disposición de re orientar sus prácticas pedagógicas como consecuencia de la reflexión hecha tras conocer los hallazgos de la investigación. En adición, ningún docente considera la negación a tal posibilidad y solo un 16,7% que respondieron “tal vez”, podrían presentar resistencia al cambio sugerido.

* Por su parte, el alcance de la meta 3 se da al promover la generación de posibles estrategias para mejorar en la academia de cada institución sin desconocer las relaciones de las singularidades de cada estudiante y de las relaciones encontradas entre los desempeños per se. En ese sentido, es de anotar que la palabra interdisciplinariedad, trabajo investigativo, mejorar la lectura, trabajo en proyectos, uso de la tecnología, fomento de la creatividad, enseñanza diferenciada, transversalidad, entre otras, surgieron como resultado de la socialización de la investigación.

Por último, se resume que la mayoría de los profesores consideran es importante la idea central del trabajo presentado, también que el estudio de la singularidad de los estudiantes, es fiable y que, se puede convertir el estudio presentado, en un fundamento para mejorar prácticas pedagógicas que conlleven el desarrollo de competencias básicas mediante la integración de docentes para la generación de ideas que se puedan materializar en guías o secuencias didácticas interdisciplinarias. De ahí que, en una fase posterior, se logró el trabajo en equipo e interdisciplinar a través de la generación de la reflexión sobre el trabajo realizado.

En la figura 64 se evidencia un espacio de debate, discernimiento y uso de los hallazgos de la investigación para, mediante la cooperación entre las cinco áreas relacionadas con las competencias básicas, crear propuestas que en sí mismas, ya son consecuencia del anhelo de mejorar prácticas pedagógicas en las dos instituciones que fueron objeto de estudio.





Figura 64. Trabajo interdisciplinar en I.E. La Merced y Ramón Alvarado
Fuente: elaboración propia

Por su parte, la figura 65 muestra un producto del trabajo interdisciplinar realizado y cimienta el primer avance en la perspectiva del presente trabajo que es: promover la mejora en las prácticas pedagógicas para fomentar el desarrollo de competencias básicas.

Mto ¹		Paso ²	DESCRIPCIÓN DEL PASO DENTRO DEL RESPECTIVO MOMENTO DE LA CLASE	COMP
INICIO	1	Formulación y análisis del problema ¿Cómo incide el nivel de escolaridad de los padres de familia o adultos en el desempeño académico de estudiantes (sexto y undécimo)? - Se selecciona esta temática de acuerdo al análisis que se hace de la tabla N° 17. Importancia de aspectos sociodemográficos.		Comunicación
	2	Lluvia de ideas y delimitación del problema ✓ Envece para determinación de escolaridad. ✓ Análisis comparativo entre desempeños académicos previos vs escolaridad / Observar y describir que aspectos muestran una tendencia que incide en el desempeño escolar.		Comunicación Argumentar
DESARROLLO	3	Investigación Información conocida y desconocida del problema + Desempeños académicos previos. + En algunos estudiantes se debe conocer ciertas características sociodemográficas. + Se desconoce la incidencia si hay relación entre las variables que se van a tomar en cuenta.		Reservar
	4	Planteamiento de posibles soluciones ✓ Impacto de forma más precisa y concreta los apoyos realizados en escuelas de padres frente a la posibilidad de mejorar los desempeños académicos.		Comunicación razonar Argumentar Resolver Problemas
EVALUACIÓN	5	Solución y presentación de conclusiones + Se espera que, culturalmente, se genere un lenguaje que permita al estudiante proyectarse académicamente más allá de la educación básica.		Comunicación Resolución Problemas
Criterios de evaluación: * Determinar si los promedios académicos en los estudiantes tienden a mejorar posterior a la aplicación de las estrategias * Aplicar un test de proyección o perspectiva del futuro para los hijos.				
Recursos: Computadores, Internet, cuestionarios, Salas de clases...				

Figura 65. Trabajo interdisciplinar pro - desarrollo de competencias básicas

Fuente: elaboración propia

El formato empleado (Ver anexo 7) ha sido modificado de Hoyos (2021) y resume en cinco pasos la aplicación del aprendizaje basado en problemas, además, incorpora elementos necesarios para la estructuración de una o varias clases bien orientadas conforme se esclarece el propósito general del problema que se abordaría.

En la figura 65, justamente se plantea por parte de cinco docentes de la Institución Educativa la Merced, un posible problema que se genera observando las relaciones entre

los aspectos sociodemográficos y de desempeño, resumidos en la tabla 17 y a los que todos los docentes de la institución educativa, tuvieron acceso. De este modo, pudieron observar que el predictor: escolaridad de la madre (ESC_M) aparecía en varias de las columnas de las competencias tomadas en el trabajo, por lo que decidieron abrir un espacio para ahondar en tal relación.

Los docentes consideran que, desde el estudio de las escolaridades de los padres, se puede pensar en reajustar las escuelas de padres para que las temáticas tratadas tengan que ver con un cambio en la proyección académica que los padres podrían generar en sus hijos. De ahí que consideren hacer un estudio interdisciplinar que destaque la importancia de la cultura y la posible incidencia del contexto familiar en los desempeños académicos de los estudiantes.

Por otro lado, en la figura 66 se deja una evidencia del planteamiento hecho por los docentes de la Institución Educativa Ramón Alvarado. En ella, se puede –similar a la I. E. La Merced– apreciar la aceptación, interés y anhelo de influir en la consecución de mejores prácticas pedagógicas conforme a lo expuesto en la socialización de la investigación. Los docentes consideraron interesante indagar sobre ¿cómo las actividades complementarias y la lectura por gusto, podrían fortalecer desempeños académicos en los estudiantes de la I. E. Ramón Alvarado? Tal cuestión surge también como derivación de la visualización de las relaciones enmarcadas en los árboles de decisión y del esquema inherente a la relevancia de los predictores.

Con base a lo anterior, la investigación permitió generar un primer espacio de convergencia entre docentes de áreas diferentes que concurren en varios puntos pero que desde este trabajo se destaca principalmente el que se haya mostrado disposición para cambiar, para trabajar con otras áreas y para considerar a los estudiantes como sujetos históricos, sociales y activos, que están supeditados a múltiples factores que afectan sus aprendizajes pero que, como docentes investigadores, podrían solventar paulatinamente conforme se promueva la mejora en las prácticas educativas.

Mto ³		Paso ⁴	DESCRIPCIÓN DEL PASO DENTRO DEL RESPECTIVO MOMENTO DE LA CLASE	COMP	
MAESTRÍA EN ESTUDIOS INTERDISCIPLINARIOS DE LA COMPLEJIDAD					
I. E. RAMON ALVARADO SANCHEZ					
PROPUESTA INTERDISCIPLINAR: DESARROLLO DE COMPETENCIAS BÁSICAS					
Ejemplo de abordaje de temáticas de diferentes áreas, para el desarrollo de competencias básicas usando ABP.					
Grado: _____		Tiempo estimado de aplicación: _____		Áreas vinculadas: CN_ CS_ MAT_ ING_ LEC_	
CONTENIDOS A ABORDAR DESDE CADA UNA DE LAS ÁREAS				CRITERIO GENERAL DE EVALUACIÓN	
C. Naturales	C. Sociales	Matemáticas	Inglés	Lec. Crítica	Análisis de mejoramiento de los desempeños académicos de los estudiantes que participan en actividades curriculares o que poseen hábitos de lectura.
Ecosistemas	Entorno y Sociedad	Juegos Mentales	Rutinas diarias	Lenguaje y Pensamiento	
Mto ³		Paso ⁴	DESCRIPCIÓN DEL PASO DENTRO DEL RESPECTIVO MOMENTO DE LA CLASE	COMP	
INICIO	1	<u>Formulación y análisis del problema</u> ¿Cómo el desarrollo de actividades complementarias y la lectura como pasatiempo, pueden fortalecer los desempeños académicos de los estudiantes de secundaria de la institución educativa Ramon Alvarado Sanchez?			Comunicativas
	2	<u>Lluvia de ideas y delimitación del problema</u> - Búsqueda de que actividades complementarias realizan los estudiantes - Tipos de géneros literarios favoritos de los estudiantes - Análisis comparativo de los desempeños académicos de los estudiantes partiendo de aquellos que leen como pasatiempo y aquellos que realizan solo actividades curriculares			Comunicativas Interpretativas
DESARROLLO	3	<u>Investigación Información conocida y desconocida del problema</u> - Se conocen los desempeños de los estudiantes. - Se deben conocer cuáles son las actividades extra clase que realizan los estudiantes y con que periodicidad la realizan. - De los estudiantes que leen como entretenimiento con que frecuencia lo hacen y cuantas lecturas realizan al mes. - No existe claridad sobre una relación intrínseca entre los desempeños académicos y las actividades extracurriculares que los estudiantes realizan.			Interpretativas Propositivas
	4	<u>Planteamiento de posibles soluciones</u> - La institución en convenio con la institución cultural Baracoa brinde espacio a todos los estudiantes para la realización de múltiples actividades (Canto, guitarra, danzas, percusión, cuerda y viento). - Dar a conocer a los estudiantes las aplicaciones de lectura (WhatsApp) en donde pueden consultar distintos géneros literarios.			Interpretativas Propositivas
EVALUACIÓN	5	<u>Solución y presentación de conclusiones</u> Se espera que los estudiantes se desarrollen de forma académica, intelectual y personal adquiriendo hábitos saludables que le permitan adaptarse a las transformaciones actuales de la sociedad.			Comunicativas Argumentativas Interpretativas
Criterios de evaluación: * Cuadro comparativo de los desempeños académicos de los estudiantes que afectan alguna actividad extracurricular y aquellos que no la realizan. * Tabla de mejora de desempeños de los estudiantes que iniciaron a participar alguna actividad extracurricular					
Recursos: computadores, internet, textos, salón, instrumentos.					

Figura 66. Trabajo interdisciplinar pro - desarrollo de competencias básicas

Fuente: elaboración propia

8. CONCLUSIONES

De manera inicial, la investigación muestra desde el análisis de antecedentes, no solo que los problemas en el ámbito educativo pueden abordarse desde perspectivas diferentes a las tradicionales, es decir, desde la observación directa y la comprobación de hipótesis que el método científico ha referenciado como la manera general de hacer ciencia; también se genera en la tabla 1, un recorrido por los trabajos que, implícita o explícitamente, afrontan distintos problemas en el campo educativo con el uso de herramientas computacionales, dentro de los cuales se resaltan el de Ramirez (2021), Florez y Pulido (2020), Contreras et al (2020) y Méndez y López (2019), cuyo fundamento es la incertidumbre y no precisamente el determinismo, el trabajo con entropía de información y el uso de una perspectiva global para el análisis de problemas antes que la segmentación de los mismos para tratar por partes lo que se comporta como sistema.

Conforme a lo anterior, el recorrido que se enmarca en la teoría expuesta en el capítulo 4 de este trabajo, permite concluir que las ciencias de la complejidad, se apoyan en herramientas tecnológicas y computacionales para, justamente considerar la solución de problemas conforme se analizan, tratan, vislumbran e indagan todas las relaciones, interconexiones, correlaciones o interdependencias posibles entre las variables que engloba el problema que se aborda. En este sentido, el problema de socavar la posibilidad de encontrar relaciones entre aspectos de tipo categóricos y cuantitativos, estriba en el estudio de la complejidad, la educación y los sistemas inteligentes.

Consecuentemente, a partir del análisis de antecedentes y de los referentes teóricos referenciados, puede inferirse que las herramientas machine learning toman relevancia en el sistema educativo debido a los beneficios que aportan para el tratamiento de información en altos volúmenes. Así, tales herramientas permiten entre otras cosas, describir, y representar distintos conglomerados de datos, pero, en adición, desde buenas prácticas de programación, aportan elementos significativos para la toma de decisiones en el ámbito académico y social de las instituciones. Lo anterior, porque usan modelos a base de probabilidades que podrían orientar mejor las acciones dentro de la búsqueda de mejoras educativas.

En este orden de ideas, las ya mencionadas herramientas computacionales (ML) brindan luz que ilumina el inextricable mundo del Big Data con lo cual se genera conocimiento que esclarece el cómo abordar fenómenos educativos complejos. En el caso de los propósitos del presente trabajo, se produjo una orientación útil para reformular prácticas pedagógicas en favor de la enseñanza por competencias junto a la valoración de la singularidad de los estudiantes, lo anterior, gracias al establecimiento de lo que podrían ser relaciones entre variables de distintas naturalezas.

Dentro de los aspectos procedimentales de la investigación, se puede concluir que: Posterior a la caracterización de la población mediante la recolección de la información; el no contar con una base de datos balanceada ni muestras suficientes de todas las clases a predecir, usando algoritmos ML, conlleva a la no fiabilidad en las métricas de rendimiento de los algoritmos. Lo anterior debido al sesgo de información que yace en los “dataset” y a los errores inherentes tras probar los algoritmos.

Las relaciones en nuestro entorno educativo, no son simples de analizar debido a la complejidad inherente a las posibles interrelaciones. Por ello, métodos tradicionales para resolver situaciones dinámicas y en las que influyen distintas variables de diversa naturaleza, no son precisamente los más adecuados.

Las métricas de incertidumbre se convierten en una herramienta imprescindible para la validación de efectividad de los algoritmos predictores. Es decir, si los algoritmos pueden generar buenos rendimientos en las predicciones encontradas, entonces, se deduce que sí podrían existir relaciones entre las variables de entrada y la de salida tomada en la investigación que se esté desarrollando. Con ello, se evita la trivialidad del conocimiento generado con relación a las interdependencias extraídas y las cuales, sin estos algoritmos y las respectivas métricas de validación, sería imposible de vislumbrar por métodos convencionales.

En consecuencia, la presente investigación manifiesta que sí pueden establecerse relaciones entre los aspectos sociodemográficos de los estudiantes objeto de estudio y sus respectivos desempeños académicos en competencias básicas. No obstante, se aclara que, cada investigación, conforme a la naturaleza del problema de estudio y de

los datos tratados, aún y cuando las métricas de rendimiento no sean las esperadas –e incluso si lo fuesen- es criterio del investigador, considerar la interpretación de los resultados que pueden arrojar desde la herramienta computacional utilizada.

Las interrelaciones no son lo único que puede inferirse mediante el uso de herramientas ML, en adición, se puede cuantificar el nivel de incidencia entre las relaciones encontradas. Así, la implementación del algoritmo DecisionTreeClassifier permitió establecer relaciones y la posibilidad de jerarquizar la importancia de las mismas respecto a cada una de las variables de salida estudiadas.

La aceptación de uso de un algoritmo u otro, en el campo de la ciencia de datos, va más allá de cuán potente puede considerarse este. Es decir, algoritmos muy estructurados y que representan un alto coste computacional, no son garantía de que son los más apropiados para realizar predicciones; esto porque, como se mencionaba en el párrafo anterior, la naturaleza de las variables –entre otros aspectos inherentes a las bases de datos– son las que condicionan el rendimiento e interpretación de los algoritmos.

Los árboles de decisión no solo conllevaron la interpretación de relaciones entre variables sociodemográficas y de desempeño, también ayudaron a discernir algunas relaciones entre las competencias académicas.

El establecimiento de relaciones entre variables sociodemográficas y de desempeño, junto a las relaciones encontradas entre los desempeños per se, permitió que los docentes reflexionaran sobre la importancia de las herramientas computacionales, la necesidad de abordar los problemas desde su complejidad y el hecho de que, los hallazgos suscitan un cambio en las prácticas pedagógicas para considerar la mejora en el desarrollo de competencias. En vista de las interrelaciones entre áreas, los docentes consideraron entre las propuestas, en pro de la mejora, el trabajo transversal e interdisciplinar a través del aprendizaje basado en problemas y proyectos.

Dentro de los aspectos que se consideran en la prospectiva de la investigación, se recalca que el análisis de relaciones entre asignaturas, identificando una de ellas como variable objetivo, se puede extender a todas las áreas del conocimiento que se orientan en las instituciones. Otro aspecto que se propone es el análisis de ocupaciones de los

egresados en relación con sus características sociodemográficas y desempeños académicos previamente recopilados, para considerar básicamente dos situaciones: cómo podría orientarse un estudiante respecto a su porvenir y si existe la posibilidad de prever las decisiones profesionales posteriores con base a un análisis similar al presentado en la presente investigación.

9. BIBLIOGRAFÍA

- Álvarez, L., Cárdenas, A., Frías, P. y Villamizar, S. (2007). Actitudes hacia la violencia social entre iguales y su relación con variables sociodemográficas en tres grupos de estudiantes de secundaria de la ciudad de Bucaramanga. *Revista Colombiana de Psicología*. No. 16. Pp. 127-133.
- Álvarez, A., & Del Río, P. (1990). Educación y desarrollo: la teoría de Vigotsky y la zona de desarrollo próximo. *Desarrollo psicológico y educación*, 2, 93-120.
https://www.researchgate.net/profile/AmeliaAlvarez/publication/348559478_Aprendizaje_y_desarrollo_La_teoría_de_la_actividad_y_la_ZDP/links/60049afa45851553a05072dc/Aprendizaje-y-desarrollo-La-teoría-de-la-actividad-y-la-ZDP.pdf
- Álvarez-Ramírez, L. Y., & Cáceres-Hernández, L. (2010). Resiliencia, rendimiento académico y variables sociodemográficas en estudiantes universitarios de Bucaramanga (Colombia). *Psicología iberoamericana*, 18(2), 37-46.
<https://www.academia.edu/download/81165111/ripsic18-2.pdf#page=39>
- Ariza-López F. J., Rodríguez-Avi, J., Alba-Fernández, V. (2018). *Control estricto de matrices de confusión por medio de distribuciones multinomiales*, GeoFocus (Artículos), nº 21, p. 215-226. ISSN: 1578-5157. <http://dx.doi.org/10.21138/GF.591>
- Barrios-Vargas C. y Cáceres-Soto M. (2017). Competencias de lenguaje y matemática según pruebas icfes vs rendimiento académico de los estudiantes de Unicolombo. *Adelante-ahead*, Vol 2, No 2. Pp. 11-37.
<http://ojs.unicolombo.edu.co/index.php/adelante-ahead/article/view/33>
- Baumrind, D. (1991). The influence of parenting style on adolescent competence and substance use. *Journal of Early Adolescence*, 11 (1), 56-95.
<https://journals.sagepub.com/doi/abs/10.1177/0272431691111004>
- Bausili Llamas, E. (2021). Machine Learning para el diagnóstico de COVID-19 [Trabajo de Grado]. Madrid, España.
<https://repositorio.comillas.edu/xmlui/handle/11531/53241>
- Beunza, J. J., & Puertas, E. (2020). Tipos de algoritmos. *Manual práctico de inteligencia artificial en entornos sanitarios*, 35.

https://books.google.com/books?hl=en&lr=&id=88nSDwAAQBAJ&oi=fnd&pg=PA35&dq=machine+learning,+tipos+de+aprendizaje&ots=6Q8dGGZ6T_&sig=wDjkify1m62oZMRtuCuJZus4kho

Bezanilla, M. (2003). El proyecto Tuning y las competencias específicas. *Seminario Internacional. Orientaciones pedagógicas para la convergencia europea de Educación Superior*. Universidad de Deusto.

Borja-Robalino, R., Monleón-Getino, A., & Rodellar, J. (2020). Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E30), 184-196.
https://www.researchgate.net/profile/Antonio-Monleon-Getino/publication/342009715_Estandarizacion_de_metricas_de_rendimiento_para_clasificadores_Machine_y_Deep_Learning/links/5ede3d0392851cf13869078e/Estandarizacion-de-metricas-de-rendimiento-para-clasificadores-Machine-y-Deep-Learning.pdf

Britos, P. (2008). *Procesos de explotación de información basados en sistemas inteligentes* (Doctoral dissertation, Universidad Nacional de La Plata).
<http://sedici.unlp.edu.ar/handle/10915/4142>

Brito-Jiménez, I. T., & Palacio-Sañudo, J. (2016). Calidad de vida, desempeño académico y variables sociodemográficas en estudiantes universitarios de Santa Marta-Colombia. *Duazary*, 13(2), 133-141.

Cadreja, M. (1990). *John Dewey: propuesta de un modelo educativo. I. Fundamentos*. *Aula Abierta*, (55), 61-88. <https://dialnet.unirioja.es/descarga/articulo/2781489.pdf>

Carrion, J. (2017). Diferencia entre dato información y conocimiento.
<http://148.202.167.116:8080/jspui/handle/123456789/869>

Carvajal, C., González, J., & Sarzoza, S. (2018). Variables Sociodemográficas y Académicas Explicativas de la Deserción de Estudiantes en la Facultad de Ciencias Naturales de la Universidad de Playa Ancha (Chile). *Formación universitaria*, 11(2), 3-12. <https://dx.doi.org/10.4067/S0718-50062018000200003>

Castaño-Villar, F. (2015). Carabaña, J.(2015). La inutilidad de PISA para las escuelas. Madrid: Catarata. *Trabajo Social Global-Global Social Work*, 5(9), 111-116.
<https://revistaseug.ugr.es/index.php/tsg/article/view/3884>

- Castro-Sánchez, M., Zurita-Ortega, F.; Chacón-Cuberos, R. (2019). Motivation towards sport based on socio-demographic variables in university students from Granada. *Journal of Sport and Health Research*. 11(1):55-68.
<https://digibug.ugr.es/handle/10481/59177>
- Castro López, R. I. (2020). Aplicación de técnicas de Machine Learning para el estudio de deserción temprana y egreso oportuno en estudiantes de Ingeniería de la Facultad de Ciencias Físicas y Matemáticas [Trabajo de Grado]. Santiago de Chile, Chile. <https://repositorio.uchile.cl/handle/2250/178598>
- Cázares, R. (2008). El enfoque por competencias en educación. *Revista Ide@s CONCYTEG*, 3(39), 53-64. https://www.researchgate.net/profile/Rocio-C-3/publication/267553115_El_enfoque_por_competencias_en_educacion/links/5495d5140cf29b944824132d/El-enfoque-por-competencias-en-educacion.pdf
- Chagoya, E. R. (2008). Métodos y técnicas de investigación. <https://www.gestiopolis.com/metodos-y-tecnicas-de-investigacion>.
- Chica, E. (2011). Una propuesta de evaluación para el trabajo en Grupo mediante rúbrica. *EA, Escuela abierta: revista de Investigación Educativa*, ISSN 1138-6908, N° 14, 2011, pp. 67-82.
- Concejero, P. (2004). Comparación de modelos de curvas ROC para la evaluación de procedimientos estadísticos de predicción en investigación de mercados. Tesis Doctoral. Universidad Complutense de Madrid. <http://concejero.wikidot.com/local--files/tesis/04-comparacion%20curvas%20ROC.pdf>
- Contreras, L. E., Fuentes, H. J., & Rodríguez, J. I. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación universitaria*, 13(5), 233-246.
https://scielo.conicyt.cl/scielo.php?pid=S071850062020000500233&script=sci_arttext&tlng=e
- Correa, M., Bielza, C., Pamies-Teixeira, J., & Alique López, J. R. (2008). Redes Bayesianas vs redes neuronales en modelos para la predicción del acabado superficial. Madrid, España. <https://digital.csic.es/handle/10261/13826>
- De-La-Hoz, E., De-La-Hoz, E., & Fontalvo, T. (2019). Metodología de aprendizaje automático para la clasificación y predicción de usuarios en ambientes virtuales de

- educación. *Información tecnológica*, 30(1), 247-254. https://scielo.conicyt.cl/scielo.-php?pid=S071807642019000100247&script=sci_arttext
- Delgado, I., Espinoza, J., & Fonseca, J. (2017). Ansiedad matemática en estudiantes universitarios de Costa Rica y su relación con el rendimiento académico y variables sociodemográficas. *Propósitos y Representaciones*, 5(1), 275-324. <http://funes.uniandes.edu.co/13497/>
- Delgado, S. (2022). Aprende Python. https://aprendepython.es/_downloads/907b5202c1466977a8d6bd3a2641453f/aprendepython.pdf
- Dewey, J. (1995). *Democracia y educación: una introducción a la filosofía de la educación*. Ediciones Morata. https://books.google.com/books?hl=en&lr=&id=s8KsHz4q7ZIC&oi=fnd&pg=PA13&dq=dewey+educaci%C3%B3n&ots=q0ZCzTqAG-&sig=-IjWhLagn7u7lyATYLnHYEuML_c
- De Tejada Lagonell, M. (2012). Variables sociodemográficas según turno escolar, en un grupo de estudiantes de educación básica: un estudio comparativo. *Revista de Pedagogía*, 33(92), 235-269. <https://www.redalyc.org/pdf/659/65926546002.pdf>
- De Ullibarri Galparsoro, L., & Pita Fernández, S. (1999). Medidas de concordancia: el índice de Kappa. *Cad Aten Primaria*, 6, 169-171. https://www.academia.edu/download/33328322/Indice_de_kappa.pdf
- Elizalde Prada, Ó. A. (2013). Aproximación a las ciencias de la complejidad. *Revista de la Universidad de La Salle*, (61), 45-66. <https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1370&context=ruls>
- Fernández Gay, C. (2013). Rendimiento escolar y contexto social en educación primaria. [Tesis de maestría]. <http://repositorio.ual.es/bitstream/handle/10835/2301/Trabajo.pdf>
- Fernández-González, N. (2016). UNESCO (2015). Replantear la educación ¿Hacia un bien común mundial? París: UNESCO. 93 pp. ISBN-978-92-3-300018.6. *Journal of Supranational Policies of Education*. https://repositorio.uam.es/bitstream/handle/10486/671277/JOSPOE_4_15.pdf?s

- Fernández, M., Alcaraz, N., y Sola, M. (2017). Evaluación y pruebas estandarizadas: una reflexión sobre el sentido, utilidad y efectos de estas pruebas en el campo educativo. *Revista Iberoamericana de Evaluación Educativa*. 10(1), 51-67. <https://doi.org/10.15366/riee2017.10.1.003>
- Finn, J. (1989). Withdrawing from school. *Review of Educational Research*, 59(2), 1171-142. <https://journals.sagepub.com/doi/abs/10.3102/00346543059002117>
- Florez, C. y Pulido, K. (2020). Estrategia Didáctica mediada por Técnicas de Machine Learning para Potenciar la Habilidad de Pensamiento Crítico Interdisciplinar en Procesos Académicos y Actitudinales. <https://repositoriotesiscomplejidad.blogspot.com/>
- Fuentes Fino, R. J. (2022). Mejora de la incertidumbre al usar datos fuera de la distribución (OOD) en un modelo semi-supervisado de aprendizaje profundo. <https://riuma.uma.es/xmlui/handle/10630/23802>
- Gamarra Gómez, F. (2020). MODELO BASADO EN MACHINE LEARNING PARA EL NEURORRENDIMIENTO ACADÉMICO DE ESTUDIANTES UNIVERSITARIOS. *REVISTA CIENCIA Y TECNOLOGÍA - Para el Desarrollo - UJCM*, 5(9), 10-18. doi:<http://dx.doi.org/10.37260/rctd.v5i9.137>
- Gatica Vásquez, J. E. (2020). Metodologías de arribos y su impacto en las métricas de un estudio de simulación sobre capacidad de almacenamiento en un centro de distribución de consumo masivo [Tesis de maestría]. <http://repositorio.udec.cl/jspui/handle/11594/476>
- Gil-Vera, V. D., & Seguro-Gallego, C. (2022). Machine learning aplicado al análisis del rendimiento de desarrollos de software. *Revista Politécnica*, 18(35), 128-139. <https://doi.org/10.33571/rpolitec.v18n35a9>
- Gómez, P., Castro, P., Mora, M. F., Pinzón, A., Torres, F., & Villegas, P. (2014). Estándares básicos de competencias. Comparación con el estudio PISA y cuestiones para su ajuste. <http://funes.uniandes.edu.co/6885/>
- Gonzalo, P. (2019). Aplicación de algoritmos machine learning para la predicción del beneficio del cliente a partir de métricas de Google Analytics. <https://core.ac.uk/download/pdf/250406428.pdf>

- González Sánchez, J., y Peñaloza, M. (2021). Identificación y predicción de estudiantes en riesgo de deserción académica por medio de modelos basados en Machine Learning. <https://repository.libertadores.edu.co/handle/11371/4188>
- González, C. (2004). Sistemas inteligentes en la educación: una revisión de las líneas de investigación y aplicaciones actuales. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 10(1), 3-22.
<https://www.redalyc.org/pdf/916/91610101.pdf>
- Guzmán Marín, F. (2017). Problemática general de la educación por competencias. *Revista Iberoamericana de Educación*.
<https://redined.educacion.gob.es/xmlui/handle/11162/174630>
- Hawking, S. (2003). *A hombros de gigantes*. Editorial Crítica. Barcelona-España.
- Hernández - Mosqueda, J., Tobón - Tobón, S., y Guerrero - Rosas, G. (2016). *Hacia una evaluación integral del desempeño: las rúbricas socioformativas*. Ra Ximhai, 12(6),359-376. <https://www.redalyc.org/articulo.oa?id=461/46148194025>
- Heuter, O. (2016). PHP7 Desarrollar un sitio web dinámico e interactivo
- Hoyos, M. (2020). *Aplicación didáctica del modelo ABP para el análisis y comprensión del concepto de límite de una función introduciendo Geogebra con estudiantes de 1º Bachillerato* [Tesis de postgrado, UNIR, España]. Archivo digital.
- Joaquín, R. (2020). Gradient Boosting con Python by, available under a Attribution 4.0 International (CC BY 4.0)
https://www.cienciadedatos.net/documentos/py09_gradient_boosting_python.html
- Jones- Ortiz, C. V. ., & Guzmán – Seraquive, J. E. . (2022). Análisis de las técnicas de machine learning aplicadas en la detección de fraudes bancarios. *Revista Científica Ciencia Y Tecnología*, 22(33). <https://doi.org/10.47189/rcct.v22i33.516>
- Loinaz, M. (2001). Sistemas inteligentes en el ámbito de la educación. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 5(12), 5-12.
<https://www.redalyc.org/pdf/925/92551202.pdf>
- Loli, A. E., Jara, E. T. D., Del Carpio, J., & Elsa La Jara, G. (2010). Actitudes de creatividad y emprendimiento en estudiantes de la Universidad Nacional de Ingeniería y su relación con algunas variables socio demográficas. *Revista de investigación en psicología*, 13(2), 139-151.

<https://dialnet.unirioja.es/servlet/articulo?codigo=3751470>

Maldonado, C. (2017). Educación compleja: Indisciplinar la sociedad. En *Revista Educación y Humanismo*, 19(33), 234-252.

<http://dx.doi.org/10.17081/eduhum.19.33.2642>

Maldonado, C. (2020). *Camino a la complejidad, Revoluciones – científicas e industriales, Investigación en complejidad*. Edit, Oscar Azmitia. 1° ed. Bogotá – Colombia.

Martinez y Palencia. (2021). Modelo de minería de datos para el análisis de la productividad y crecimiento personal en las mujeres emprendedoras: el caso de la Asociación las Rosas, *SUMA DE NEGOCIOS*, 12(26), 23-30.
<http://doi.org/10.14349/sumneg/2021.V12.N26.A3>

Mayorga, D., Melgarejo, M., & Obregon, N. (2016). A Fuzzy Clustering based method for the spatiotemporal analysis of criminal patterns. In 2016 *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 738-744). IEEE.

Maya, E. (2014). Métodos y técnicas de investigación.

<http://www.librosoa.unam.mx/handle/123456789/2418>

Méndez, A. (2007). Terminología pedagógica específica al enfoque por competencias: el concepto de competencia. *Innovación educativa*.
<https://redined.educacion.gob.es/xmlui/handle/11162/75735>

Méndez, O. y López, J. (2019). Técnicas de Machine Learning para la predicción de desempeño académico en el Desarrollo del espacio proyectivo del Pensamiento Espacial. [Tesis de maestría, Universidad Pedagógica Nacional]. Bogotá, Colombia.
<https://docs.google.com/document/d/11-pYeDpzAg3Y9wRbczjyEENjtQXo2cJH/edit#>

MEN (2006). Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas. Bogotá. https://www.mineducacion.gov.co/1621/articulos-340021_recurso_1.pdf

Monge-López, D., Bonilla, R., & Aguilar-Freyan, W. (2017). El Inventario de Estrategias de Autorregulación: traducción al español, características psicométricas preliminares y su relación con variables sociodemográficas en una muestra de estudiantes universitarios. *Avances en psicología latinoamericana*, 35(1), 61-78.

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-47242017000100005

Morin, E. (1999). *Los siete saberes necesarios para la educación del futuro*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura - 7 place de Fontenoy - 75352 París 07 SP – Francia.

<http://repositorio.minedu.gob.pe/handle/20.500.12799/1448>

Murcia Ramírez, J. (s.f). Modelo de máxima entropía para la predicción de criminalidad con técnicas de Machine Learning.[Trabajo de Grado]. Universidad de los Andes, Colombia. <http://hdl.handle.net/1992/55674>

Murga-Menoyo, M. (2015). Competencias para el desarrollo sostenible: las capacidades, actitudes y valores meta de la educación en el marco de la Agenda global post-2015. *Foro de Educación*.

<https://redined.educacion.gob.es/xmlui/handle/11162/177103>

Nieto, Y., García-Díaz, V., & Montenegro, C. (2019). Decision-making model at higher educational institutions based on machine learning. *JUCS-Journal of Universal Computer Science*, 25, 1301. <https://lib.jucs.org/article/22660/download/pdf/>

NÚÑEZ PÉREZ, María Cruz Silvia. El docente en el enfoque por competencias. *Pensamiento. Papeles de Filosofía*, [S.l.], n. 1, p. 177-186, jun. 2013. ISSN 1870-6304. Disponible en: <https://revistapensamiento.uaemex.mx/article/view/340>

Ortiz, J., Rojas, P. (2018). Diseño de un aplicativo web para el control de asistencia en lenguaje Preprocesador de Hipertexto. http://repository.ucc.edu.co/bitstream/20.500.12494/10474/1/2019_dise%C3%B1o_lenguaje_hipertexto.pdf

Ortiz Toribio, E. (2021). Incorporación de métricas difusas en el algoritmo RANSAC: implementación y análisis de comportamiento [Tesis de maestría]. <http://hdl.handle.net/11201/158307>

Patiño, R. G. (2017). La calidad, las competencias y las pruebas estandarizadas: una mirada desde los organismos internacionales. *Educación y ciudad*, (33), 159-170. <https://dialnet.unirioja.es/servlet/articulo?codigo=6232098>

Pérez, D., & González, C. (2007). *Minería de Datos. Técnicas y Herramientas*. B & N. <https://books.google.com/books?hl=en&lr=&id=wz->

- D_8uPFCEC&oi=fnd&pg=PR4&dq=Miner%C3%ADa+de+Datos.+T%C3%A9cnica
s+y+Herramientas&ots=TjW6Bk8taL&sig=3tcAZXQDTZDPKFJAxI46Jo-VOYc
- Perrenoud, P. (1999). Construir competencias desde la escuela, Dolmen Ediciones SA, Santiago de Chile.
- Perrenoud, P. (2008). Construir las competencias, ¿es darle la espalda a los saberes? *Revista de docencia universitaria*, 6(2).
<https://revistas.um.es/redu/article/view/35261>
- Peña, P., Muñoz L., Nuñez, E. y Torres, A. (2019). Modelo basado en minería de datos para el análisis de los intereses cognoscitivos y los factores psicosociales que inciden en el aprendizaje de los estudiantes. [Tesis de maestría, Universidad Surcolombiana]. Huila, Colombia. <https://repositoriotesiscomplejidad.blogspot.com/>
- Piaget, J. (1981). La teoría de Piaget. *Infancia y aprendizaje*, 4(sup2), pp. 13-54.
<https://www.tandfonline.com/doi/pdf/10.1080/02103702.1981.10821902>
- Puchol, A. S., NÚÑEZ, R. N., & Rodríguez, J. (2008). Elaboración de un cuestionario para la evaluación de competencias genéricas en estudiantes universitarios. *Apuntes de psicología*, 26(1), 35-49.
<https://redined.educacion.gob.es/xmlui/handle/11162/135003>
- Quinlan, J. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
<https://hunch.net/~coms-4771/quinlan.pdf>
- Quinlan, J. (1990). Learning logical definitions from relations. *Machine learning*, 5(3), 239-266. <https://link.springer.com/content/pdf/10.1007/BF00117105.pdf>
- Ramírez Martínez, D. E. (2021). Análisis del Machine Learning como estrategia didáctica para el mejoramiento de Competencias Básicas en Educación Secundaria en Entornos Virtuales de Aprendizaje. <https://gredos.usal.es/handle/10366/146077>
- Ramírez, L. Y. Á., Quintero, A. P. C., Aguilar, P. F., & Méndez, S. V. (2007). Actitudes hacia la violencia social entre iguales y su relación con variables sociodemográficas en tres grupos de estudiantes de secundaria de la ciudad de Bucaramanga. *Revista Colombiana de Psicología*, 16, 127-137.
<https://revistas.unal.edu.co/index.php/psicologia/article/view/1010>

- Rodríguez, A., León, S., & Guerra, Y. (2007). Competencias ciudadanas aplicadas a la educación en Colombia. *Revista educación y desarrollo social*, 1(2), 140-157. <http://revistas.unimilitar.edu.co/index.php/reds/article/view/700>
- Rojas, E. M. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E28), 586-599. <https://search.proquest.com/openview/c7e24c997199215aa26a39107dd2fe98/1?pq-origsite=gscholar&cbl=1006393>
- Rojas, L. (2013). Predicción de la dificultad de la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica. *Revista Digital: Matemática, Educación e Internet*, 13(1). <https://revistas.tec.ac.cr/index.php/matematica/article/view/1627>
- Ros, I. (2009). La implicación del estudiante con la escuela. *Revista de Psicodidáctica*, 14(1), 79-92. <https://ojs.ehu.eus/index.php/psicodidactica/article/download/252/248>
- Ruiz Morales, Y., García García, M., Biencinto López, C. M., & Carpintero Molina, M. E. (2017). Evaluación de competencias genéricas en el ámbito universitario a través de entornos virtuales: Una revisión narrativa. *Revista electrónica de investigación y evaluación educativa*. <https://redined.educacion.gob.es/xmlui/handle/11162/135003>
- San Martín R., V. (2002). La formación en competencias: el desafío de la educación superior en Iberoamérica. *Revista Iberoamericana De Educación*, 29(1), 1-8. <https://doi.org/10.35362/rie2912945>
- Sánchez-Lorite, S. (2012). La importancia de la familia en Educación Infantil. [Trabajo de fin de grado]. Universidad Internacional de la Rioja.
- Sánchez Castrillón, J. D., Vallejo, P., & Tabares Betancur, M. S. (2020). *Hacia un método de predicción de resultados de evaluación en un contexto de micro aprendizaje* (Doctoral dissertation, Universidad EAFIT). <https://repository.eafit.edu.co/handle/10784/29617>
- Sancho, G. M. (2010). Asesoría académica universitaria: Perfil de competencias básicas y genéricas. *Revista Electrónica "Actualidades Investigativas en Educación"*, 10(2), 1-16. <https://www.redalyc.org/pdf/447/44717910004.pdf>

- Schmidt, Q. (2006). Estándares básicos de competencias en lenguaje, matemáticas, ciencias y ciudadanas: guía sobre lo que los estudiantes deben saber y saber hacer con lo que aprenden [1].
https://edumediaepot.gei.de/bitstream/handle/11163/1921/788071114_2006_A.pdf?sequence=6
- Sucar, L. E., & Tonantzintla, M. (2006). Redes bayesianas. *Aprendizaje Automático: conceptos básicos y avanzados*, 77, 100. <https://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>
- Tobón, S. (2003). Las competencias en el sistema educativo: de la simplicidad a la complejidad. Bogotá: CIFE.
<http://files.nathalyeismenia.webnode.com.ve/200000134-bff07c1e3d/LAS%20COMPETENCIAS%20Y%20EL%20PENSAMIENTO%20COMPLEJO.doc>
- Torío, S.L., Peña, J.V.C. y Rodríguez, M.D.C. (2008). Estilos educativos parentales. Revisión bibliográfica y reformulación teórica, Teoría de la Educación. *Revista Interuniversitaria*, 20, 151-178. <https://gredos.usal.es/handle/10366/71805>
- Uribe, M. I. G., & Márquez, M. G. (2022). Clima social, familiar, escolar y conductas de riesgo en adolescentes. *Revista de Psicología de la Universidad Autónoma del Estado de México*, 11(23), 231-258.
<https://revistapsicologia.uaemex.mx/article/view/18057>
- Urteaga, I., Siri, L., & Garófalo, G. (2020). Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea. *RIED. Revista Iberoamericana de Educación a Distancia*, 23(2), 147-167.
<https://www.redalyc.org/jatsRepo/3314/331463171008/331463171008.pdf>
- Velásquez, S., Celis, J., & Hernández, C. (2017). Evaluación contextualizada como estrategia docente para potenciar el desarrollo de competencias matemáticas en Pruebas Saber. *Eco Matemático Journal of Mathematical Sciences*, 8(S1), 33-37.
<http://funes.uniandes.edu.co/23384/>
- Vigotsky, L. (1988). El desarrollo de los procesos psicológicos superiores. México: Editorial Crítica, Grupo Editorial Grijalbo.

<https://www.bibliopsi.org/docs/carreras/obligatorias/CFP/educacional/erausquin/Unidad%202/Vigotsky%2%20el%20desarrollo%20de%20los%20p.%20cap%204.pdf>

Woiski, E. R. (2015). Data Analysis and Machine Learning in Python. In *HEALTH MANAGEMENT OF ENERGY SYSTEMS WORKSHOP* (p. 9). Universidade Estadual Paulista, Brazil. https://www.researchgate.net/profile/Amarildo-Paschoalini/publication/292576257_Peridynamics_using_Finite_Elements/links/56af472908ae19a385174231/Peridynamics-using-Finite-Elements.pdf#page=10

Zambrano-Villalba, C., & Almeida-Monge, E. (2017). Clima social familiar y su influencia en la conducta violenta en los escolares. *Revista ciencia UNEMI*, 10(25), 97-102. <https://www.redalyc.org/journal/5826/582661258010/582661258010.pdf>

Vilca Masco, H. (2020). Predicción del nivel de estrés en estudiantes universitarios utilizando técnicas de machine learning. <https://repositorio.upeu.edu.pe/handle/20.500.12840/4096>

Zamorano, R. (2018). Comparativa y análisis de algoritmos de aprendizaje automático para la predicción del tipo predominante de cubierta arbórea. [https://eprints.ucm.es/id/eprint/48800/1/Memoria%20TFM%20Machine%20Learning_Juan_Zamorano_para_difundir%20\(2\).pdf](https://eprints.ucm.es/id/eprint/48800/1/Memoria%20TFM%20Machine%20Learning_Juan_Zamorano_para_difundir%20(2).pdf)

10. ANEXOS

10.1 Anexo 1: Cronograma de actividades

Tabla 20. Cronograma de actividades

Objetivo	Acciones	Fase (F)/Mes								
		F1	F2	F3	F4	F5	F6	F7	F8	
		Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct
Caracterizar desde aspectos socio demográficos los estudiantes de grado décimo de las instituciones Ramón Alvarado y La Merced	Caracterización de la población y muestreo para el estudio. Establecimiento de la relevancia de las variables sociodemográficas dentro de la singularidad de los estudiantes	X								
Interrelacionar variables sociodemográficas y de desempeño académico, implementando una herramienta machine learning que pueda llegar a predecir dificultades en el desarrollo de las competencias básicas de estudiantes de grado décimo a través de un entorno virtual.	Establecer de manera coherente, las variables sociodemográficas y de desempeños que se van a utilizar, tomando en consideración la naturaleza de las mismas y las respectivas categorías			X						
	Estipular los datos que estarían población y el porcentaje considerado para muestras con los que se va a trabajar en los algoritmos			X						
	Establecer los datos a recopilar y la manera de hacerlo, el tiempo de recopilación y el formato en que se manejarán éstos. Por ejemplo, se puede usar un Google forms y tratar los datos con la extensión .xlsx. (un solo formato).				X					
	Recopilados los datos, depurar los que estén incompletos o que llegue a establecerse, no son acordes a los propósitos investigativos				X					
	Buscar herramientas machine learning que existan en el mercado y que puedan permitir analizar datos para hacer informes descriptivos y predictivos					X				
	Realizar un filtro en la búsqueda anterior con indagación bibliográfica que de fundamento a la prueba de algunos algoritmos de clasificación servir en la investigación y de acuerdo a la naturaleza de los datos obtenidos (Ver tabla 2)					X				
	Probar algoritmos ML de clasificación cuyas entradas a relacionar serán las características socio demográficas de					X				

	los estudiantes y la salida sean los niveles de desempeños en competencias básicas								
	Considerar el –o los– modelos que adicional a la muestra de interrelaciones entre las variables, se pueda integrar a un entorno virtual					X			
	Investigar sobre los diferentes códigos, lenguajes y métodos que utilizan las herramientas para hacer análisis predictivo						X		
	Codificar un aplicativo Web usando alguno de los lenguajes de programación, para que funcione como entorno virtual educativo y que además pueda vincularse a algoritmo de predicción de desempeños académicos						X	X	
Evaluar la incertidumbre de la información manejada en el entorno virtual para considerar la fiabilidad de la herramienta <i>machine learning</i> en la predicción de dificultades en competencias básicas de los estudiantes de grado décimo de las instituciones educativas Ramón Alvarado y La Merced	Analizar las métricas de incertidumbre que arrojen las herramientas machine learning y valorar la fiabilidad en el propósito de predecir niveles de desempeños en competencias básicas de los estudiantes					X	X	X	X
	Evaluar la aceptación -en la comunidad educativa de las dos instituciones- del interfaz propuesto dentro del aplicativo Web a través de una encuesta de satisfacción sobre la utilidad del sistema								X
	Realizar una prueba del entorno virtual tanto a nivel descriptivo como predictivo en la que pueda valorarse la integración efectiva en términos operativos del algoritmo ML								X
Presentar una estrategia que permita reorientar practicas pedagógicas que puedan ayudar a mejorar falencias en el desarrollo de competencias básicas	Construir una estrategia pedagógica que permita usar los hallazgos encontrados para posibilitar la mejora en el desarrollo de competencias básicas								X

Fuente: Elaboración propia

10.2 Anexo 2: Gráfica de las instituciones

Institución Educativa Ramón Alvarado



Institución Educativa La Merced



10.3 Anexo 3: Cuestionario real aplicado

https://docs.google.com/forms/d/1IXjDqMcUJPDihBLsAhe23DcnZ8_7P8e_RdK70T-ZAB4/edit?usp=drivesdk

10.4 Anexo 4: Dataset

Primer dataset utilizado
https://drive.google.com/file/d/1vwAnJ7iU8yXyw6tSIZ9LtGaZtBqC1fy_/view?usp=sharing
Segundo dataset utilizado
https://drive.google.com/file/d/1fIK6Miz1OgE6qZX0fyQuVDZ-rOAVB_6o/view?usp=sharing

10.5 Anexo 5: Codificación de los algoritmos en Python

DecisionTreeClassifier
https://drive.google.com/file/d/1-nzTodL6nro8VQl5EsSa0t4dWRDnCl5/view?usp=share_link
K-Nearest Neighbors
https://drive.google.com/file/d/1Fqtr_chqhgAH8G2BTwSvpggg7WdREGIK/view?usp=share_link
Logistic Regression
https://drive.google.com/file/d/1N2Y0VAhegg7sne8kwYcMKY8vMTmzqmCp/view?usp=share_link
Support Vector Machine
https://drive.google.com/file/d/1XMRDDdkZQK9wTVd77pZULzzXrRHfYFT0/view?usp=sharing
Neural Net
https://drive.google.com/file/d/11eFQA2cvuhkLMKf-cmy_TsnEwYhJ44Zb/view?usp=sharing

10.6 Anexo 6: Dirección Web del aplicativo generado

<http://edu-tech.duckdns.org:7080/maestria/>

10.7 Anexo 7: Formato para trabajo interdisciplinar

Mto ¹		Paso ²	DESCRIPCIÓN DEL PASO DENTRO DEL RESPECTIVO MOMENTO DE LA CLASE	COMP
INICIO	1	<u>Formulación y análisis del problema</u>		
	2	<u>Lluvia de ideas y delimitación del problema</u>		
DESARROLLO	3	<u>Investigación Información conocida y desconocida del problema</u>		
	4	<u>Planteamiento de posibles soluciones</u>		
EVALUACIÓN	5	<u>Solución y presentación de conclusiones</u>		
Criterios de evaluación: * *				
Recursos:				

¹ Momento de la clase

² Paso de la estrategia utilizada: ABP (Aprendizaje Basado en Problemas/Proyectos)