CAPÍTULO 3

Marco teórico conceptual

OMNISCENS PUBLISHING

CAPÍTULO 3

Marco teórico conceptual

Theoretical and conceptual framework

DOI: https://doi.org/10.71112/rsbwvb15

Resumen

Este capítulo explora la aplicación de algoritmos de aprendizaje automático —como Random Forest, Naive Bayes y Árboles de Decisión— para predecir el rendimiento académico. Se revisan antecedentes internacionales y nacionales que demuestran la efectividad de estas técnicas en la identificación temprana de estudiantes en riesgo. Además, se fundamenta en teorías educativas como la Taxonomía de Bloom y el Modelo de Productividad de Walberg, que destacan factores cognitivos, afectivos y ambientales clave en el aprendizaje. El marco integra conceptos de minería de datos educativos para mejorar la toma de decisiones en entornos académicos.

Palabras clave: Aprendizaje Automático, Rendimiento Académico, Minería de Datos Educativos, Taxonomía de Bloom, Modelo de Walberg, Predicción.

Abstract

This chapter explores the application of machine learning algorithms—such as Random Forest, Naive Bayes, and Decision Trees—to predict academic performance. It reviews international and national background studies demonstrating the effectiveness of these techniques in the early identification of at-risk students. Furthermore, it is grounded in educational theories like Bloom's Taxonomy and Walberg's Educational Productivity Model, which highlight key cognitive, affective, and environmental factors in learning. The framework integrates concepts from educational data mining to enhance decision-making in academic settings.

Keywords: Machine Learning, Academic Performance, Educational Data Mining, Bloom's Taxonomy, Walberg's Model, Prediction.

Antecedentes del problema

a) Antecedentes internacionales

Parhizkar et al., (2023), En la investigación realizada “Predicción del desempeño de los estudiantes utilizando algoritmos de clasificación de minería de datos: evaluación de la generalización de modelos desde el aspecto geográfico” Irán, describe que el incremento de la productividad en los sistemas educativos es de gran importancia, hoy en día los investigadores están interesados en predecir el rendimiento académico de los estudiantes; Esto se hace para mejorar la productividad general del sistema educativo mediante la identificación efectiva de los estudiantes cuyo rendimiento está por debajo del promedio. Esta preocupación universal se ha combinado con la ciencia de datos, lo que ha llevado a la creación de un área de investigación interdisciplinaria llamada Minería de Datos Educativos. Uno de los temas recientes que ha sido abordado por los investigadores es la formación de modelos generalizables desde diferentes aspectos como el género, la especialidad, la geografía, etc. Por lo tanto, en esta investigación utilizamos métodos de aprendizaje automático para predecir el rendimiento de los estudiantes, haciendo énfasis en el entrenamiento de modelos generalizables desde el punto de vista geográfico. Para ello, se diseñó un cuestionario de 37 preguntas, a través del cual se recogieron 536 respuestas, de las cuales 111 fueron internacionales y 425 nacionales. De acuerdo con la literatura, el rendimiento de los estudiantes se determina principalmente en función del GPA (promedio de calificaciones) de todo el curso. En esta investigación, se recopiló información sobre el promedio de los encuestados en cursos de pregrado y posgrado en forma de tres clases. Después de una revisión final de los modelos empleados en estudios previos, los principales modelos seleccionados y utilizados para fines de clasificación incluyeron SVM, CNN, Adaboost, RF, SVM y DT. La selección de características se realiza mediante XGBoost, bosque aleatorio y SVML1. La principal cuestión investigada en este estudio es la generalización de los modelos entrenados con datos nacionales (iraníes) y probados con datos internacionales (no iraníes). Los resultados experimentales muestran que los mejores modelos entrenados con conjuntos de datos específicos recopilados en esta investigación tuvieron generalización en comparación con los resultados de los modelos base que se entrenaron y probaron con datos nacionales. Por su parte, los modelos Random forest y CNN muestran el mejor rendimiento con un promedio de precisión y una puntuación F de 73,5 y 68,5, respectivamente.

Morales Hernández et al. (2022), realizaron un estudio titulado "Algoritmos de aprendizaje automático para la predicción del logro académico" en colegios ubicados en el estado de

Tlaxcala, México. En este estudio, se desarrollaron dos aplicaciones inteligentes: una red neuronal multicapa (perceptrón multicapa [MLP]) y un modelo de potenciación del gradiente (GB), con el objetivo de predecir el nivel de productividad académica en los cursos de español y matemáticas para alumnos del grado sexto del nivel primario y tercero de secundaria, cuyas variables pertenecieron a las evaluaciones nacionales del logro académico de los colegios escolares del estado mencionado de México. Es importante mencionar que se utilizaron 13 variables de entrada, cuya importancia fue determinada por el algoritmo de bosque aleatorio (RF). Las ententes inteligentes MLP y GB fueron entrenados y probados con un conjunto de datos de 11,036 registros pertenecientes a los escolares en el sistema escolar entre 2008 y 2011. Los modelos inteligentes ejecutaron predicciones en función de los registros de los periodos 2008 y 2011. En la asignatura de español, el modelo MLP superó al modelo GB con una precisión global de clasificación (PG) de 70.1 % en 2008 y 61.1 % en 2011. En referencia al curso de matemáticas, el modelo GB mostró mejores resultados, con una PG de 68.8 % en 2008 y 63.5 % en 2011. Asimismo, se pudo identificar una fuerte correlación entre el puntaje en del curso español y el grado de logro académico en el curso de matemáticas. Es importante destacar que los puntajes en español y matemáticas fueron relevantes, en comparación con factores contextuales como el sexo, beca y el turno escolar. Conforme a la muestra analizada, se observó que, en ambas asignaturas, el sexo femenino superó al sexo masculino en los grados de logro académico elemental, bueno o excelente; pero se invertiría en el grado de logro insuficiente.

Yağcı (2022), en su modelo de estudio denominado Minería de datos educativos: predicción de la productividad académica de los estudiantes de una universidad estatal turca mediante algoritmos inteligentes, propone predecir calificaciones de los exámenes finales de los estudiantes de pregrado, en donde referencia las calificaciones de las evaluaciones parciales para dicho estudio, asimismo utiliza algoritmos ML Random Forest, Vecinos más Cercanos, Regresión Logística, Naive Bayes y algoritmos de k-vecino más cercano para dicha investigación, para ello utilizó una muestra de 1,854 estudiantes del curso de idioma turco I en una universidad estatal en Turquía, durante el semestre de otoño de 2019-2020, cuyo resultado del modelo propuesto logró una precisión de clasificación del 70% al 75%; cabe mencionar que para las predicciones consideró tres tipos de parámetros calificaciones de las evaluaciones parciales, datos del departamento y datos de la facultad, en donde atribuye que dichos datos son muy importantes para el análisis de aprendizaje en la educación superior a fin de contribuir los procesos de toma de decisiones. El autor menciona que la contribución de la investigación está orientada a la predicción temprana de estudiantes con alto grado de fracaso y determina los métodos de aprendizaje automático más efectivo.

Nedeva y Pehlivanova (2021), realiza el trabajo de investigación "Análisis de rendimiento de los estudiantes utilizando algoritmos de aprendizaje automático en WEKA", cuyo objetivo es pronosticar la productividad de los estudiantes de la Universidad Trakia de Stara Zagora de Bulgaria, con la finalidad de apoyar y mejorar los resultados educativos, asimismo identifica características más significativas (12 atributos con impacto en el desempeño de los estudiantes) que afectan el desempeño de los estudiantes, luego eligen el algoritmo inteligente más eficiente para predecir su desempeño. Se comparó la eficiencia de cuatro algoritmos de clasificación: BayesNet (BN), Perceptrón multicapa (MLP), Optimización mínima secuencial (SMO) y Árbol de decisión (J48), los mejores resultados se obtuvieron utilizando el algoritmo Multilayer Perceptron MLP, aplicado a 12 atributos, con valores principales: Tasa TP = 0,974; Precisión = 0,976, Medida F = 0,974 y Exactitud = 97,39 %, el cual concluyo con una mejor precisión en el pronóstico.

Yildiz y Börekci (2020), en la investigación realizada acerca de la Predicción de la productividad académica de los estudiantes mediante técnicas de aprendizaje automático, desarrolla una visión de los datos educativos recopilados en estudiantes del noveno grado, para ello se utilizaron datos con información demográfica de estudiantes y familias, rutinas de estudio, comportamientos de asistencia y creencia epistemologías sobre la ciencia, dicha investigación tuvo como objetivo resolver un problema de clasificación para ello intenta estimar clases de aprobado y desaprobado en los exámenes, asimismo dicho estudio hizo una predicción comparativa de precisión con los algoritmos de clasificación supervisada, definiendo que variables eran más efectivas en la formación académica, adicionalmente se examinó el coeficiente de ganancia de información de las variables para determinar los factores que afectan la precisión de la predicción.

En el estudio, se comparó la precisión de predicción de los algoritmos de clasificación supervisada y se definió qué variables eran efectivas en la formación de clases. Cuando se comparó la precisión de predicción de los algoritmos inteligentes, los resultados indicaron que el algoritmo de red neuronal 98,6 %) obtuvo la puntuación más alta. La tasa de precisión de los otros algoritmos son K vecinos más cercanos (KNN) (86,2 %), regresión logística (78,4 %), SVM (90,3 %), árbol de decisión (91,9 %), bosque aleatorio (Random forest) (90,0 %) e Naive Bayes (81,7 %). Se examinó el coeficiente de ganancia de información de las variables para determinar los factores que afectan la precisión de la predicción. Se puede concluir que existió una relación entre estas variables y el éxito académico. Los estudios sobre estas variables apoyarán el éxito académico de los estudiantes. Finalmente, el autor recalca que las variables de investigación apoyaran el éxito académico de los estudiantes.

b) Antecedentes nacionales

Gismondi (2021), en la investigación titulada *"Modelo predictivo basado en machine learning como soporte para el seguimiento académico del estudiante universitario",* cuya finalidad fue mejorar los resultados en la educación universitaria, propone aplicar la inteligencia artificial, machine learning y Deep learning, para ello considera atributos relevantes a fin de proponer un modelo de predicción de aprendizaje profundo con el algoritmo de redes neuronales con 2, 3, 4, 5, 6 y 7 capas con el respectivo entrenamiento y prueba, para ello se concluyó que se obtuvo mejores resultados con redes neuronales de 6 capas con un 98.97% de precisión en el entrenamiento y 81.73 de precisión en el conjunto de prueba.

Gismondi y Huiman (2021), elaboran la investigación *"Características para un modelo de predicción de la deserción académica universitaria. Caso Universidad Nacional del Santa",* proponen un *modelo* predictivo a fin de disminuir la deserción académica de estudiantes de la Universidad Nacional de Santa, para ello propone atributos que formaran parte de un modelo inteligente basado en Machine Learning, para ello se realizó una revisión documental y métodos estadísticos; finalmente concluyen la propuesta con seis características académicas y doce indicadores dentro de las características no académicas.

Díaz-Landa et al. (2021), propone la investigación realizada "Rendimiento académico de estudiantes en Educación Superior: predicciones de factores influyentes a partir de árboles de decisión", cuyo objetivo es predecir el rendimiento académico de estudiantes de maestrías en educación, para ello empleó técnicas de árbol de decisión, minería de *datos* y algoritmo J48 en la herramienta WEKA, asimismo considero dimensiones educativas, familiares, socio económicos, hábitos personales y costumbres de una población de 237 estudiantes de una universidad pública Peruana, luego efectuó el entrenamiento del modelo, logrando establecer que el algoritmo de clasificación bayesiano J48 una efectividad del 66.6%, identificando asimismo mediante técnicas inteligentes los factores influyentes en el rendimiento académico: enseñanza, convenientes horarios de clase, adecuada relación social docente-estudiante, nivel académico.

Quiñones y Quiñones (2020), llevaron a cabo una investigación titulada "Rendimiento académico empleando minería de datos", cuyo propósito fue predecir el rendimiento académico de los estudiantes de la especialidad de Industrias Alimentarias en la Universidad Nacional de Jaén, mediante modelos de minería de datos. Los datos históricos en el estudio fueron recopilados a través de formularios y registros de las oficinas dicha universidad. La metodología aplicada fue CRISP-DM; cuyos pronósticos de los modelos inteligentes se realizaron en el software weka cuyos resultados fueron superiores al 83% de efectividad.

Chiok (2017), presenta una investigación titulada "Predicción del rendimiento académico aplicando técnicas de minería de datos, curso de Estadística General de la Universidad Nacional Agraria La Molina (UNALM)", donde acentúa que el rendimiento académico de los estudiantes es una de las principales preocupaciones para las instituciones de educación superior. El estudio destaca que las técnicas de minería de datos (TMD) aplicadas a entornos educativos se han convertido en una herramienta eficaz para pronosticar la productividad académica, a fin de identificar los factores más influyentes en el aprendizaje de los estudiantes y apoyar a los docentes a optimar el proceso enseñanza aprendizaje con acciones más eficaces y oportunas. La investigación aplico TMD como el algoritmo de regresión logística, algoritmo árboles de decisión, algoritmo de redes bayesianas y el algoritmo de redes neuronales, cuyos datos de entrenamiento pertenecieron a los estudiantes de la asignatura de estadística general de UNALM durante los semestres académicos 2013-II y 2014-I, con la finalidad de pronosticar la evaluación final de aprobación o reprobación del futuro en estudiantes del curso. Los resultados del estudio revelan que el algoritmo Naive Bayes alcanzó la mayor tasa de efectividad, con un 71.0% de precisión.

Bases teóricas o científicas

a) Teoría de la productividad educativa

Taxonomía de bloom de los objetivos de aprendizaje

Según Gogus (2012), la taxonomía es el proceso científico de clasificar y organizar; asimismo en el sentido académico los objetivos de aprendizaje es el logro y demostración que se espera del estudiante luego de un periodo de tiempo. La taxonomía de Bloom desarrolla una jerarquía de objetivos educativos que ***contenía categorías en el dominio cognitivo***, en donde menciona al conocimiento, luego a la comprensión, seguido de la aplicación, posteriormente al análisis, seguido de la síntesis y finalmente a la evaluación (Amer, 2006, como se citó en Rosales Sánchez et al., 2020).

Sepúlveda et al. (2020) menciona que la taxonomía de Bloom se convirtió en un componente esencial para la estructura y comprensión de las actividades educativas basadas en el proceso enseñanza aprendizaje en donde identifico dominios de las actividades educativas que a continuación se describen:

- ***Dominio cognitivo***: Adquisición de conocimientos y habilidades intelectuales (Conocimiento).
- ***Dominio afectivo***: Se refiere a la integración de credos e ideas, fundamentada en las actitudes y sentimientos que participan en el proceso. (Actitud).

- ***Dominio psicomotor***: Basado en la adquisición de movimientos coordinados, tanto manuales como físicos (Habilidades).

Bloom también define dominios y subdominios dentro de las seis categorías principales (niveles), organizándolas de lo simple a lo complejo y a partir de lo definido hacia lo abstracto (Krathwohl, 2002). Además, se asumía que el dominio de una categoría más sencilla era esencial para poder dominar la siguiente categoría de mayor complejidad (Ver Tabla 1).

Tabla 1

Taxonomía de Bloom de los objetivos de aprendizaje para el dominio cognitivo

Dominio	Niveles
Cognitivo simple o bajo	1. ***Conocimiento***, capacidad de memorizar y recordar hechos específicos. 2. ***Comprensión***, capacidad de interpretar y demostrar comprensión básica.
Cognitivo superior	*3.* ***Aplicación***, capacidad de aplicar conceptos y separar conceptos en componentes. *4.* ***Análisis***, capacidad de analizar conceptos *5.* ***Síntesis***, capacidad para combinar elementos y formar un todo. *6.* ***Evaluación***, capacidad de emitir juicios sobre el valor de lo aprendido.
Habilidades del pensamiento crítico.	

Nota: extraído de (Gogus, 2012)

Asimismo, Serumena D. et al. (2021), destaca que el domino afectivo esta caracterizado por las emociones, valores, apreciación, entusiasmo, interés, motivación y actitudes, dichas categorías están ordenadas del más simple al más complejo:

1ro. ***Recepción***. capacidad de prestar atención y responder a estimulación adecuada, la aceptación es el más bajo.

2do. ***Sensible***, los estudiantes se involucran de manera afectiva, participan y toman interés.

3ro. ***Valor adherido***, se refiere al valor o importancia de atracción a ciertos objetos o eventos con reacciones a la aceptación, rechazo o ignorar el objeto. El objetivo está basado en "actitud / aprecio", en la valoración de lo bueno y lo malo.

4to. ***Organización***, centrado a la unificación de valores, actitudes que generan conflictos internos que va determinando el comportamiento que es representado en una filosofía de vida y la capacidad de formar un sistema de valores y cultura organizacional.

5to. ***Caracterización***, se refiere al carácter y la fuerza vital de la persona, considerando los valores están evolucionados para que el comportamiento se vuelva consistente y predecible.

Finalmente, los dominios psicomotores incluyen el aspecto físico y coordinación, habilidades motoras y capacidades físicas:

1ro. ***Imitación***, cuando empieza a responder de manera similar a los observados.

2do. ***Manipulación***, se refiere a la habilidad de seguir direcciones, apariencias, movimientos seleccionados determinando una apariencia a través de la práctica.

3ro. ***Idoneidad***, adquiere mayor precisión, proporción y certeza en la apariencia.

4to. ***Articulación***, establece coordinación de movimientos de manera exacta y logrando consistencia esperada entre diferentes movimientos.

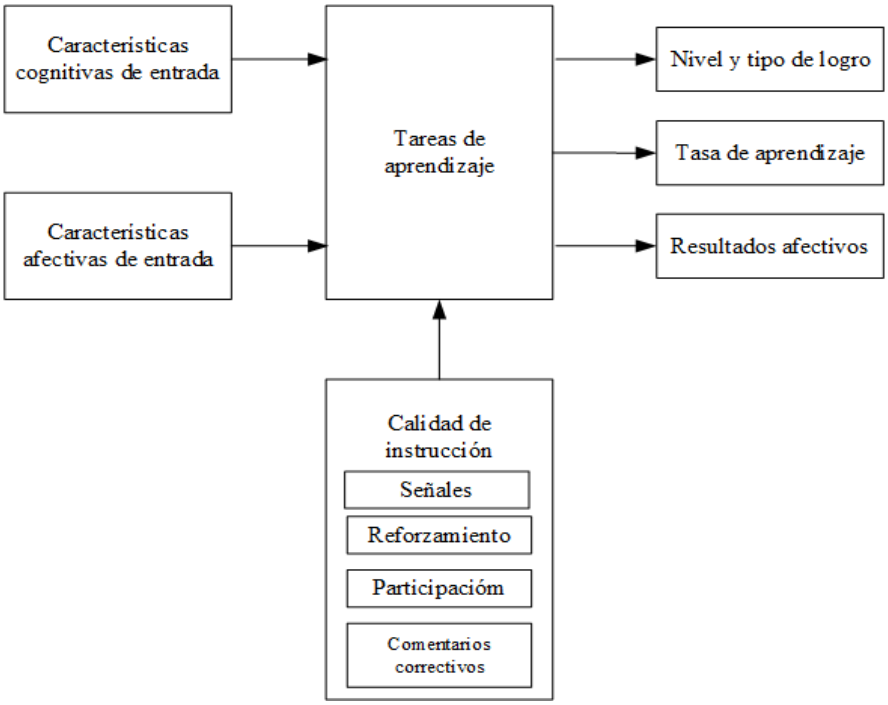
5to. ***Experiencia***, su conducta expresa menor gasto energético físico y psíquico.

Modelo de aprendizaje bloom

Según, Seel (2012) el modelo de aprendizaje de Bloom propuesto en 1976 , tiene una característica de carácter cíclico de la instrucción, lo cual significa que los resultados obtenidos de una tarea, se convierte en el aporte del anterior; el modelo está representado por cuatro dimensiones: características cognitivas de entrada, características afectivas de entrada, calidad de instrucción, los resultados de aprendizaje se especifican por el nivel y tipo de logro, tasa de aprendizaje o resultados afectivo (Figura 1).

Figura 1

Modelo de aprendizaje de Bloom



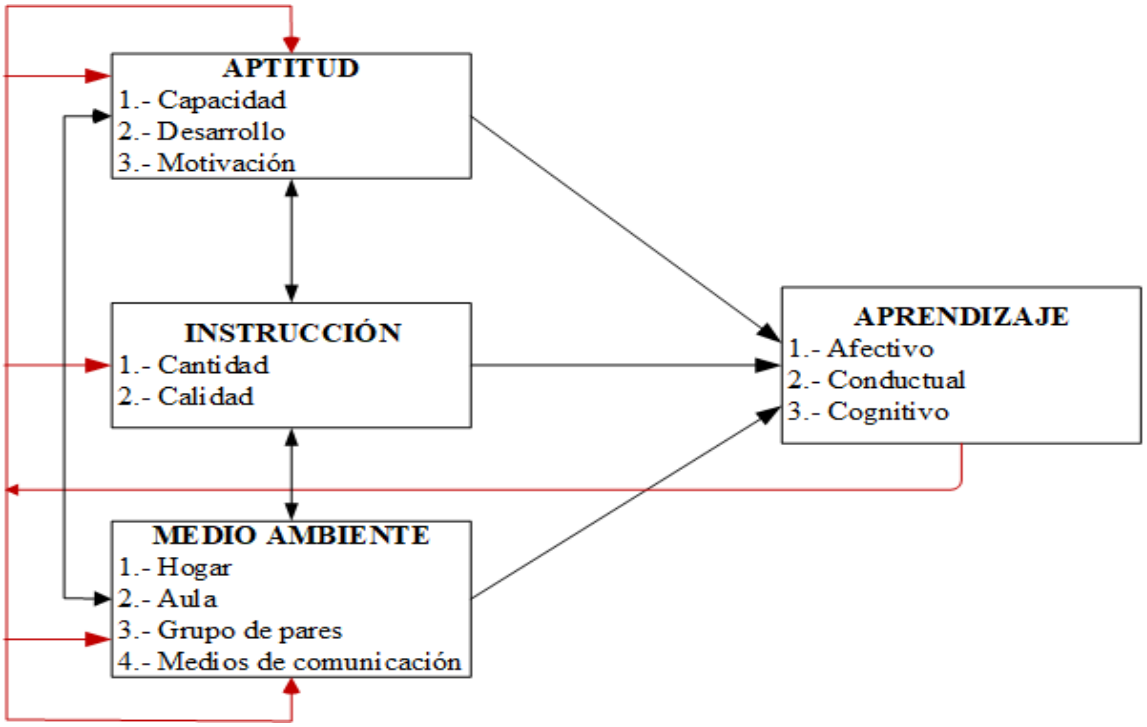
Nota: Extraído de (Seel, 2012a)

Modelo de la productividad educativa

Modelo propuesto por Walberg (1984), basado en el rendimiento académico del estudiante, propone cuatro componentes (Figura 2) que se combinan e influyen entre sí, con la finalidad de determinar el rendimiento final del estudiante, *asimismo recalca si el estudiante tiene habilidades favorables y ambientes favorables bajo estímulos, entonces el aprendizaje puede ser positivo en términos de actitudes, comportamientos y conocimientos.* Es importante considerar que la teoría de Walberg menciona que las características psicológicas influyen en los resultados de aprendizaje individual del estudiante en lo cognitivo, conductual y actitudinal (Mazana et al., 2019), (Peconcillo Jr et al., 2020), por tanto dichas características psicológicas influyen directamente en el rendimiento académico (Reynolds y Walberg, 1992; Wang et al., 1993, como se citó en Abid et al., 2021), es decir que las nueve variables (grafico 4) contribuyen al rendimiento académico (Walberg y Tsai, 1985; Wang et al., 1993, como se citó en Abid et al., 2021).

Figura 2

Modelo de productividad educativa



Nota: obtenido de (Walberg, 1984)

a) Aptitud del estudiante

La aptitud se refiere a las habilidades y capacidades innatas de los estudiantes. Incluye tanto las habilidades cognitivas (como la inteligencia, el razonamiento lógico y la memoria) como las habilidades no cognitivas (como la motivación, la perseverancia y la autoestima). El comportamiento del estudiante puede influir en su aptitud al afectar su nivel de compromiso, su actitud hacia el aprendizaje y su disposición para superar desafíos, asimismo, implica:

- **Capacidad**: Se refiere a la habilidad innata de un individuo para aprender y comprender nuevos conceptos. Los indicadores de capacidad pueden incluir la rapidez con la que se adquieren nuevas habilidades cognitivas.
- **Motivación**: La disposición y el interés del estudiante para participar activamente en el proceso de aprendizaje. Los indicadores de motivación podrían incluir la persistencia en la resolución de problemas y el entusiasmo por aprender.

- **Desarrollo**: Se refiere al progreso y crecimiento del estudiante a lo largo del tiempo. Indicadores de desarrollo pueden ser la mejora continua en el desempeño académico y el desarrollo de habilidades sociales.

b) Instrucción

Se refiere al tiempo que el estudiante se involucra en el aprendizaje. Asimismo, a la calidad de experiencia educativa, que incluye aspectos psicológicos y curriculares, los cuales involucra:

- **Cantidad**: La cantidad de tiempo y recursos dedicados a la enseñanza. Indicadores de cantidad pueden incluir la duración de las clases, la cantidad de material didáctico proporcionado y el tiempo total de instrucción.
- **Calidad**: La efectividad de la enseñanza. Indicadores de calidad podrían abarcar la claridad en la presentación de conceptos, la retroalimentación constructiva y la adaptación de métodos de enseñanza a las necesidades individuales de los estudiantes.

c) Medio Ambiente

Son factores constantes que afectan al aprendizaje como el clima psicológico y estímulos educativos: el hogar, grupo social del aula, compañeros fuera de la institución educativa, usos del tiempo extra escolar (actividades libres de realizar) los cuales involucra:

- **Hogar**: El entorno familiar en el que crece el estudiante. Indicadores del hogar podrían incluir el apoyo de los padres, la disponibilidad de recursos educativos en casa y la estabilidad del entorno familiar.
- **Entorno comunitario**: La influencia del entorno fuera del hogar. Indicadores del entorno comunitario pueden abarcar el acceso a servicios educativos adicionales, la presencia de modelos a seguir en la comunidad y la seguridad del entorno.
- **Medios de comunicación**: El impacto de los medios de comunicación en la educación. Indicadores de medios de comunicación podrían incluir el acceso a recursos educativos a través de medios electrónicos, la calidad de los programas educativos y el tiempo dedicado a actividades educativas frente a las pantallas.

d) Aprendizaje

En esta etapa esta referido a los logros obtenidos o conocimientos nuevos durante el proceso del aprendizaje como un proceso de producción y pueden ser de distintos tipos afectivo, conductual o cognitivo. El aprendizaje será mayor cuando el ambiente es cooperativo con

orientaciones claras y con objetivos y con estímulos necesarios para el aprendizaje, cuyo indicador resaltante es:

- **Logro***:* de detalla el valor del nivel de éxito alcanzado por el estudiante en términos de conocimientos adquiridos y habilidades desarrolladas. Indicadores de logro pueden abarcar calificaciones académicas, resultados en pruebas estandarizadas y la aplicación exitosa de conocimientos en situaciones prácticas.

Modelo de aprendizaje escolar de carroll

El modelo de Carroll propuesto en 1960 (Figura 3) destaca los roles distintivos de las habilidades, aptitudes para el logro del aprendizaje; la eficacia se determina en función del tiempo para aprender y del tiempo realmente dedicado al aprendizaje (Guill et al., 2021; Rosales Sánchez et al., 2020).

Ecuación funcional de grado de aprendizaje:

$$\text{Grado de aprendizaje} = \frac{\text{tiempo real necesario para el aprendizaje}}{\text{tiempo realmente necesario para el aprendizaje}}$$

Para ello referencia al aprendizaje como activo cuantificado del éxito logrado durante un periodo fijo de tiempo, el mismo que puede variar dicha tarea de aprendizaje desde minutos, días, semanas o meses para tareas complejas, asimismo, dependerá de otras variables: comprensión y aptitud del estudiante para el logro del encargo académico (Seel, 2012b).

Asimismo, Seel, menciona que Carroll (1989) establece la existencia de cinco variables en dos clases principales concebidas para influir en el grado de aprendizaje; ello se resume de la siguiente manera:

Variables instructivas:

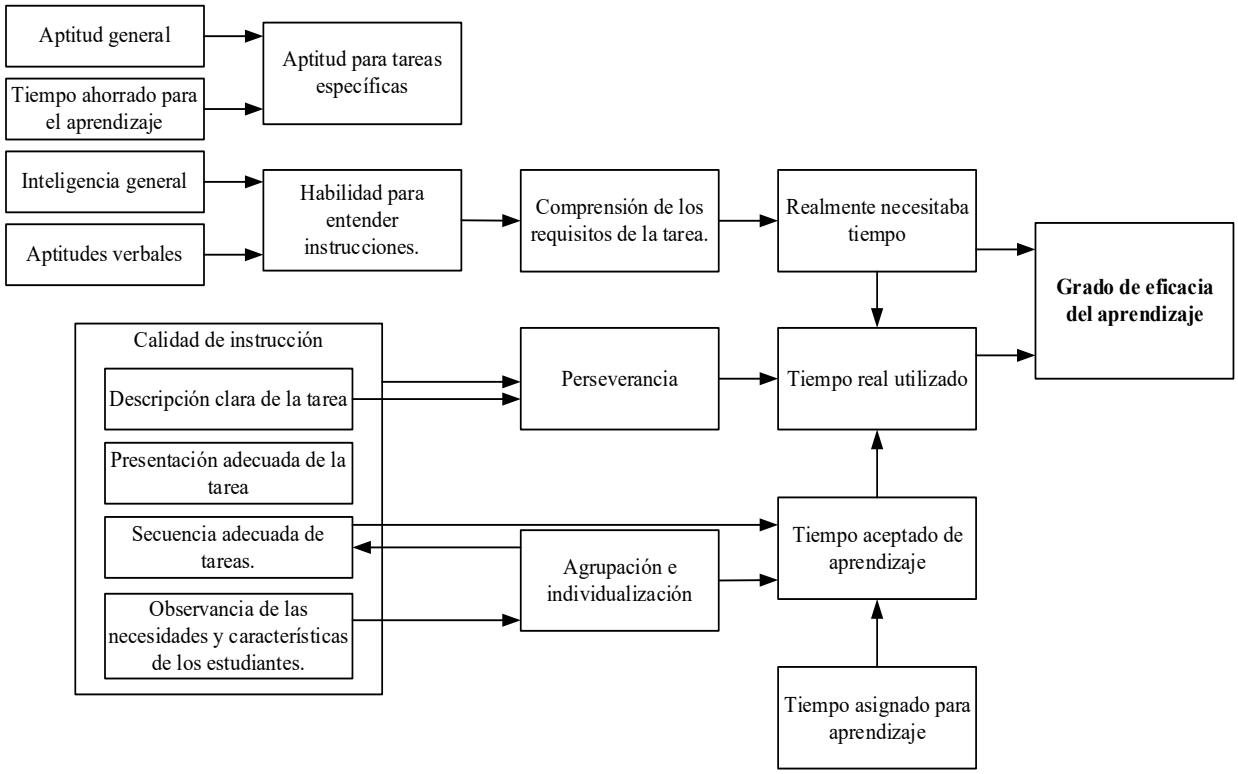
- Calidad de la instrucción (trabajo presentado)
- Oportunidad (tiempo de aprendizaje)

Variables de diferencias individuales

- Aptitud del estudiante para aprender la tarea en función del tiempo, capacidad y logros previos.
- Capacidad de entendimiento de la instrucción (inteligencia y habilidad verbal)
- La "perseverancia" del estudiante: el tiempo como inversión de aprendizaje.

Figura 3

Modelo de aprendizaje Carroll



Nota: obtenido de (Seel, 2012b)

b) Teoría de algoritmo de aprendizaje

La inteligencia es la capacidad de aprender, entender y pensar en las cosas. Es una capacidad intelectual general que implica la capacidad de razonar, planificar, resolver complicaciones, pensar abstractamente, percibir ideas y aprender (Sauce et al., 2022). Asimismo, tiene la capacidad para llegar a la salida más adecuada a pesar de todas las probabilidades en una variedad de situaciones (Vostroknutov et al., 2018); puede adquirir conocimientos para superar dificultades y producir resultados apropiados presentados(Zhao, 2022); por otro lado, la Inteligencia Artificial (IA) se puede determinar cómo inteligencia ejecutada por un sistema fabricado o hecho por el hombre. Es un intento de hacer que una máquina se comporte de maneras que se llamarían inteligentes si un humano hiciera lo mismo. La Inteligencia Artificial (IA) es un campo diferente donde los investigadores persiguen una amplia gama de complicaciones, utilizan una selección de métodos y persiguen un espectro de objetivos

científicos, según Wang (2019) se debe establecer cuatro criterios para para una definición funcional de la IA basado en especificar su uso común, establecer un límite claro, conducir a una investigación exitosa y debe ser lo más simple posible el desarrollo un sistema inteligente. El objetivo principal de un sistema basado en Inteligencia Artificial (IA) es alcanzar un nivel humano de inteligencia(Vigo et al., 2022), (Dong et al., 2020); a pesar de existir diferentes definiciones, el entendimiento común de la IA está enfocada a máquinas y computadoras con la finalidad de ayudar a la humanidad a resolver problemas y facilitar los procesos de trabajo (Tai, 2020).

En la rama de IA encontramos el enfoque del aprendizaje automático (Machine Learning (ML)) basado en una colección de algoritmos que intentan extraer patrones de los datos y asociar dichos patrones con clases discretas de muestras en los datos, asimismo dentro del enfoque encontramos métodos como ML supervisado, no supervisado y por refuerzo, que intentan encontrar una secuencia de acciones con la finalidad de lograr un objetivo específico (Jovel y Greiner, 2021), Para (Kelleher, Mac Namee, & D´Arcy, 2015), el aprendizaje automático es una disciplina científica que capacita a las máquinas para desarrollar técnicas de aprendizaje. Este proceso automatizado extrae patrones de los datos y construye modelos que posibilitan la predicción mediante algoritmos supervisados, estableciendo conexiones entre características descriptivas actuales y objetivos basados en un conjunto de ejemplos o instancias históricas. Este proceso consta de dos pasos principales: data set, aprendizaje supervisado y su posterior pronóstico.

Por otra parte (Mitchell, 1997), establece que el aprendizaje automático se fundamenta en conceptos de diversas disciplinas, como la disciplina de la inteligencia artificial, la disciplina de la probabilidad y la estadística, la disciplina de la complejidad computacional, la teoría de la información, teoría de psicología y neurobiología, teoría del control y la disciplina de la filosofía. Según Jovel y Greiner (2021), la implementación de estos conceptos a través de algoritmos computacionales mejora el rendimiento en función de la experiencia de aprendizaje; para Janiesch et al. (2021) los algoritmos inteligentes tienen la capacidad para aprender de los datos conforme a entrenamientos específicos del problema a fin de automatizar el proceso de creación del modelo y posteriormente resolver tareas asignadas.

Según Bangato J. I. (2020) menciona que es una subdisciplina de la Inteligencia Artificial basado en “cómo construir programas de computadora que mejoran automáticamente adquiriendo experiencia”, menciona que la implementación desarrollada con ML no necesita de reglas o funciones explicitas del lenguaje de programación, sino que se desarrolla de manera automática, asimismo, destaca de los algoritmo inteligentes utilizan pocos recursos durante el

proceso de entrenamiento de grandes volúmenes de datos y cuyo aprendizaje se destaca de manera autónoma. Clasifica al aprendizaje automático en dos tipos: una referida al aprendizaje supervisado o al aprendizaje no supervisado y la otra refiere al aprendizaje por reforzamiento de los algoritmos.

Podemos mencionar que los algoritmos inteligentes más utilizados son: los algoritmos de árboles de decisión, algoritmos de regresión lineal, algoritmos de regresión logística, algoritmos k Nearest Neighbor, algoritmos PCA / Principal Component Analysis, algoritmos SVM, algoritmos gaussian Naive Bayes, algoritmos K-Means, algoritmos redes neuronales artificiales, algoritmos de aprendizaje profundo.

Aprendizaje Supervisado

Según, Jiang et al. (2020) el aprendizaje supervisado consiste en métodos diseñados para predecir o clasificar un resultado acorde a un objetivo específico, asimismo el aprendizaje se utiliza para describir tareas de predicción porque el objetivo es pronosticar/clasificar un resultado específico de interés; por otro lado, Stern et al.(2020) sostiene que los algoritmos de aprendizaje supervisado pueden aprender características sutiles que distinguen una clase de ejemplos de entrada de otra; adicionalmente, Pugliese et al. (2021) menciona que el aprendizaje supervisado se basa en tareas para aprender una función establecida, la misma que ha sido asignada una entrada a una salida en función de pares de entrada-salida de la muestra; dicho proceso de aprendizaje está basado en datos representados por un dominio de datos con similitudes genéricas (Münch et al., 2020).

Según Lampropoulos y Tsihrintzis (2015), el aprendizaje supervisado se encarga de tareas predictivas, de clasificación y constituye la categoría principal de aprendizaje. En el aprendizaje supervisado, los objetos que están relacionados con un concepto específico son pares de patrones de entrada y salida. Esto significa que los datos que pertenecen al mismo concepto ya están asociados con valores objetivo, como, por ejemplo, clases que definen las identidades de conceptos, por su parte Bagnato J.I. (2020), en el aprendizaje supervisado, los datos de entrenamiento contienen etiquetas, o la respuesta deseada. Determinar si el correo entrante es spam es un buen ejemplo. Una de las muchas características que deseamos entrenar debe ser si la entrada es spam, denotada por un 1 o un 0. Otro ejemplo es cuando tenemos que añadir el precio que descubrimos en nuestro conjunto de datos y tenemos que estimar valores numéricos, como el precio de una casa, basándonos en sus características (metros cuadrados, número de habitaciones, calefacción, distancia al centro, etc.).

El algoritmo k-vecinos más cercanos, algoritmo regresión lineal, algoritmo regresión logística, algoritmo máquinas de vectores de soporte, algoritmo clasificadores bayesianos, algoritmo arboles de decisión, algoritmo bosque aleatorio, algoritmo redes neuronales y los algoritmos de aprendizaje profundo son los algoritmos más utilizados en el aprendizaje supervisado.

Aprendizaje No Supervisado

El modelo no supervisado según Lee et al. (2022) es el sistema que desarrolla el aprendizaje sin intervención humana para predecir; dicho modelo no construye ningún modelo predictivo, para la evaluación de nuevos datos, existen dos opciones; ya sea el mapeo en el espacio agrupado o el agrupamiento o la reducción de la dimensión se realiza con todos los datos una vez más(Tebani y Bekri, 2020); los algoritmos no supervisados se utilizan con frecuencia para determinar las etiquetas de los grupos de datos que no tienen etiquetas (KOÇOĞLU, 2022).

Según Zheng (2019) el desarrollo del aprendizaje no supervisado se basa en agrupar los datos conglomerados donde dichos elementos son muy similares y los elementos de diferentes conglomerados muestran diferencias mucho mayores. Debido a que los elementos no tienen etiquetas de clase, que se conocen en tareas de clasificación o aprendizaje supervisado, la agrupación en clústeres se reconoce como aprendizaje automático no supervisado.

Asimismo, Lampropoulos y Tsihrintzis (2015) explica que se trata de comprender o encontrar una descripción concisa de datos mediante el mapeo pasivo o la agrupación de datos de acuerdo con algunos principios de orden. Esto significa que los datos constituyen solo un conjunto de objetos donde no se dispone de una etiqueta para definir el concepto asociado específico como en el aprendizaje supervisado. Así, el objetivo del aprendizaje no supervisado es crear grupos-clusters de objetos similares de acuerdo con un criterio de similitud y luego inferir un concepto que se comparte entre estos objetos. Asimismo, el aprendizaje no supervisado incluye algoritmos que tienen como objetivo proporcionar una representación de espacios de alta a baja dimensión, preservando la información inicial de los datos y ofreciendo un cálculo más eficiente. Por su parte (Bagnato, 2020) establece que a diferencia del aprendizaje supervisado, el aprendizaje no supervisado no incluye etiquetas predeterminadas con los datos de entrenamiento. Más bien, el algoritmo se encarga de ordenar y evaluar automáticamente los datos, encontrando estructuras o patrones ocultos en la información. Este método es utilizado comúnmente en la segmentación de usuarios en sitios web o aplicaciones, donde los usuarios son agrupados por la inteligencia artificial basándose en características compartidas que el sistema reconoce por sí mismo. La agrupación de K-Means, el análisis de componentes principales y la detección de anomalías son algunos de los algoritmos más

importantes del aprendizaje no supervisado; todos ellos son esenciales para identificar patrones subyacentes y descifrar datos complejos sin necesidad de supervisión humana.

Aprendizaje por refuerzo

Según, Esteso et al. (2022) el aprendizaje por refuerzo es similar a la forma en que aprenden los humanos y los animales, es decir que el aprendizaje por refuerzo dentro del contexto más amplio de la inteligencia artificial permite que los agentes algorítmicos reemplacen a los seres humanos en el mundo real, destacando su participación en diferentes escenarios casas, edificios etc., cuyos dominios hasta ahora consideradas más allá de las capacidades actuales (Al-Ani y Das, 2022); por otro lado Xiang et al. (2021) menciona que el aprendizaje por refuerzo (RL) es un enfoque para simular el proceso de aprendizaje natural del ser humano, es decir imita el proceso de aprendizaje de los humanos, entrena cometiendo y luego evitando errores,, puede resolver algunos problemas que los métodos convencionales no pueden resolver, en algunas tareas, también tiene la capacidad de superar a los humanos; un aspecto destable según, Singh et al.(2022) es que el sistema de aprendizaje por refuerzo, no se proporcionan pares de entrada y salida, al estado actual del sistema se le asigna una meta específica y un conjunto de acciones permitidas y restricciones para esperar sus resultados.

Según, Lampropoulos y Tsihrintzis (2015), el aprendizaje por refuerzo implica realizar acciones para lograr un objetivo, el agente aprende por ensayo y error al realizar una acción, asimismo espera una recompensa; de esta manera el método es eficiente para desarrollar estrategias de acción dirigidas a objetivos. El aprendizaje por refuerzo se inspiró en teorías psicológicas relacionadas y está fuertemente relacionado con los ganglios basales del cerebro. Las metodologías de aprendizaje por refuerzo se relacionan con problemas donde el agente de aprendizaje no sabe a priori lo que debe hacer, por lo tanto, el agente debe descubrir una política de acción que maximice la "ganancia esperada". El aprendizaje por refuerzo difiere del aprendizaje supervisado porque en el aprendizaje por refuerzo no se presentan pares de entrada / salida, ni se corrigen explícitamente acciones subóptimas, sino que los agentes en un momento específico caen en un estado y con esta información seleccionan una acción, como consecuencia, el agente recibe por su acción una señal de refuerzo o recompensa; finalmente, Bagnato Juan Ignacio (2020), describe que durante este tipo de sistemas, el agente se comporta como un "agente autosuficiente" que debe descubrir un entorno desconocido, denominado "espacio", para decidir los movimientos a realizar mediante un procedimiento de ensayo y error. A través de este procedimiento, el agente aprende de forma independiente, recibiendo recompensas por los movimientos correctos y consecuencias por los errores, lo que le permite

averiguar la forma más beneficiosa de llevar a cabo responsabilidades como recorrer un camino, resolver un rompecabezas o tal vez cultivar técnicas de recreación en entornos como Pac-Man o Flappy Bird. Este aprendizaje continuo permite al agente formular la mejor estrategia viable, denominada "política", que le ayudará a tomar decisiones para maximizar las recompensas en el menor tiempo posible y de la forma más ecológica posible. Las políticas, por tanto, definen los movimientos que deben realizarse en cada situación concreta que se le presente al agente. Entre las máximas modas usadas habitualmente en este método está Q-Learning, que se basa totalmente en los Procesos de Decisión de Markov (MDP), esenciales para la toma de decisiones en entornos en los que no siempre se dispone de registros completos y el agente debe adaptarse continuamente a las circunstancias cambiantes.

c) Pasos para construir un modelo de machine learning

Simplilearn (2022), señala que la creación de un modelo de aprendizaje automático es un proceso que suele constar de las siguientes fases y no se limita al uso de algoritmos o bibliotecas de aprendizaje:

- **Recolectar los datos.** La información puede obtenerse de diversas fuentes, como bases de datos, sitios web, API y otros recursos en línea. También podemos utilizar datos que sean de dominio público o utilizar otros dispositivos que recopilen los datos por nosotros. Las posibilidades que tenemos para recopilar datos son infinitas. Aunque este proceso parece sencillo, es uno de los más complicados y de los que más tiempo requieren.
- **Preprocesar los datos.** Tras obtener los datos, debemos asegurarnos que dichos datos sean homogéneos, estén limpios y tienen el formato correcto para poder alimentar nuestro algoritmo inteligente. Antes de poder utilizar los datos, es casi seguro que tendremos que completar una serie de operaciones de preparación. Pero esta parte suele ser mucho más fácil que la anterior.
- **Elegir el modelo.** Un modelo de aprendizaje automático determina el éxito del resultado que se obtiene luego de ejecutar el algoritmo con los datos recopilados. Es importante elegir un modelo relevante para desarrollar dicha tarea en cuestión. Recordemos que existen muchos aportes de modelos inteligentes significativos por parte de científicos e ingenieros para diferentes tareas como reconocimiento de voz, reconocimiento de imágenes, predicción, etc. Aparte de esto, también debe ver si su modelo es adecuado para datos numéricos o categóricos y elegir en consecuencia.

- **Entrenar el algoritmo.** La capacitación del entrenamiento, se basa en preparar los datos para el algoritmo inteligente a fin encontrar patrones y realizar pronósticos, dando como resultado que el modelo aprenda de los datos para que pueda realizar el conjunto de tareas. La idea es que, a medida que pase el tiempo, los algoritmos sean capaces de "aprender" información de los datos que proporcionamos, lo que permitirá al modelo hacer predicciones más precisas.
- **Evaluar el algoritmo.** En este punto, evaluamos el rendimiento del modelo utilizando un nuevo conjunto de datos que no se utilizó durante el entrenamiento. Esto es importante, ya que, si se utilizara el mismo conjunto de datos, el modelo se limitaría a repetir patrones aprendidos previamente, lo que daría lugar a una precisión engañosamente alta. Podemos determinar el rendimiento del modelo en escenarios reales y su capacidad de generalización con mayor precisión probándolo con datos que no se hayan visto antes.
- **Ajuste de parámetros.** Luego de la creación y evaluación del modelo, se debe verificar la precisión, esto se hace ajustando los parámetros en el modelo. Los parámetros son las variables en el modelo que el programador generalmente decide, es un valor particular que se debe ajustar a fin de obtener la precisión máxima.
- **Utilizar el modelo.** Por último, el modelo se aplica al problema real, lo que permite evaluar su eficacia en un entorno real. Es posible que haya que revisar y modificar los procesos anteriores a la luz de los resultados.

d) Teoría de algoritmos de aprendizaje automático

Árboles de decisión

Las tareas de regresión y clasificación son algunas de las muchas funciones que pueden ejecutar los árboles de decisión. Se trata de algoritmos extremadamente potentes que pueden adaptarse a conjuntos de datos modernos y complicados (Géron, 2019); según Hajjej et al. (2022) el objetivo principal del algoritmo es el aprendizaje inductivo basado en la observación y el razonamiento lógico, se utiliza el árbol para simbolizar la información del proceso de aprendizaje inductivo; los árboles se pueden representar mediante nudos, hojas y ramas en una representación visual, cuya clasificación comienza con un nodo raíz que representa el atributo del que se restan todos los demás atributos; asimismo (Mienye et al., 2019), describe que una vez construido el árbol de decisión se utilizan para clasificar los datos de prueba; manifiesta que el conjunto de datos diseñado para entrenar el árbol de decisión sirve como entrada para

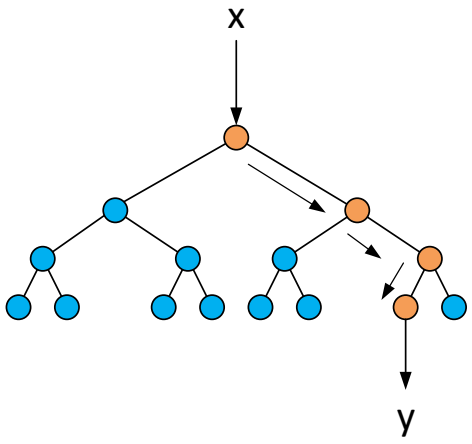
los algoritmos y se compone de objetos y varios atributos; según, Kaun et al. (2021) el algoritmo del árbol de decisión tiene un gran potencial para clasificar la calidad del conocimiento.

Russell y Norvig (2008), una de las mejores técnicas para crear algoritmos de aprendizaje, según su libro, es el árbol de decisión. Este método utiliza como entrada un conjunto de atributos para caracterizar un objeto o circunstancia y, a continuación, emite una "decisión", es decir, un valor predicho basado en la entrada. Aunque en este contexto nos centraremos en las entradas discretas, las cualidades de entrada pueden ser continuas o discretas. El procedimiento se denomina clasificación cuando se aprende una función de valores discretos y regresión cuando se trabaja con una función continua. El resultado puede ser discreto o continuo, en referencia a las clasificaciones booleanas será un punto en discusión debido a que cada ejemplo se clasifica como verdadero o falso.

Para llegar a una decisión, un árbol de decisión crea una secuencia de pruebas sucesivas (véase la Figura 4). Las ramas que se extienden desde cada nodo interno del árbol están etiquetadas con los valores potenciales de cada propiedad, y cada nodo interno corresponde a una prueba basada en el valor de una de las propiedades. El valor que se devolverá en caso de que se alcance un nodo específico está representado por los nodos hoja del árbol. La gente encuentra que los árboles de decisión son bastante intuitivos de usar; de hecho, muchos cientos de páginas de instrucciones que instruyen sobre cómo hacer tareas específicas, como la reparación de un vehículo, están organizadas en su totalidad como árboles de decisión.

Figura 4

Un árbol de decisión para decidir una opción



Nota: Comportamiento de la toma de decisiones

Por lo tanto, según (Fu et al., 2023) la **definición conceptual** del algoritmo de Árboles de Decisión es un método de aprendizaje automático supervisado utilizado para la clasificación y regresión. Con este método, las decisiones y sus posibles resultados se representan mediante un modelo de árbol que adopta la forma de una estructura arbórea. En el árbol, cada nodo interno representa un atributo o característica, cada rama una regla de decisión y cada hoja el resultado o previsión.

Dado que el árbol de decisión que se produce se interpreta y muestra de forma tan sencilla, el método de los árboles de decisión es muy conocido por su capacidad de aprendizaje interpretable. También puede utilizarse para resolver problemas de regresión y clasificación, y puede manejar una gran variedad de tipos de variables.

Asimismo, el algoritmo desarrolla dos fases para realizar predicciones:

- **Aprendizaje:** El aprendizaje del algoritmo de Árboles de Decisión implica construir el árbol de decisión utilizando un conjunto de datos de entrenamiento. El proceso se puede dividir en varios pasos:
    - **a) Selección de atributos:** Se eligen los atributos más relevantes o informativos del conjunto de datos para construir el árbol. Esto se puede hacer utilizando diversas técnicas, como la ganancia de información o el índice Gini.
    - **b) División de nodos**: El árbol se construye dividiendo los nodos en función de los atributos seleccionados. Cada división se basa en una regla de decisión, que puede ser una comparación de valores o una condición lógica.
    - **c) Criterio de parada:** La construcción recursiva del árbol se lleva a cabo hasta que se cumple un criterio de parada, como la pureza de clase, el número mínimo de instancias en un nodo o la profundidad máxima.
    - **d) Poda del árbol:** Una vez construido el árbol, puede aplicarse una técnica de poda para evitar el sobreajuste. La poda implica la eliminación de nodos o ramas que no mejoran significativamente la precisión del árbol.
- **Asertividad**: establecer asertividad en la predicción del algoritmo de Árboles de Decisión se refiere a la precisión o exactitud de las predicciones realizadas por el modelo de árbol. Es una medida de qué tan bien el árbol puede clasificar o predecir correctamente nuevas instancias o ejemplos que no se utilizaron durante el proceso de entrenamiento. Por ello asertividad en la predicción puede evaluarse utilizando diversas métricas, dependiendo del tipo de problema, para ello la clasificación incluyen el grado de precisión (Charbuty y Abdulazeez, 2021).

Naive Bayes

Según Ilić et al. (2022), explica que, aunque los algoritmos típicos de aprendizaje automático se basan en datos fiables, el clasificador Naïve Bayes (NB) funciona bien con datos poco claros, asimismo menciona que el modelo NB es muy popular y efectivo en muchos problemas complejos a pesar de su simplicidad, lo cual es una gran ventaja para su uso; el modelo pertenece a la familia de clasificadores probabilísticos basados en la Teoría de Bayes, cuya característica principal de este clasificador es la suposición de que todas las variables son condicionalmente independientes, razón por la cual se llama 'Naïve'; el uso del algoritmo de aprendizaje automático Naive Bayes está centrado en el desarrollo de problemas de clasificación (Mandal y Jana, 2019).

Villalba (2018) destaca que el modelo es sencillo, robusto y cuya conclusión se deriva directamente de los datos. Su tratamiento se basa en un estadístico bayesiano básico de probabilidad condicional (Barus, 2021), y Villalba destaca que se asume por simplificación que las variables son todas eventos independientes.

Prabhakaran(2018), explica que una técnica probabilística de aprendizaje automático NB puede aplicarse a una serie de aplicaciones de categorización. Son comunes aplicaciones como el análisis de sentimientos, la clasificación de documentos, el filtrado de spam y otras. El nombre proviene del hecho de que se basa en los escritos del reverendo Thomas Bayes. Como se supone que las características que entran en el modelo son independientes entre sí, se utiliza el término "ingenuo". En otras palabras, alterar el valor de una característica no tiene ningún impacto o efecto directo sobre los valores de las demás características que utiliza el algoritmo.

Se trata de un modelo probabilístico con un método fácilmente codificable que permite realizar pronosticos a la velocidad del rayo, incluso en tiempo real. Su rápida capacidad de respuesta lo convierte en una solución altamente escalable, y desde hace tiempo es el algoritmo de referencia para aplicaciones del mundo real en las que es esencial generar respuestas instantáneas a las peticiones de los usuarios.

Mohammed et al. (2017) menciona que los clasificadores bayesianos ingenuos son clasificadores probabilísticos simples basado en aplicar el teorema de Bayes con el supuesto de una (ingenua) independencia entre sus atributos. La siguiente ecuación establece el teorema de Bayes en términos matemáticos:

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$$

dónde:

A y B son eventos

P (A) y P (B) Probabilidad de A y B

P (A | B), Probabilidad posterior de que se dé A dado B

P (B | A), Probabilidad posterior de que se dé B dado A

Roman (2019) menciona que este tipo de algoritmo se conoce como una clase específica de algoritmos de clasificación de aprendizaje automático. Se basan en el método de clasificación estadística del "teorema de Bayes". Debido a su suposición simplista (es decir, que todas las variables predictoras son independientes entre sí), estos modelos suelen denominarse algoritmos "ingenuos". Dicho de otro modo, su suposición subyacente es que no existe correlación entre la existencia de una determinada característica en un conjunto de datos y la presencia de cualquier otra característica. A pesar de esta simplificación, estos algoritmos destacan por su facilidad de construcción de modelos y su excelente rendimiento, que puede atribuirse a su sencillez. El experto Brownlee (2018) en inteligencia artificial, destaca la utilización de usar la probabilidad para hacer predicciones. Quizás el ejemplo más utilizado se llama algoritmo Naive Bayes. No solo es fácil de entender, sino que también logra resultados sorprendentemente buenos.

Por ello según Barus (2021), lo ***define conceptualmente*** al algoritmo Naive Bayes como el método de aprendizaje automático supervisado basado en el teorema de Bayes. Aunque el algoritmo Naive Bayes también puede aplicarse a tareas de regresión, su uso principal se centra en problemas de clasificación. La base de este algoritmo radica en el concepto de independencia condicional de los atributos, el cual significa que, dado un valor de clase específico, cada atributo se considera independiente de los demás. Utilizando la regla de Bayes, Naive Bayes calcula la probabilidad de que una instancia pertenezca a una clase particular. Para lograr esto, el algoritmo estima las probabilidades condicionales y las probabilidades a priori de diversas clases y atributos dentro del conjunto de entrenamiento, permitiéndole asignar con eficacia una clase a cada nueva instancia basada en los datos disponibles

Luego, utiliza estas probabilidades para realizar predicciones sobre nuevas instancias, seleccionando la clase con la probabilidad más alta.

Asimismo, el algoritmo establece indicadores de aprendizaje y asertividad:

- **Aprendizaje,** Utilizando como base un conjunto de datos de entrenamiento, el algoritmo Naive Bayes estima las probabilidades a priori y condicionales de los atributos y las clases. El proceso se puede dividir en los siguientes pasos (Parth Shukla, 2023):
    a) **Recopilación de datos:** Se recopila un conjunto de datos de entrenamiento que consta de instancias etiquetadas con sus respectivas clases.

b) **Estimación de probabilidades a priori:** Se calculan las probabilidades a priori de las clases, es decir, la probabilidad de que un caso elegido al azar pertenezca a una clase determinada.

c) **Estimación de probabilidades condicionales:** Se calculan las probabilidades condicionales de los atributos dados los valores de la clase. En el caso del Naive Bayes, se asume independencia condicional entre los atributos, por lo que estas probabilidades se pueden estimar de manera independiente para cada atributo.

d) **Predicción de nuevas instancias:** Una vez que se han estimado las probabilidades, se utilizan para realizar predicciones sobre nuevas instancias. Se calcula la probabilidad posterior de cada clase para una instancia dada y se selecciona la clase con la probabilidad más alta como la predicción.

- **Asertividad,** Se refiere a la precisión o exactitud de los pronósticos realizados por el modelo. Es una medida de qué tan bien el algoritmo puede clasificar correctamente nuevas instancias o ejemplos que no se utilizaron durante el proceso de entrenamiento. Asertividad en la predicción puede evaluarse utilizando la métrica de precisión dependiendo del problema de clasificación específico y del umbral de decisión seleccionado (Vadlamudi et al., 2023).

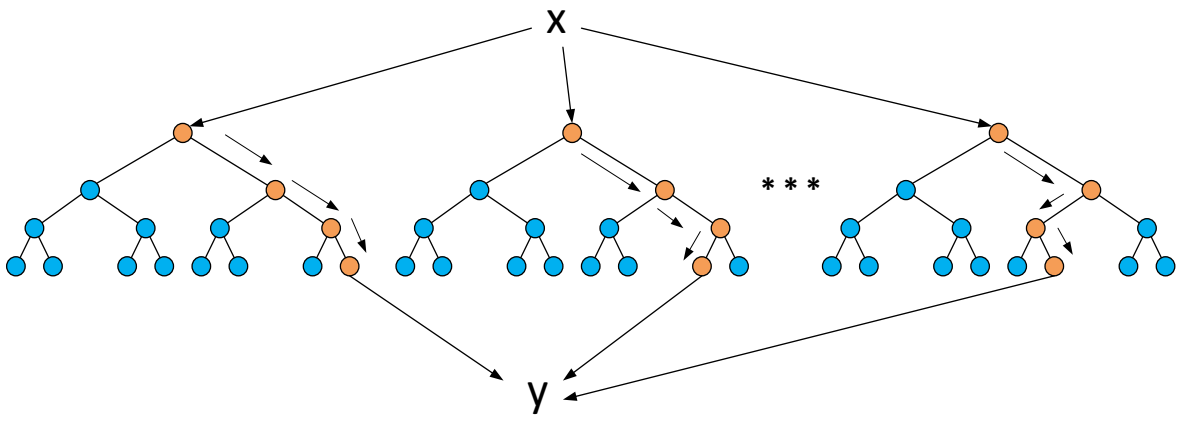
Random Forest

(Pellegrino et al., 2021) explica que Random forest (RF) es un algoritmo ML que se define por un conjunto de métodos de aprendizaje para clasificación, regresión y otras tareas, opera mediante varios árboles de decisión (Pirneskoski et al., 2020), en la fase de entrenamiento y generando las clases (problema de clasificación) o predicción media (problema de regresión) de los árboles individuales; el algoritmo construye cada uno de sus árboles de decisión, entrenándolos a todos con un subconjunto de los datos del problema (Figura 5), elige aleatoriamente N registros de los datos y aprende en múltiples árboles que se construyen y entrenan aleatoriamente de acuerdo con el principio de embolsado (Ong et al., 2022), asimismo Pellegrino asegura que todos los árboles de decisión tengan una experiencia diferente del problema. Una vez entrenados todos los árboles de decisión, RF toma sus decisiones de acuerdo con el problema de clasificación o regresión a resolver votando todos sus árboles de decisión. La aleatoriedad RF se encuentra en el muestreo sobre el que se construyen los árboles

y en la selección de variables sobre las que se realiza la segmentación. Es por eso que RF es una caja negra en muchos casos.

Figura 5

Random Forest



Nota: Comportamiento de la toma de decisiones

(Schonlau y Zou, 2020) conceptualmente, el algoritmo de aprendizaje automático Random Forest es una técnica de aprendizaje supervisado que puede utilizarse tanto para aplicaciones de regresión como de clasificación. Las predicciones pueden realizarse con mayor precisión y resistencia gracias a su funcionamiento, que combina varios árboles de decisión independientes.

Asimismo, el algoritmo establece dos componentes de desarrollo:

- **Aprendizaje**: se refiere al método mediante el cual el modelo Random Forest aprende de los datos entrenados a fin de establecer pronósticos cada vez más precisas y ampliamente aplicables. Según, Kaur Arora (2022), el algoritmo de aprendizaje automático Random Forest se desarrolla de los siguientes pasos:
  - **a)** **Preparación de los datos:** Al igual que los demás algoritmos inteligentes, se requiere preparar los datos de entrenamiento, lo cual incluye la selección y transformación de los atributos relevantes y la codificación adecuada de las variables categóricas.
  - **b)** **Construcción de los árboles de decisión:** Basándose en los datos de entrenamiento, Random Forest crea una colección de árboles de decisión. Para

crear cada árbol se utiliza un método denominado "muestreo bootstrap". Consiste en seleccionar muestras al azar del conjunto de entrenamiento y sustituirlas. Además, en cada nodo del árbol, se selecciona un subconjunto aleatorio de atributos para dividir los datos.

c) **Votación o promedio de resultados:** Una vez que se han construido todos los árboles, Random Forest realiza la predicción combinando los resultados de cada árbol. En el caso de tareas de clasificación, se utiliza un esquema de votación para determinar la clase final, donde cada árbol emite un voto y la clase más votada se selecciona como predicción. En el caso de tareas de regresión, se toma el promedio de los valores predichos por cada árbol.

- **Asertividad**: Se refiere de la precisión y fiabilidad de las predicciones del modelo en el pronóstico del algoritmo Random Forest. Debido a que Random Forest combina múltiples árboles de decisión, que a su vez aprenden de diferentes subconjuntos de datos y atributos, tiende a tener una mayor capacidad predictiva en comparación con un solo árbol de decisión. La combinación de predicciones de diferentes árboles a través de votación o promedio ayuda a reducir el sesgo y la varianza, lo que puede conducir a un mayor asertividad en las predicciones (Andrade y Freitas, 2023.

Definición de términos básicos

Productividad educativa

Cecilia et al. (2020) menciona que la productividad educativa es la medida en que los estudiantes logran sus metas educativas a corto o largo plazo. Esto se mide comúnmente a través de un examen externo o interno, así como una evaluación continua en forma de pruebas, asignaciones, proyectos, debates y trabajos prácticos y trimestrales.

Modelo de Productividad Educativa de Walberg:

Es un marco teórico que identifica y analiza múltiples variables que afectan el rendimiento académico y la calidad de la educación. Walberg propone que la productividad educativa se ve influida por factores como el tiempo de instrucción, la calidad del personal docente, el clima escolar, la motivación del estudiante, los recursos educativos y otros elementos que interactúan para determinar el éxito educativo (Zambrano, 2022).

Capacidad del estudiante

Destaca la capacidad del estudiante es un contexto amplio de habilidades cognitivas, emocionales y sociales que influyen en la capacidad de aprender y desarrollarse en contextos educativos y más allá. También abarca la disposición para aprender, la motivación, capacidad para resolver problemas, inventiva, flexibilidad y otros rasgos que favorecen el éxito en los entornos de aprendizaje y en la vida en general (Figueroa-Abarzúa et al., 2022).

Aptitud del estudiante

Es un concepto integral que engloba las habilidades, actitudes y disposiciones que contribuyen al desempeño académico y al desarrollo personal del estudiante. Es importante tener en cuenta que la aptitud puede ser moldeada y desarrollada a lo largo del tiempo mediante experiencias educativas, prácticas y esfuerzos individuales (Herrera y Sánchez, 2019).

Motivación

La motivación del estudiante implica la fuerza interna que impulsa al individuo a comprometerse y esforzarse en sus estudios, superar desafíos y persistir en la consecución de sus objetivos educativos. Este concepto reconoce que la motivación no es uniforme para todos los estudiantes y puede variar según factores individuales, contextuales y situacionales (Avila, 2022).

Desarrollo

El desarrollo cognitivo implica cómo los individuos perciben, procesan y utilizan la información del entorno para comprender el mundo que les rodea y adaptarse a él. Este concepto está estrechamente relacionado con la maduración del cerebro, la interacción con el entorno y las experiencias de aprendizaje a lo largo de la vida (Ahmed Khan et al., 2023).

Ambiente de aprendizaje

El término "entorno de aprendizaje" se refiere a un concepto más amplio que incluye factores sociales, emocionales y culturales, además de la disposición física del aula. La relación profesor-alumno, la cooperación entre compañeros, la accesibilidad a los materiales didácticos, la flexibilidad del plan de estudios, el uso de la tecnología y la atmósfera emocional del aula son componentes de este entorno.(Castañon, 2023).

Rendimiento académico

Desde un punto de vista conceptual, el rendimiento académico es más amplio que las notas que se obtienen en las evaluaciones, los exámenes o las valoraciones formales, aunque son indicadores importantes. También incluye elementos adicionales que muestran lo bien que los estudiantes aprovechan las oportunidades de aprendizaje, como la creatividad, la capacidad de

resolución de problemas, la capacidad de análisis y síntesis y la participación activa en clase (Grasso Imig, 2020).

Datos

Es una representación de hechos, conceptos o instrucciones en forma formalizada, apta para su comunicación, interpretación o procesamiento, asimismo es una información creada por un usuario, como documentos, imágenes o grabaciones de sonido (Rigdon, 2016).

Dataset o colección de datos

Es una colección coherente de datos con criterios de selección y estructura interna bien definido (Oxford University, 2016); tambien es definido como una ccolección de información relacionada formada por elementos separados que pueden ser tratados como una unidad en el manejo de datos (Rigdon, 2016).

Información

Según Oxford University (2018) la información es un conjunto de datos con significado que es capaz de hacer que la mente humana cambie su opinión sobre el estado actual del mundo real. En ciencia e ingeniería, información es todo lo que contribuye a reducir la incertidumbre del estado de un sistema; en este caso, la incertidumbre suele expresarse de forma objetivamente medible. Es importante mencionar que la información debe distinguirse de cualquier medio que sea capaz de transportarla. Un medio físico (como un disco magnético) puede transportar un medio lógico (datos, como símbolos binarios o de texto). El contenido de información de cualquier objeto físico, o datos lógicos, no puede medirse ni discutirse hasta que se sepa qué rango de posibilidades existía antes y después de que fueran recibidos.

La información radica en la reducción de la incertidumbre resultante de la recepción de los objetos o datos, y no en el tamaño o complejidad de los objetos o datos en sí. Las preguntas sobre la forma, la función y la importancia semántica de los datos solo son relevantes para la información en la medida en que contribuyan a la reducción de la incertidumbre. Si un memorándum idéntico se recibe dos veces, no transmite el doble de información que la primera vez: la segunda no transmite ninguna información, a menos que, por acuerdo previo, el número de ocurrencias se considere significativo.

Patrón

Laplante (2001) menciona que es una regla que especifica clases de equivalencia de secuencias textuales. Un patrón a menudo se expresa en una notación como una expresión regular o alguna notación equivalente. Los lenguajes como SNOBOL tenían especificaciones de patrones elaborados que incluían reglas para volver a intentar una coincidencia después de un

intento fallido ("reglas de retroceso"). Los lenguajes modernos que incluyen patrones de coincidencia de cadenas como operaciones primitivas incluyen awk y Perl.

Algoritmo

Según, Oxford University (2018) un algoritmo es una colección predeterminada de pautas o directivas creadas para completar una tarea, como completar un cálculo, en una cantidad limitada de pasos. Un componente clave de la programación automatizada es la notación formal de algoritmos; muchas ideas que se aplican en programas se aplican en algoritmos y en forma contraria. Cuando un algoritmo se puede calcular con éxito, se dice que es eficiente. La base de la teoría de algoritmos es la investigación de la existencia de algoritmos capaces de calcular determinadas cantidades. Puede resultar difícil demostrar la solidez de un algoritmo o incluso determinar su objetivo preciso. La validación de algoritmos, un procedimiento que confirma que un algoritmo cumple sus funciones, se utiliza con frecuencia en la práctica. Esto involucra poner a prueba el algoritmo en una serie de escenarios para garantizar su funcionamiento adecuado en cada uno de ellos. Se puede estar seguro de que el algoritmo es fiable eligiendo un conjunto de pruebas adecuado.

Machine learning (ml)

Raynor (999) define a Machine Learning como la capacidad de un programa para adquirir o desarrollar nuevos conocimientos o habilidades. El estudio de Machine Learning se centra en el desarrollo de métodos de cálculo para descubrir nuevos conocimientos a partir de datos.

BBVA (2019) define como el área de la inteligencia artificial que se ocupa del aprendizaje automático sin necesidad de programación especial. una característica crucial que permite a los sistemas reconocer patrones en los datos para poder generar predicciones. Numerosas aplicaciones, como los contenidos sugeridos de Netflix o Spotify, las respuestas inteligentes de Gmail o el reconocimiento de voz de Siri y Alexa, utilizan esta técnica.

Aprendizaje supervisado

Un subcampo de la inteligencia artificial y el aprendizaje automático se denomina aprendizaje supervisado o aprendizaje automático supervisado. Se distingue por el uso de conjuntos de datos etiquetados para entrenar algoritmos de clasificación precisa de datos o predicción de resultados. A lo largo de este procedimiento, el modelo aprende de los datos que recibe y modifica sus ponderaciones según sea necesario para garantizar un buen ajuste del modelo. Las organizaciones pueden utilizar el aprendizaje supervisado para resolver una serie de problemas complejos del mundo real, como clasificar automáticamente los correos electrónicos no solicitados en una carpeta diferente de la bandeja de entrada (IBM Cloud Education, 2020).

Aprendizaje no supervisado

El aprendizaje no supervisado, usualmente acreditado como aprendizaje automático no supervisado, utiliza algoritmos de aprendizaje automático para agrupar y examinar conjuntos de datos no etiquetados. Sin la ayuda de una persona, estos algoritmos pueden encontrar patrones ocultos o agrupar datos. Por su capacidad para identificar patrones en los datos, es perfecto para actividades como la ejecución de estrategias de venta cruzada, el reconocimiento de imágenes, la segmentación de clientes y el análisis exploratorio de datos (IBM Cloud Education, 2020)**.**

Clasificación

Predecir a qué clase pertenece un conjunto concreto de puntos de datos es el proceso de clasificación. Estas clases también pueden denominarse etiquetas, categorías u objetivos. El proceso de aproximación de una función de mapeo (f) que conecta variables de salida discretas (Y) con variables de entrada (X) se conoce como modelización de clasificación predictiva. (Asiri, 2018)**.**

Laplante (2001) menciona que es el proceso de encontrar reglas de clasificación, adicionalmente menciona que es un área de minería de datos que intenta predecir la categoría de datos categóricos mediante la construcción de un modelo basado en algunas variables predictoras.

Referencias

Abid, N. y Ali, R. y Akhter, M. (2021). Exploring gender-based difference towards academic enablers scales among secondary school students of Pakistan. *Psychology in the Schools*, *58*(7), 1380–1398. https://doi.org/10.1002/pits.22538

Ahmed Khan, Z. y Adnan, J. y Adnan Raza, S. (2023). *Cognitive Learning Theory and Development: Higher Education Case Study*. https://doi.org/10.5772/intechopen.110629

Al-Ani, O. y Das, S. (2022). Reinforcement Learning: Theory and Applications in HEMS. *MATHEMATICS & COMPUTER SCIENCE, Artificial Intelligence & Robotics*.

Andrade, M. S. y Freitas, J. C. de. (2023). Analysis of Performance Metrics on the Conjuncture of Intrusions in IEEE 802.11 Networks with Machine Learning at Hospital N.S.C. In *Connecting Expertise Multidisciplinary Development For The Future*. Seven Editora. https://doi.org/10.56238/Connexpemultidisdevolpfut-116

Avila, Á. (2022). *La motivación en los estudiantes: 3 Claves para aumentarla*. https://www.enriccorberainstitute.com/blog/motivacion-en-los-estudiantes/

Bagnato Juan Ignacio. (2020). *Aprende Machine Learning en Español Teoría + Práctica Python*. 1–184.

Barus, S. P. (2021). Implementation of Naïve Bayes Classifier-based Machine Learning to Predict and Classify New Students at Matana University. *Journal of Physics: Conference Series*, *1842*(1), 012008. https://doi.org/10.1088/1742-6596/1842/1/012008

BBVA. (2019). *"Machine learning": ¿qué es y cómo funciona?*

Brownlee, J. (2018). *Machine Learning Algorithms From Scratch* (v.1.8, Ed.).

Carroll, J. B. (1989). The Carroll Model: A 25-Year Retrospective and Prospective View. *Educational Researcher*, *18*(1), 26. https://doi.org/10.2307/1176007

Caselli Gismondi, H. E. (2021). *Modelo predictivo basado en machine learning como soporte para el seguimiento académico del estudiante universitario* [Universidad Nacional del Santa]. http://repositorio.uns.edu.pe/bitstream/handle/UNS/3804/52337.pdf?sequence=5&isAllowed=y

Castañon, E. R. (2023). *¿Qué es un ambiente de aprendizaje según autores?* https://www.centrobanamex.com.mx/que-es-un-ambiente-de-aprendizaje-segun-autores/

Cecilia, O. N. y Bernedette, C.-U. y Emmanuel, A. E. y Hope, A. N. (2020). Enhancing students academic performance in Chemistry by using kitchen resources in Ikom, Calabar. *Educational Research and Reviews*, *15*(1), 19–26. https://doi.org/10.5897/ERR2019.3810

Charbuty, B. y Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Díaz-Landa, B. y Meleán-Romero, R. y Marín-Rodriguez, W. (2021). Rendimiento académico de estudiantes en Educación Superior: predicciones de factores influyentes a partir de árboles de decisión. *Telos Revista de Estudios Interdisciplinarios En Ciencias Sociales*, *23*(3), 616–639. https://doi.org/10.36390/telos233.08

Dong, Y. y Hou, J. y Zhang, N. y Zhang, M. (2020). Research on How Human Intelligence, Consciousness, and Cognitive Computing Affect the Development of Artificial Intelligence. *Complexity*, *2020*, 1–10. https://doi.org/10.1155/2020/1680845

Esteso, A. y Peidro, D. y Mula, J. y Díaz-Madroñero, M. (2022). Reinforcement learning applied to production planning and control. *International Journal of Production Research*, 1–18. https://doi.org/10.1080/00207543.2022.2104180

Figueroa-Abarzúa, C. y Meza-Vásquez, S. y Estrada-Lagos, R. (2022). Desarrollo de capacidad argumentativa en estudiantes universitarios, mediante uso del debate como estrategia didáctica. *South Florida Journal of Development*, *3*(6), 6328–6346. https://doi.org/10.46932/sfjdv3n6-001

Fu, M. y Zhang, C. y Hu, C. y Wu, T. y Dong, J. y Zhu, L. (2023). Achieving Verifiable Decision Tree Prediction on Hybrid Blockchains. *Entropy*, *25*(7), 1058. https://doi.org/10.3390/e25071058

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent System* (Ni. Tache, Ed.; 2nd ed.).

Gogus, A. (2012). Bloom's Taxonomy of Learning Objectives. In *Encyclopedia of the Sciences of Learning* (pp. 469–473). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_141

Grasso Imig, P. (2020). *Rendimiento académico: un recorrido conceptual que aproxima a una definición unificada para el ámbito superior*. https://fh.mdp.edu.ar/revistas/index.php/r_educ/article/view/4165

Guill, K. y Ömeroğulları, M. y Köller, O. (2021). Intensity and content of private tutoring lessons during German secondary schooling: effects on students' grades and test achievement. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-021-00581-x

Hajjej, F. y Alohali, M. A. y Badr, M. y Rahman, M. A. (2022). A Comparison of Decision Tree Algorithms in the Assessment of Biomedical Data. *BioMed Research International*, *2022*, 1–9. https://doi.org/10.1155/2022/9449497

Herrera Acosta, C. E. y Sánchez Pinedo, L. D. (2019). Metodología De Aprendizaje Y Sistema De Evaluación Para Alcanzar Resultados En El Proceso Educativo. *Evaluación de La Calidad Educativa*. https://www.igobernanza.org/index.php/IGOB/Contacto

IBM Cloud Education. (2020). *¿Qué es el aprendizaje supervisado?* IBM. https://www.ibm.com/es-es/think/topics/supervised-learning

Ilić, M. y Srdjević, Z. y Srdjević, B. (2022). Water quality prediction based on Naïve Bayes algorithm. *Water Science and Technology*, *85*(4), 1027–1039. https://doi.org/10.2166/wst.2022.006

Janiesch, C. y Zschech, P. y Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jiang, T. y Gradus, J. L. y Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, *51*(5), 675–687. https://doi.org/10.1016/j.beth.2020.05.002

Jovel, J. y Greiner, R. (2021). An Introduction to Machine Learning Approaches for Biomedical Research. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.771607

Kaun, C. y Jhanjhi, N. Z. y Goh, W. W. y Sukumaran, S. (2021). Implementation of Decision Tree Algorithm to Classify Knowledge Quality in a Knowledge Intensive System. *MATEC Web of Conferences*, *335*, 04002. https://doi.org/10.1051/matecconf/202133504002

Kaur Arora, S. (2022). *Top Steps To Learn Naive Bayes Algorithm*. https://hackr.io/blog/top-steps-to-learn-naive-bayes-algorithm

KOÇOĞLU, F. Ö. (2022). Research on the success of unsupervised learning algorithms in indoor location prediction. *International Advanced Researches and Engineering Journal*, *6*(2 (under construction)), 148–153. https://doi.org/10.35860/iarej.1096573

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. In *Theory Into Practice* (Vol. 41, Issue 4). https://doi.org/10.1207/s15430421tip4104_2

Lampropoulos, A. S. y Tsihrintzis, G. A. (2015). *Machine Learning Paradigms* (Vol. 92). Springer International Publishing. https://doi.org/10.1007/978-3-319-19135-5

Laplante, P. A. (2001). *Dictionary of computer science, engineering, and technology*. www.crcpress.com

Lee, J. y Warner, E. y Shaikhouni, S. y Bitzer, M. y Kretzler, M. y Gipson, D. y Pennathur, S. y Bellovich, K. y Bhat, Z. y Gadegbeku, C. y Massengill, S. y Perumal, K. y Saha, J. y Yang, Y. y Luo, J. y Zhang, X. y Mariani, L. y Hodgin, J. B. y Rao, A. (2022). Unsupervised machine learning for identifying important visual features through bag-of-words using histopathology data from chronic kidney disease. *Scientific Reports*, *12*(1), 4832. https://doi.org/10.1038/s41598-022-08974-8

Mandal, L. y Jana, N. D. (2019). A Comparative Study of Naive Bayes and k-NN Algorithm for Multi-class Drug Molecule Classification. *2019 IEEE 16th India Council International Conference (INDICON)*, 1–4. https://doi.org/10.1109/INDICON47234.2019.9029095

Mazana, M. Y. y Montero, C. S. y Casmir, R. O. (2019). Investigating Students' Attitude towards Learning Mathematics. *International Electronic Journal of Mathematics Education*, *14*(1). https://doi.org/10.29333/iejme/3997

Mienye, I. D. y Sun, Y. y Wang, Z. (2019). Prediction performance of improved decision tree-based algorithms: a review. *Procedia Manufacturing*, *35*, 698–703. https://doi.org/10.1016/j.promfg.2019.06.011

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill. https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf

Mohammed, M. y Khan, M. B. y Mohammed Bashier, E. B. (2017). *Machine Learning Algorithms and Applications*.

Morales Hernández, M. Á. y González Camacho, J. M. y Robles Vásquez, H. y Del Valle Paniagua, D. H. y Durán Moreno, J. R. (2022). Algoritmos de aprendizaje automático para la predicción del logro académico. *RIDE Revista Iberoamericana Para La Investigación y El Desarrollo Educativo*, *12*(24). https://doi.org/10.23913/ride.v12i24.1180

Münch, M. y Raab, C. y Biehl, M. y Schleif, F.-M. (2020). Data-Driven Supervised Learning for Life Science Data. *Frontiers in Applied Mathematics and Statistics*, *6*. https://doi.org/10.3389/fams.2020.553000

Nedeva, V. y Pehlivanova, T. (2021). Students' Performance Analyses Using Machine Learning Algorithms in WEKA. *IOP Conference Series: Materials Science and Engineering*, *1031*(1), 012061. https://doi.org/10.1088/1757-899X/1031/1/012061

Ong, A. K. S. y Chuenyindee, T. y Prasetyo, Y. T. y Nadlifatin, R. y Persada, S. F. y Gumasing, Ma. J. J. y German, J. D. y Robas, K. P. E. y Young, M. N. y Sittiwatethanasiri, T. (2022). Utilization of Random Forest and Deep Learning Neural Network for Predicting Factors Affecting Perceived Usability of a COVID-19 Contact Tracing Mobile Application in Thailand "ThaiChana." *International Journal of Environmental Research and Public Health*, *19*(10), 6111. https://doi.org/10.3390/ijerph19106111

Oxford University. (2016). *A Dictionary of Computer Science*.

Parhizkar, A. y Tejeddin, G. y Khatibi, T. (2023). Student performance prediction using datamining classification algorithms: Evaluating generalizability of models from geographical aspect. *Education and Information Technologies*, *28*(11), 14167–14185. https://doi.org/10.1007/s10639-022-11560-0

Parth Shukla. (2023). *Naive Bayes Algorithms: A Complete Guide for Beginners*. https://www.analyticsvidhya.com/blog/2023/01/naive-bayes-algorithms-a-complete-guide-for-beginners/

Peconcillo Jr, L. B. y D. Peteros, E. y O. Mamites, I. y T. Sanchez, D. y L. Tenerife, J. J. y L. Suson, R. (2020). Structuring Determinants to Level Up Students Performance. *International Journal of Education and Practice*, *8*(4), 638–651. https://doi.org/10.18488/journal.61.2020.84.638.651

Pellegrino, E. y Jacques, C. y Beaufils, N. y Nanni, I. y Carlioz, A. y Metellus, P. y Ouafik, L. (2021). Machine learning random forest for predicting oncosomatic variant NGS analysis. *Scientific Reports*, *11*(1), 21820. https://doi.org/10.1038/s41598-021-01253-y

Pirneskoski, J. y Tamminen, J. y Kallonen, A. y Nurmi, J. y Kuisma, M. y Olkkola, K. T. y Hoppu, S. (2020). Random forest machine learning method outperforms prehospital National Early Warning Score for predicting one-day mortality: A retrospective study. *Resuscitation Plus*, *4*, 100046. https://doi.org/10.1016/j.resplu.2020.100046

Prabhakaran, S. (2018). *How Naive Bayes Algorithm Works?* Machine Learning. https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/

Pugliese, R. y Regondi, S. y Marini, R. (2021). Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Science and Management*, *4*, 19–29. https://doi.org/10.1016/j.dsm.2021.12.002

Quiñones, L. y Quiñones, Y. L. (2020). Rendimiento académico empleando minería de datos. *Espacios*, *41*(44), 277–285. https://doi.org/10.48082/espacios-a20v41n44p17

Raynor, W. J. (1999). *The International Dictionary of Artificial Intelligence*.

Rigdon, J. C. (2016). *Dictionary of Computer and Internet Terms* (Microsoft Corporation, Ed.; 1st Edition). http://www.wordsrus.info

Roman, V. (2019). *Algoritmos Naive Bayes: Fundamentos e Implementación*. Ciencia de Datos. https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fudamentos-e-implementaci%C3%B3n-4bcb24b307f

Rosales Sánchez, E. M. y Rodríguez Ortega, P. G. y Romero Ariza, M. (2020). Conocimiento, demanda cognitiva y contextos en la evaluación de la alfabetización científica en PISA. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, *17*(2), 1–22. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2020.v17.i2.2302

Russell, S. y Norvig, P. (2008). *Inteligencia Artificial Un Enfoque Moderno* (D. F. Aragón, Ed.; 2nd ed.).

Sauce, B. y Liebherr, M. y Judd, N. y Klingberg, T. (2022). The impact of digital media on children's intelligence while controlling for genetic differences in cognition and socioeconomic background. *Scientific Reports*, *12*(1), 7720. https://doi.org/10.1038/s41598-022-11341-2

Schonlau, M. y Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal: Promoting Communications on Statistics and Stata*, *20*(1), 3–29. https://doi.org/10.1177/1536867X20909688

Seel, N. M. (2012a). Bloom's Model of School Learning. In *Encyclopedia of the Sciences of Learning* (pp. 466–469). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_979

Seel, N. M. (2012b). Carroll's Model of School Learning. In *Encyclopedia of the Sciences of Learning* (pp. 501–503). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_980

Sepúlveda, A. y Minte, A. y Díaz-Levicoy, D. (2020). Caracterización de preguntas en libros de texto de Ciencias Naturales en Educación Primaria chilena. *Educação e Pesquisa*, *46*. https://doi.org/10.1590/s1678-4634202046224118

Serumena D. y Utan F. y Poernomo M. (2021). The Effectiveness of Social Media as An Online Learning Pattern in Improving the 3 Domains of Student Intellectual Ability During the Pandemic (Covid-19). *Advances in Engineering Research*.

Simplilearn. (2022). *The Complete Guide to Machine Learning Steps*. https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-steps

Singh, V. y Chen, S.-S. y Singhania, M. y Nanavati, B. y kar, A. kumar y Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda. *International Journal of Information Management Data Insights*, *2*(2), 100094. https://doi.org/10.1016/j.jjimei.2022.100094

Stern, M. y Arinze, C. y Perez, L. y Palmer, S. E. y Murugan, A. (2020). Supervised learning through physical changes in a mechanical system. *Proceedings of the National Academy of Sciences*, *117*(26), 14843–14850. https://doi.org/10.1073/pnas.2000807117

Tai, M.-T. (2020). The impact of artificial intelligence on human society and bioethics. *Tzu Chi Medical Journal*, *32*(4), 339. https://doi.org/10.4103/tcmj.tcmj_71_20

Tebani, A. y Bekri, S. (2020). High-throughput omics in the precision medicine ecosystem. In *Precision Medicine for Investigators, Practitioners and Providers* (pp. 19–31). Elsevier. https://doi.org/10.1016/B978-0-12-819178-1.00003-4

Vadlamudi, P. S. y Gunasekaran, M. y Nagalakshmi, T. J. (2023). An Analysis of the Effectiveness of the Naive Bayes Algorithm and the Support Vector Machine for Detecting Fake News on Social Media. *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, 726–731. https://doi.org/10.1109/IITCEE57236.2023.10090978

Vigo, R. y Zeigler, D. E. y Wimsatt, J. (2022). Uncharted Aspects of Human Intelligence in Knowledge-Based "Intelligent" Systems. *Philosophies*, *7*(3), 46. https://doi.org/10.3390/philosophies7030046

Villalba, F. (2018). Naive Bayes- clasificación bayesiano ingenuo. In *Aprendizaje supervisado en R*. https://fervilber.github.io/Aprendizaje-supervisado-en-R/ingenuo.html

Vostroknutov, A. y Polonio, L. y Coricelli, G. (2018). The Role of Intelligence in Social Learning. *Scientific Reports*, *8*(1), 6896. https://doi.org/10.1038/s41598-018-25289-9

Walberg, H. J. (1984). *Improving the Productivity of America's Schools*.

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Xiang, X. y Foo, S. y Zang, H. (2021). Recent Advances in Deep Reinforcement Learning Applications for Solving Partially Observable Markov Decision Processes (POMDP) Problems Part 2—Applications in Transportation, Industries, Communications and Networking and More Topics. *Machine Learning and Knowledge Extraction*, *3*(4), 863–878. https://doi.org/10.3390/make3040043

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11. https://doi.org/10.1186/s40561-022-00192-z

Yildiz, M. y Börekci, C. (2020). Predicting Academic Achievement with Machine Learning Algorithms. *Journal of Educational Technology and Online Learning*. https://doi.org/10.31681/jetol.773206

Zambrano Aranda, G. (2022). Análisis de la satisfacción con respecto a una maestría de una universidad de Lima desde la perspectiva de sus egresados. *Revista Educación y Sociedad*, *3*(5), 9–22. https://doi.org/10.53940/reys.v3i5.89

Zhao, W. (2022). Inspired, but not mimicking: a conversation between artificial intelligence and human intelligence. *National Science Review*, *9*(6). https://doi.org/10.1093/nsr/nwac068

Zheng, Y. (2019). Identification of microRNAs From Small RNA Sequencing Profiles. In *Computational Non-coding RNA Biology* (pp. 35–82). Elsevier. https://doi.org/10.1016/B978-0-12-814365-0.00012-9