



# Clasificación de estudiantes basada en habilidades cognitivas y Predicción de rendimientos de aprendizaje usando la máquina Modelos de aprendizaje

T. PanduRanga Vital<sup>1, \*</sup>, K. Sangeeta<sup>2</sup> y Kalyana Kiran Kumar<sup>3</sup>

<sup>1</sup>Profesor asociado, Depto. de CSE, Instituto Aditya de Tecnología y Gestión, Tekkali-532 201, Srikakulam (Dt.), AP, India,

<sup>2</sup> Asistente Sr. Profesor, Depto. de CSE, Instituto Aditya de Tecnología y Gestión, Tekkali-532 201, Srikakulam (Dt.), AP, India,

<sup>3</sup>Profesor, Dpto. de EEE, Instituto Aditya de Tecnología y Gestión, Tekkali-532 201, Srikakulam (Dt.), AP, India,

Recibido el 9 de marzo de 2020, revisado el 30 de julio de 2020, aceptado el 13 de noviembre de 2020, publicado el 1 de enero de 2021

**Resumen:** La educación es el parámetro vital del país para el desarrollo en áreas divergentes como el cultivo, económico, político, salud, etc. El objetivo principal de cualquier Instituto educativo (universidades, colegios, escuelas) es aumentar la capacidad de aprendizaje de los estudiantes y sus habilidades para su contribución total a la sociedad. En estos días, el tema de investigación del "proceso de aprendizaje y desarrollo de habilidades de los estudiantes" requiere una atención muy necesaria para el mejoramiento de la sociedad.

El desempeño del estudiante depende de su capacidad de aprendizaje y está influenciado por muchos factores. En este artículo, analizamos las diferentes categorías de inclinaciones de los estudiantes que son muy rápidas, rápidas, moderadas y lentas. Para esto, llevamos a cabo la capacitación y las pruebas y usamos las características de simpatía, nivel de conocimiento, razonamiento y habilidades de las materias básicas para los 313 estudiantes de ingeniería en AITAM, Tekkali, afiliado a JNTUK, India de 2017 a 2019. Recopilamos información sobre el personal, académico, nivel cognitivo y datos demográficos de los estudiantes. En este experimento, estamos realizando un análisis estadístico y clasificando a los estudiantes en 4 tipos de estudiantes y aplicando las diferentes técnicas de aprendizaje automático (ML), y elegimos el mejor algoritmo de ML para predecir las tasas de aprendizaje de los estudiantes. Esto lleva a realizar clases de recuperación con nuevos métodos de enseñanza para estudiantes de aprendizaje moderado y lento. El documento propuesto acomoda las diferencias individuales de los alumnos en términos de nivel de conocimiento, preferencias de aprendizaje, habilidades cognitivas, etc. Para esto, aplicamos 5 algoritmos ML que son Naive Bayes, árboles de clasificación (CT), k-NN, C4.5 y SVM. Según el análisis de ML, el algoritmo k-Nearest Neighborhood (k-NN) es más eficiente que otros algoritmos donde los valores de precisión y predicción están más cerca del 100%.

**Palabras clave:** educación, aprendizaje automático, aprendizaje del estudiante, desempeño del estudiante

## 1. INTRODUCCIÓN

La educación es la columna vertebral de cualquier país o sociedad [16]. Es una de las medidas del desarrollo social, económico y político del país. Este documento utiliza datos de estudiantes en tiempo real del departamento de ingeniería informática, Instituto Aditya de Tecnología y Gestión, Tekkali en el distrito de Srikakulam. El estudio involucra experimentos para comprender la influencia de los atributos cognitivos en el rendimiento académico. Los estudiantes se clasifican en aprendices muy rápidos, aprendices rápidos, aprendices promedio y aprendices lentos utilizando algoritmos de clasificación y descubriendo el mejor modelo de predicción. El documento propuesto acomoda las diferencias individuales de los

estudiantes en términos de nivel de conocimiento, preferencias de aprendizaje, habilidades cognitivas, etc.

La predicción del desempeño de los estudiantes es una tarea desafiante, ya que depende de muchos factores, como las calificaciones, el desempeño en clase, los datos demográficos y las características emocionales. Los docentes necesitan pronosticar el desempeño futuro de un estudiante en base a sus desempeños anteriores, identificando a los estudiantes débiles en una etapa temprana para que se pueda facilitar material adicional y atención especial para evitar el riesgo de fallar.

Más análisis, comparamos con otros trabajos en profundidad.

La Sección 2 brinda descripciones detalladas de otros trabajos de investigación relevantes para el análisis del desempeño de los estudiantes con la máquina.



algoritmos de aprendizaje con diferentes conjuntos de datos con varios puntos de referencia ML. La Sección 3 proporciona el modelo propuesto y los materiales en los que se analizan los diferentes algoritmos de ML. La sección 4 y la sección 5 proporcionan un análisis comparativo detallado de los resultados experimentales y la conclusión del trabajo y las propuestas de trabajo futuro.

## 2. REVISIÓN DE LA LITERATURA

En esta sección, revisamos 120 artículos de revistas de renombre como IEEE, Elsevier y Springer, entre otras.

Se presentan algunos de los trabajos que están muy relacionados con este experimento. Fok et al. [1] aplicó el motor analítico Deep Learning (DL) para evaluar el desempeño del estudiante. El estudio involucra el análisis de la influencia del rendimiento académico y sus actividades extracurriculares como servicios, artes y su conducta.

El experimento logró una precisión que osciló entre el 80 % y el 91 %. Un total de 2000 conjuntos de datos de tamaño se divide en un 75 % de datos de entrenamiento y un 25 % de datos de prueba. El modelo de aprendizaje profundo de flujo de Tensor está configurado de manera óptima para lograr la mayor precisión de predicción. Ma et al. [2] consideró la dependencia entre los atributos de los estudiantes para inicializar los coeficientes de los algoritmos de aprendizaje automático utilizando reglas de coeficientes de inicialización. Esto ayuda a una convergencia más rápida de los algoritmos. En esto, el algoritmo de búsqueda de cuadrícula se ajusta a los algoritmos DT y SVM y obtiene el modelo optimizado.

Sekeroglu et al. [3] en esta clasificación se aplica la regresión de vector de soporte (SVR), la retropropagación, LSTM para la clasificación y predicción del rendimiento de los estudiantes, y también el clasificador de aumento de gradiente. Para ello se utilizan dos conjuntos de datos diferentes SPD [14] y SAPD [15], uno para predicción y otro para clasificación. Cortés et al. [4] usó un conjunto de datos de estudiantes en tiempo real compuesto por estudiantes de instituciones académicas líderes en la India para la predicción de estudiantes

rendimiento. Usaron las características de los estudiantes como CGPA, rendimiento de laboratorio, etc., para clasificarlos en cuatro grupos, como deficiente, promedio, bueno y excelente. El árbol de decisión convencional ha mejorado su funcionalidad a través de funciones de asociación y factores normalizados.

Yu et al. [7] aplicó el análisis de sentimientos sobre comentarios autoevaluados basados en texto dados por los estudiantes para la predicción del desempeño de los estudiantes. Los resultados experimentales muestran que la información de sentimiento de estos comentarios proporciona precisión en la predicción. Ellos consideraron datos estructurados, como finalización de tareas, asistencia y calificaciones de exámenes para su experimento. Soror et al. [8] también aplicó técnicas de minería de textos para predecir el desempeño de los estudiantes. Utilizaron los comentarios proporcionados por los estudiantes y aplicaron los métodos K-means y LSA (análisis semántico latente) para la predicción. Los métodos de medición de superposición y similitud se utilizan junto con LSA y k-means para la mejora. La precisión observada fue del 66,4 % para k-medias y del 78,5 % para el método de superposición relativa. Aziz et al. [9] investigaron bases de datos educativas utilizando técnicas de DM para identificar patrones.

Recolectaron datos en tiempo real de I-B.Tech. estudiantes en CSE y concibió características relacionadas con sus registros académicos, historia familiar y demografía. Aplicaron técnicas de ML como Bayes ingenuo basado en reglas y DT para la predicción del rendimiento de los estudiantes. La técnica de clasificación basada en reglas logró una alta precisión en comparación con los otros dos modelos con un valor de precisión del 71,3 %. Zhang et al. [10] centrado en identificar

estudiantes en riesgo para que se puedan tomar medidas tempranas para aumentar la retención de estudiantes. Aplicaron algoritmos de DM y NLP para observar el desempeño académico de un estudiante.

Mohsin et al. [11] aplicó algoritmos de minería de datos en conjuntos de datos de rendimiento de programación de los estudiantes. El conjunto de datos de capacitación consta del perfil de concierto de los estudiantes universitarios de la Universidad de Utara Malasia de 4 programas distintivos que son licenciatura en TI, Multimedia, Ciencias de la Decisión y Educación en TI del año 2004/2005. Hubo 419 registros con 70 atributos sobre los que se aplicó el algoritmo de minería de reglas de asociación Apriori.

El autor aplicó técnicas de DM en un conjunto de datos de estudiantes de la Universidad de Bulgaria para predecir el rendimiento de los estudiantes. Huang et al. [13] estudió el análisis comparativo de regresión en algoritmos lineales múltiples, la red MLP, la red RBF y SVM para descubrir el mejor modelo para el pronóstico del rendimiento académico de los estudiantes. Los atributos del estudiante considerados son CGPA, notas de medio examen, etc. El resultado de los modelos es la puntuación del estudiante en el examen final. Parak et al. [14] propusieron técnicas de aprendizaje automático para agrupar y crear perfiles de estudiantes. Aplicaron el algoritmo Apriori para perfilar a los estudiantes y encontrar correlaciones entre un conjunto de elementos. K-means se emplea para agrupar a los estudiantes.

El perfil de los estudiantes se completa utilizando el rendimiento de los expedientes académicos, las calificaciones obtenidas en los exámenes parciales y parciales, algunos de los trabajos se describen en la tabla I en detalle.



TABLA I. CONTRIBUCIÓN DE LOS AUTORES EN DIFERENTES TRABAJOS

Árbitro. No.	Autor	Año	Técnicas utilizadas	Contribuciones destacadas
[12]	Kabakchieva et al.,	2013	Técnicas de MD	Proyecto de investigación de DM para la gestión universitaria destinado a revelar el potencial más elevado de las aplicaciones de DM.
[5]	Hussain et al.,	2018	red bayesiana, , PARTE, RF, J48	Rendimiento académico de los estudiantes utilizando métodos ML
[6]	Sivakumar et al. al.,	2018	Clasificadores supervisados, NN, SVM, K-NN, NB DT y DT mejorado	Estudio sobre el conjunto de datos de rendimiento académico de los estudiantes con diferentes métodos de ML y concluye que el DT mejorado es el mejor.
[19]	Z. Raihana y otros,	2018	MVS	El estudio muestra la importancia de la calidad de vida en el rendimiento académico.
[20]	Fahad Razaque y otros, Parneet	2018	Algoritmo bayesiano ingenuo	El estudio es sobre el rendimiento en académicos con técnicas de ML. Ayuda a los estudiantes a mejorar su rendimiento.
[21]	Kaur y otros,	2015	Algoritmos de minería de datos	El estudio se centra en identificar a los estudiantes lentos utilizando algoritmos de clasificación
[22]	Hasan et al.,	2018	Algoritmo de árbol forestal aleatorio	El estudio se centra en mejorar el rendimiento de los estudiantes mediante la predicción temprana.
[23]	Tomás et al., . 2017		Análisis descriptivo y de correlación	El estudio muestra la importancia de la inteligencia emocional, el afrontamiento y la ansiedad ante los exámenes cognitivos en el rendimiento académico.
[24]	Hamaideh et al., 2014		Análisis descriptivo y de correlación	Este estudio muestra el aspecto cognitivo, psicológico y personal del estudiante para el rendimiento académico con el uso de algoritmos de predicción de ML.
[25]	Ying Lin et al., 2017		Análisis descriptivo y de correlación	El estudio demuestra la importancia de la fortaleza mental en el rendimiento académico
[26]	Amirah et al.,	2015	técnicas de MD	Predicción del rendimiento del estudiante
[27]	Tjioe Marvin Christian et al.,	2014	NB, técnica de clasificación de árboles	En este estudio, los autores analizaron la educación personal, la admisión y los datos académicos de los estudiantes y proporcionaron las predicciones sobre estas características.
[28]	Mayilvaganan et al.,	2014	algoritmos de DM	El estudio se enfoca en comparar los algoritmos de clasificación utilizados para predecir el desempeño de los estudiantes en base a los exámenes semestrales.

García et al., [15] consideraron el desempeño sociodemográfico y académico para predecir el desempeño de estudiantes de 1er semestre de Ingeniería. Los estudiantes fueron clasificados en tres categorías: bajo, medio y alto.

Bajo significa estudiantes que aprobaron ninguno o hasta dos cursos, medio se refiere a tres o cuatro cursos aprobados y alto se refiere a todos los aprobados. Aplicaron el clasificador Naïve Bayes y el software Rapid Miner que conducen a una precisión del 60 %. Los autores recopilamos datos de tres universidades de Assam, India, que consisten en información socioeconómica, demográfica y académica de trescientos estudiantes con veinticuatro atributos.

Aplicaron algoritmos ML como Bayes Network, PART, RF y SVM. Se consideran los atributos que más influyen en el uso de la herramienta. Aunque, se implementó el algoritmo Apriori minando con reglas de asociación para todos los atributos. Ogunde et al. [17] analizó el impacto de los exámenes de ingreso a la universidad en las calificaciones de graduación de los estudiantes. Aplicaron el algoritmo del árbol de decisión ID3 para la predicción de las calificaciones finales. Las reglas IF-THEN se enmarcan para representar el conocimiento extraído de los árboles de decisión. Hamoud et al., [18] estudiaron el desempeño de estudiantes portugueses. Se aplicaron algoritmos de árboles de decisión y se compararon los resultados de los árboles DT, Hoeffding (VFDT), C4.5 y RP. Los resultados mostraron que el algoritmo J48 supera a los otros clasificadores y brinda una predicción precisa de los estudiantes que pueden completar con éxito los cursos de educación superior.

### 3. MODELO Y MATERIALES PROPUESTOS

En esta investigación, realizamos y analizamos las diferentes categorías de niveles de aprendizaje de los estudiantes. Según los datos, clasificamos el aprendizaje de los estudiantes en 4 formas que son lentos, moderados o promedio, rápidos y muy rápidos. Para esto, llevamos a cabo las capacitaciones y las pruebas para los estudiantes de CSE de AITAM, Tekkali, India, y recopilamos información sobre el desempeño personal, académico, en clase, preferencias de aprendizaje, etc., de los estudiantes, de 2017 a 2019. La figura muestra el modelo propuesto para predecir las tasas de aprendizaje de los estudiantes con los datos. En el nivel cognitivo, se prepara un conjunto de cuestionarios que observa las habilidades cognitivas como el pensamiento analítico, la identificación de errores y conceptos erróneos, la toma de decisiones, el nivel de conocimiento, etc. de un estudiante para recopilar los datos sin procesar. Los datos sin procesar recopilados y la información sobre su rendimiento académico, sexo, datos demográficos, sus preferencias de aprendizaje y rendimiento en clase se agregan a la tabla de datos y se convierten al formato \*.CSV para el preprocesamiento. El preprocesamiento de datos es un paso crucial que se practica para mejorar la precisión de los algoritmos. La información debe transformarse a la forma utilizada por algoritmos específicos. Los datos preprocesados se ingresan a varios algoritmos de aprendizaje automático y análisis estadístico. Este estudio compara los resultados de los experimentos realizados con los algoritmos de árbol de clasificación, C4.5, SVM, K-NN, Naïve Bayes. Los resultados se comparan utilizando parámetros de precisión, recuperación, puntuación F1 y precisión. El análisis de rendimiento se



curvas ROC. Los resultados del análisis estadístico y los informes del modelo ML se comunican a los analistas para su mayor planificación y acción.

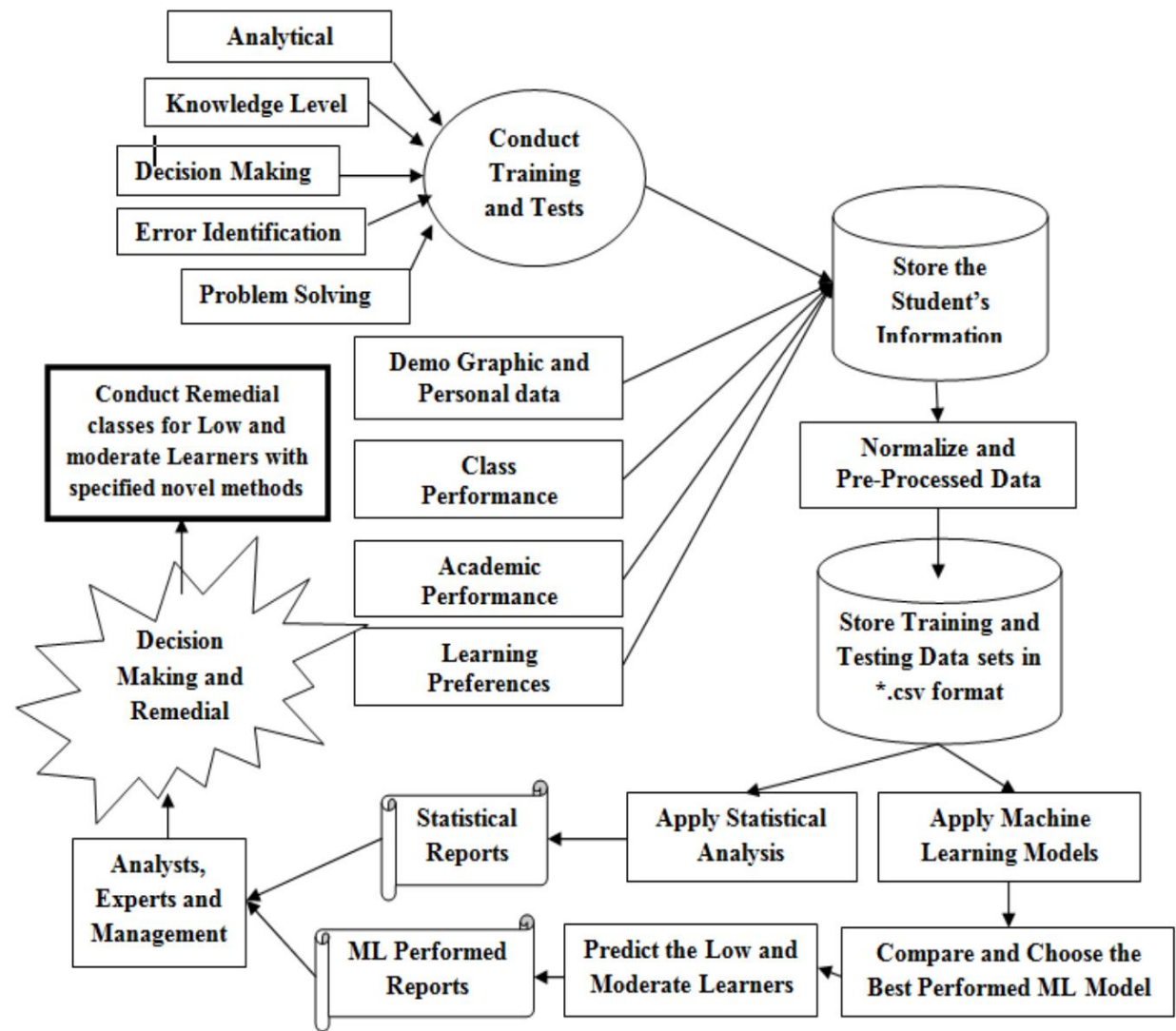


Figura 1. Propuesta de Modelo para la Predicción de los Niveles de Aprendizaje de los Estudiantes

Recopilación y descripción del conjunto de datos:

Recopilación y descripción del conjunto de datos: los datos se recopilan de AITAM College, Tekkali, AP, India. Para este experimento, participan 313 estudiantes de ingeniería, de los cuales 143 estudiantes varones y 170 estudiantes mujeres. Las capacidades totales de aprendizaje del conjunto de datos de las categorías de clase de los estudiantes (rápido, muy rápido, moderado y lento). La Tabla II muestra en detalle la descripción de los Atributos del Conjunto de Datos. Cada atributo se describe con valores nominales o discretos del tipo de datos.

Principalmente sexo, área, valores de función de preferencias de aprendizaje son videos, PDF o PPT. los restos están en valores discretos que son de 1 a 5 o de 1 a 10 valores.

TABLA II . ATRIBUTOS DEL CONJUNTO DE DATOS DESCRIPCIÓN

Características (atributos)	Tipo de datos	Descripción
Sexo	Nominal	Masculino o femenino
Área	Nominal	Rural o Urbano
Preferencias de aprendizaje	Videos nominales o PDF o PPT	
Desempeño de clase	Discreto	1 a 5 valores
CGPA	Discreto	1 a 10 valores
Pensamiento analítico	Discreto	1 a 10 valores
Nivel de conocimiento	Discreto	1 a 10 valores
Habilidades para resolver problemas	Discreto	1 a 10 valores
Toma de decisiones	Discreto	1 a 10 valores
Identificación de errores	Discreto	1 a 10 valores
Atributo de clase	Nominal	rápido, muy rápido, moderado y lento

### Clasificación Naïve Bayes:

Espera que la presencia de un aspecto inequívoco de una clase sea autónomo de cualquier otro aspecto. Según el teorema de Bayes, la probabilidad contingente viene dada por las Ecuaciones (1) y (2). Es el algoritmo más exitoso para muchas aplicaciones, como clasificación de documentos de texto, filtrado de spam, sistema de recomendación, etc.

$$P(i) = \frac{(\tilde{y})}{( )} \quad (1)$$

$$P(i) = \frac{(i)( )}{( )} \quad (2)$$

### Support Vector Machine (SVM): otro

inreible modelo de aprendizaje automático supervisado es SVM, que se puede utilizar tanto para problemas de regresión como de clasificación. Se habla de los números de características 'n' en el espacio n dimensional con cada componente representado por la estimación de una coordenada específica. Un componente de información que comprende n características se traza en este espacio n-dimensional. El punto es encontrar un hiperplano que clasifique y aumente el borde en un espacio n dimensional.

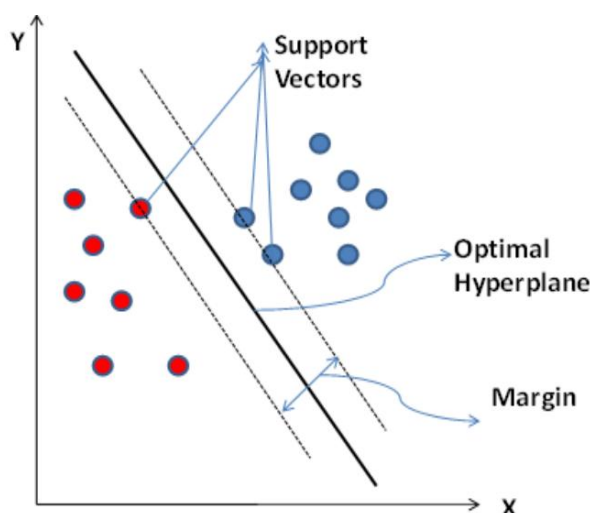


Figura 2. Análisis del clasificador SVM

### Clasificación de K-vecinos más cercanos (k-NN):

El k-NN es un método de algoritmo supervisado no paramétrico adecuado tanto para la clasificación como para la regresión. Considera los k puntos de datos más cercanos en los ejemplos de entrenamiento [12]. El resultado difiere en función del hecho de que K-NN se usa para clasificación o regresión. La salida predice la clase a la que pertenece un punto de datos en función de qué tan cerca coincide con los k vecinos más cercanos. Este es uno de los algoritmos de aprendizaje basado en instancias o de aprendizaje perezoso, porque la función considera los puntos de datos locales y todo el cálculo se difiere hasta la clasificación. Este algoritmo utiliza una función de distancia para calcular el aproximado cercano con los K vecinos más cercanos. Para

variables continuas, se utilizan medidas de distancia Euclidiana, Manhattan y Minkowski y distancia de Hamming para variables categóricas que se muestran en ecuaciones (3-5).

$$= \tilde{y} \tilde{y} \quad (3)$$

$$= \tilde{y} | \tilde{y} | \quad (4)$$

$$= (\tilde{y} | \tilde{y} | ) \quad (5)$$

### Matriz de confusión y parámetros de rendimiento:

En esto, representamos el problema de 4 clases que son Rápido, Muy Rápido, Promedio y Lento. La tabla III muestra la matriz de confusión para el conjunto de datos de aprendizaje del estudiante con el problema de 4 clases. La precisión se calcula por la diagonal de la matriz de confusión. La matriz de confusión se construye utilizando valores reales o verdaderos y valores previstos.

CUADRO III. MATRIZ DE CONFUSIÓN

clasificador	Valores reales o verdaderos				
	Rápido (F)	Muy Rápido (V)	Promedio (A)	Lento (S)	Total
Rápido (F)	FF	VF	FA	FS	T5
FV muy rápida (V)		vv	será	contra	T6
FA media (A)		DE		COMO	T7
Lento (S)	SF	SV	A	SS	T8
Total	T1	T2	T3	T4	T

Calculamos los parámetros de rendimiento como TPR True Positive Rate-Recall-Sensitivity, probabilidad de detección, potencia, FNR-False Negative Rate, Miss Rate, FPR-False Positive Rate, Fall Out, Probabilidad de falsa alarma, SPC-Especificidad, Selectividad, Tasa de verdadero negativo (TNR), PPV- Valor predictivo positivo, Precisión, FOR-Tasa de omisión falsa, LR+-Razón de probabilidad positiva, LR-Razón de probabilidad negativa, ACC-Precisión, FDR-Tasa de descubrimiento falso, NPV-Valor predictivo negativo, DOR -Razón de probabilidades de diagnóstico, F1Score 4 a 17 respectivamente.

$$= \frac{\tilde{y}}{\tilde{y} \tilde{y}} \quad (6)$$

$$= \frac{\tilde{y}}{\tilde{y}} \quad (7)$$

$$= \frac{\tilde{y}}{\tilde{y}} \quad (8)$$





$$= \frac{\ddot{y}}{\ddot{y} \ddot{y} \ddot{y}} \quad (9)$$

$$= \frac{\ddot{y}}{\ddot{y}} \quad (10)$$

$$= \frac{\ddot{y}}{\ddot{y}} \quad (11)$$

$$= \frac{\ddot{y}}{\ddot{y}} \quad (12)$$

$$( ) = \frac{\ddot{y} + \ddot{y}}{\ddot{y}} \quad (13)$$

$$= \frac{\ddot{y}}{\ddot{y}} \quad (14)$$

$$= \frac{\ddot{y}}{\ddot{y}} \quad (15)$$

$$= \frac{\ddot{y}}{\ddot{y}} \quad (16)$$

$$1 = 2 \ddot{y} \frac{\ddot{y}}{\ddot{y}} \quad (17)$$

gerentes y maestros para una decisión posterior. Los algoritmos de ML son muy útiles para predecir la categoría del estudiante con respecto a la clase objetivo (aprendices rápidos, muy rápidos, promedio (moderados), lentos).

#### 4.1 Análisis estadístico La

Tabla IV muestra los valores estadísticos de todos los atributos del conjunto de datos de aprendizaje del estudiante que son Sexo, Urbano y Rural, Preferencias de aprendizaje, Rendimiento en clase, CGPA, Pensamiento analítico, Nivel de conocimiento, Habilidades para resolver problemas, Toma de decisiones e Identificación de errores. En el experimento, utilizamos un total de 313 individuos, 143 estudiantes varones y 170 estudiantes mujeres. Las capacidades totales de aprendizaje del conjunto de datos de las categorías de clase de los estudiantes son rápido, muy rápido, moderado y lento, cuyos valores son 115 (36,7 %), 21 (6,7 %), 127 (40,6 %) y 50 (16,0 %) respectivamente. Comparativamente, las alumnas aprenden muy rápido que los alumnos varones. Los estudiantes de aprendizaje muy rápido son 21, entre los cuales el 61,9% son estudiantes mujeres y el 38,1% de los alumnos son hombres, el 27,8% son hombres y el 72,2% son estudiantes mujeres en aprendizaje rápido de 115 estudiantes rápidos.

#### 4. RESULTADOS Y DISCUSIONES

En esto, llevamos a cabo un análisis estadístico del conjunto de datos total y observamos el rendimiento de los algoritmos de ML de este conjunto de datos. En el análisis estadístico, el conjunto de datos se observa en relación con la categoría de clase y los atributos. Los resultados del análisis estadístico son muy útiles para los analistas,

CUADRO IV. ANÁLISIS ESTADÍSTICO DE ATRIBUTOS DEL CONJUNTO DE DATOS ESTUDIANTILES

Atributos	Categoría	Muy rápido	rápido	lento	Promedio	Total
Sexo	Masculino	8 (38,1%)	rápido 32	33 (66,0%)	70 (55,1%)	143 (45,2%)
	Mujer	13 (61,9%)	(27,8%) 83	17 (34,0%)	57 (44,9%)	170 (54,3%)
Urbano rural	Urbano	13 (61,9%) 8	(72,2%) 70	34 (68,0%)	67 (52,8%)	184 (58,8%)
	Rural	(38,1%)	(60,9%) 45	16 (32,0%)	60 (47,2%)	129 (41,2%)
Preferencias de aprendizaje	Videos	13(61,9%)	(39,1%) 53	25 (50,0%) 11	67 (52,8%)	158 (50,5)
	PDF	8(38,1%)	(46,1%) 43	(22,0%) 14	42 (33,1%) 18	104 (33,2%)
	ppt	0(0,0%) 1	(37,4%) 19	(28,0%) 1	(14,2%) 1	51 (16,3%)
Desempeño de clase	minimo		(16,5%) 1			1
	Mediana					2
	Significar	2 1,90±0,97	2 2,26±0,92	3 2,88±1,07	3 2,69±1,08	2,51±1,05
	máx.	4	5	5	5	5
CGPA	minimo					5
	Mediana	7	6	5	6	7.56
	Significar	8,3 8,25±0,58	7,9 7,89±0,67	6,52	7,45	7,51±0,83
	máx.	9,5 4 8	9,5 3 7	6,56±0,72	7,42±0,68	9.5
Pensamiento analítico	minimo			8,54 1	9,17 1	1
	Mediana			4	5	6
	Significar	7,62±1,36	6,33±1,69	3,96±1,71	5,06±1,67	5,52±1,94
	máx.	10	9	7	8	10
Nivel de conocimiento	minimo	4	2	0	0	0
	Mediana	8	6	3	4	4
	Significar	7,14±1,70	6,02±2,05	2,96±2,01 6	4,24±2,03	4,88±2,37
	máx.	10 2 6	10 2 6	0 2	10	10
Habilidades para resolver problemas	minimo				0	0
	Mediana				4	4
	Significar	6,48±1,84	5,06±1,65 8	2,36±1,82 6	3,65±1,61 6	4,15±2,02
	máx.	10 2	0	0	0	10
Toma de decisiones	minimo					0
	Mediana					4
	Significar	6 6,29±2,33	4 4,49±1,91	2 2,16±1,83	2 2,94±1,92	3,61±2,24

	máx.	10			10	10
	mínimo	0			0	0
Identificación de errores	Mediana	6			2	2
	Significar	5,14±2,87 10	8 0 4	6 0 0	2,35±1,88 8	2,66±2,16
	máx.		3,29±1,91 8	0,96±1,34 4		10

La Figura 3 muestra el 'Pensamiento analítico' para el aprendizaje lento, promedio, rápido y muy rápido, y Total Student. El valor medio del nivel de pensamiento analítico del conjunto de datos total es  $5,52 \pm 1,94$ ; este es el valor más alto que los estudiantes lentos y los estudiantes promedio o moderados y que los valores medios son  $3,96 \pm 1,71$  y  $5,06 \pm 1,67$  respectivamente. Los valores medios del aprendizaje muy rápido y rápido son más altos que el valor medio del conjunto de datos total, ya que los valores medios son  $7,14$  y  $6,02 \pm 2,05$ , respectivamente. Los otros valores estadísticos son

descrito en la tabla 3 así como en la figura 2 en detalle. Según el análisis, el nivel de conocimiento, las habilidades para resolver problemas y la toma de decisiones también son atributos importantes que clasifican los niveles de aprendizaje de los estudiantes. Los valores medios de estos atributos son los valores medios relativos a los estudiantes lentos a muy rápidos. Los valores medios del conjunto total de datos de Nivel de conocimiento, Habilidades para resolver problemas y Toma de decisiones son  $4,88 \pm 2,37$ ,  $4,15 \pm 2,02$  y  $3,61 \pm 2,24$ , respectivamente.

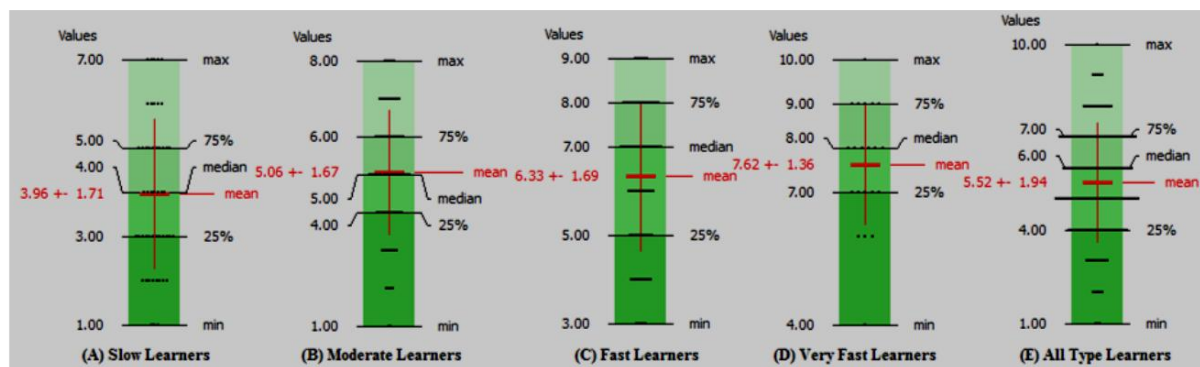


Figura 3. Análisis del Pensamiento Analítico para Estudiantes aprendices

La figura 4 muestra el análisis de algunos de los atributos del conjunto de datos total, como la categoría de clase, las preferencias de aprendizaje, el género y las categorías de área (rural o urbana). De 313 estudiantes, aprendizaje rápido y promedio o moderado (rápido 115 (36.7%), moderado o promedio 127 (40.6%)) los estudiantes son más que muy rápidos y lentos (muy rápido 21 (6.7), lento 50 (16.0%)) que se muestra en la figura 4(A).

La mayoría de los estudiantes eligen videos como su preferencia de aprendizaje y el recuento es de 158 (50.5 %), el siguiente material de aprendizaje que eligen son los pdf. En la categoría de género participan en el experimento 170 (54,3%) personas del sexo femenino y 143 (45,7%) del sexo masculino y en el área 184 son urbanas y 129 rurales.

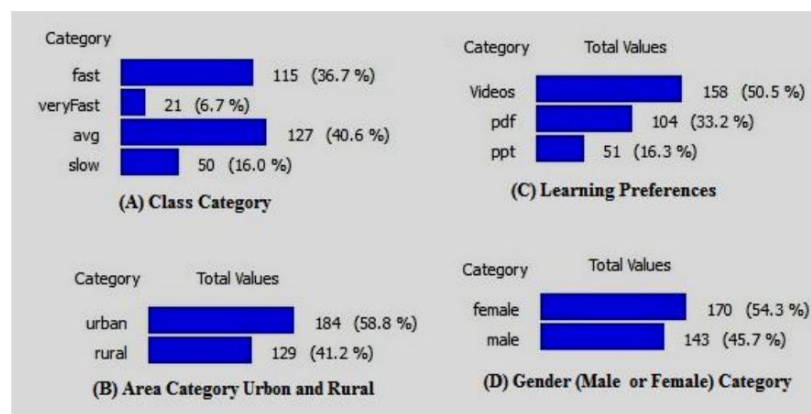


Figura 4. Análisis de atributos del conjunto de datos total



#### 4.2 Algoritmos de aprendizaje automático

Después del análisis estadístico, aplicamos los algoritmos de ML para predecir las tasas de aprendizaje de los estudiantes que son rápidas, muy rápidas, promedio (moderadas) y lentas. En esto, analizamos las precisiones de los algoritmos de ML específicos y los comparamos entre sí. Principalmente, los algoritmos enfocados son Naive Bayes, C4.5, árboles de clasificación (CT), k-NN y SVM. A modo de comparación, calculamos los parámetros de rendimiento como CA, Sens., Spec, AUC, IS, F1, Prec., Recall, Brier, MCC, etc. utilizando Confusion Matrix. La curva ROC también demuestra ser una medida crucial para el rendimiento del conjunto de datos. Finalmente, comparamos la precisión, los valores ROC y el tiempo que se toma para construir el modelo para cada algoritmo.

##### 4.2.1. Análisis del modelo Naïve Bayes (NB) para el aprendizaje de los estudiantes:

En el modelo NB, el total de instancias clasificadas es 114 de 313 en 0,06 segundos. La Tabla V muestra la Matriz de confusión del clasificador de NB.

Los valores se asignan a través de valores predichos y verdaderos. La clase de aprendizaje rápido clasificada por este algoritmo es 35 de 115 y las instancias restantes (75+3+2) se clasifican incorrectamente, además, la clase muy rápida clasificó 20 instancias de 21, la instancia clasificada perdida pasa a la clase rápida. La media

Las instancias de clase (moderadas) se clasifican incorrectamente más, 10 en rápido, 60 es muy rápido y 48 en lento, por lo que las instancias clasificadas correctamente son solo 9. Las instancias de aprendizaje lento de la clase se clasifican con mucha precisión, 50 de 50 clasificadas correctamente. La Tabla V muestra el análisis detallado de la clasificación de NB con una matriz de confusión para calcular parámetros de rendimiento como CA, Sens, Spec, AUC, IS, F1, Perc, Recall, Brier y MCC.

TABLA V. ANÁLISIS DE LA MATRIZ DE CONFUSIÓN DEL CLASIFICADOR NAÏVE BAYES

Clasificador NB	Valores verdaderos				
	clase	Promedio	muy rápido	lento	Total
rapida		20	60	0	155
Muy rápido			0	48	21
medio			9	50	127
lento			0	100	50
Total	rápido 35	110	0	46	12
					313

La Tabla VI muestra el análisis de los parámetros de desempeño del NB.

El valor de precisión promedio de todas las clases es 0,3642 por debajo de 0,5, por lo que este algoritmo no se usa para predecir las medidas de aprendizaje del estudiante. La clase lenta es muy precisa cuando los valores de sensibilidad o recuperación son 1.

La precisión promedio es muy baja donde la sensibilidad es 0.0709 y el valor MCC es 0.14. En general, Naïve Bayes es el modelo de falla para este conjunto de datos.

CUADRO VI. ANÁLISIS DEL RENDIMIENTO DEL MODELO NAÏVE BAYES

Clase	ESE		Especificidad	ABC	ES	F1	Precisión		Brier MCC	
Muy rapido	0,3642	Dirección	0,5377	0,9391	0,4325	0,2273	0,129		0,9447	0,2452
Rápido	0,3642	0,9524	0,9444	0,9391	0,4325	0,4348	<b>0,7609</b>		0,9447	0,3388
Promedio	0,3642	0,3043	<b>0,9839</b>	0,9391	0,4325	<b>0,1295</b>	0,75 0,5		0,9447	<b>0,14</b>
Lento	0,3642		0,8099	0,9391	0,4325	0,6667	<b>0,534975</b>		0,9447	0,6364
Promedio	<b>0,3642</b>	0,0709 1 0,5819	<b>0,818975</b>	<b>0,9391</b>	<b>0,4325</b>	<b>0,364575</b>		Recuperación 0,93447 0,3043	0,0709 1 0,5819	<b>0,14</b>

##### 4.2.2. Análisis del modelo k-NN para el aprendizaje del estudiante:

TABLA VII. ANÁLISIS DE LA MATRIZ DE CONFUSIÓN DEL CLASIFICADOR K-NN

K-NN clasificador	Valores verdaderos					
	clase	rápido	Muy-Rápido	promedio	lento	Total
rapida		0	0	0	0	115
Muy rápido	0 0 0	21	0	0	0	21
promedio		0	127	0	0	127
lento		0	0	0	50	50
Total	115	21	127	50	0	313

En el modelo k-Nearest Neighbor, el total de instancias clasificadas es 313 de 313 en 0,11 segundos. La Tabla VII muestra la Matriz de confusión del clasificador k-NN. Los valores se asignan a través de valores predichos y verdaderos. Todas las clases rápidas, muy rápidas, medias (moderadas) y lentas se clasificaron correctamente en las instancias 115, 21, 127 y 50 en verdadero orden positivo respecto. Entonces, **CNN K-NN** es una muy

modelo preciso y predecible para el conjunto de datos de aprendizaje de los estudiantes.

La Tabla VIII muestra el análisis de los parámetros de rendimiento de k-NN. El valor de precisión promedio de todas las clases es 1, lo que significa que la precisión es del 100%, por lo que este algoritmo se utiliza para predecir con precisión las medidas de aprendizaje del estudiante. Todas las clases son muy precisas donde los valores de sensibilidad o recuperación son 1. La precisión promedio es muy alta donde la sensibilidad es 1, el valor MCC es 1 y los valores Brier son muy bajos, es decir, 0,00047. En general, k-NN es el mejor modelo para este conjunto de datos. La Tabla VIII muestra el análisis detallado de este algoritmo.





TABLA VIII. ANÁLISIS DEL RENDIMIENTO DEL MODELO K-NN

Clase	sensor	CA	Recall	ABC 1	ES	F1	Precisión 1	Recordar	zarza	MCC
Muy rápido	1	1	1		1.7112	1		1	0,0047	1
Rápido	1	1	1	1	1.7112	1	1	1	0,0047	1
Promedio	1	1	1	1	1.7112	1	1	1	0,0047	1
Lento	1	1	1	1	1.7112	1	1	1	0,0047	1
Promedio	1	1	1	1	1.7112	1	1	1	0,0047	1

#### 4.2.3 Análisis del modelo de árboles de clasificación para el

**aprendizaje de los estudiantes:** en el modelo CT, el total de instancias correctamente clasificadas es 295 de 313 en 0,15 segundos. La Tabla IX muestra la Matriz de confusión del clasificador de CTs. Los valores se asignan a través de valores predichos y verdaderos. La clase de aprendizaje rápido clasificó correctamente 110 de las 115 instancias restantes (0+5+0) también se clasificaron incorrectamente, la clase muy rápida clasificó 15 instancias de 21, la instancia de miss clasificó 6 va a la clase rápida. La Tabla X muestra la Matriz de confusión del clasificador k-NN. Los valores se asignan a través de valores predichos y verdaderos.

TABLA IX. ANÁLISIS DE LA MATRIZ DE CONFUSIÓN DEL CLASIFICADOR CTS

Clasificador de TC	Clase	Valores verdaderos				Total
		rápido	Muy-Rápido	promedio	lento	
	rápido	120	0		0	115
	Muy rápido		15	5	0	21
	promedio		1	0	0	127
	lento		0		48	50
	Total		16	122	48	313

TABLA X. ANÁLISIS DEL RENDIMIENTO DEL MODELO CTS

Clase	CA			ABC	ES	F1	Anterior	Recordar	Brier MCC	
Muy rápido	0,9425			0,9937	1,548	0,9362	0,9167	0,9565	0,0905	0,8984
Rápido	0,9425			0,9937	1,548	0,8108	0,9375	0,7143	0,0905	0,8075
Promedio	0,9425			0,9937	1,548	0,9531	0,9457	0,9606	0,0905	0,9208
Lento	0,9425	Dirección	0,565	0,966	0,966	0,966	0,966	0,966	0,966	0,966
Promedio	0,9425	0,89785	0,9772	0,9937	1,548	0,91993	0,949975	0,89785	0,0905	0,9007

#### 4.2.4. C4.5 Modelo de Análisis para el aprendizaje del Estudiante:

En el modelo C4.5, el total de instancias correctamente clasificadas es 288 de 313 en 0,13 segundos. La Tabla XI muestra la Matriz de confusión del clasificador de CTs. Los valores se asignan a través de valores predichos y verdaderos. La clase de aprendizaje rápido clasificó correctamente 108 de 115, las instancias restantes (2+5+0) se clasificaron incorrectamente y también la clase muy rápida clasificó 14 instancias de 21, la instancia de miss clasificó 7 va a la clase rápida. Las instancias clasificadas correctamente de la clase promedio (moderada) son 119 y las instancias clasificadas incorrectamente son 5, 0 y 3 en rápido, muy rápido y lento, respectivamente. De las 50 instancias de aprendizaje lento clasificadas, 47 correctamente y 3 instancias clasificadas incorrectamente se han ido al promedio de clase más cercano. Tabla XI

muestra el análisis detallado de la clasificación de los TC con una matriz de confusión para el cálculo de los parámetros de rendimiento.

Las instancias clasificadas correctamente de clase promedio (moderadas) son 122 y las instancias clasificadas incorrectamente son 4, 1 y 0 en rápido, muy rápido y lento, respectivamente. De las 50 instancias de aprendizaje lento, 48 se clasificaron correctamente y 2 instancias de clasificación incorrecta se fueron al promedio. La Tabla IX muestra el análisis detallado de la clasificación de los TC con la matriz de confusión para el cálculo de los parámetros de desempeño. La Tabla X muestra el análisis de los parámetros de desempeño de los TC. El valor de precisión promedio de todas las clases es 0.9425 por encima del 94%, por lo que este algoritmo es bueno para predecir las medidas de aprendizaje del estudiante. La clase lenta se clasifica correctamente en comparación con otras clases con este algoritmo donde los valores de sensibilidad o recuperación son 0,96 y los valores de MCC son 0,9761. Además, los valores de especificidad y precisión son muy altos en la clase de aprendizaje lento en la que el valor es 1. El valor de Brier es igual a todas las clases en las que el valor es 0,0905. La tabla detallada de análisis de rendimiento X.

descrito en

La Tabla XII muestra el análisis de los parámetros de desempeño del C4.5. El valor de precisión promedio de todas las clases es 0.9201 por encima de 0.9, por lo que este algoritmo es bueno para predecir las medidas de aprendizaje del estudiante. El valor sensitivo o de recuperación es muy alto para la clase de aprendizaje lento que el valor es 0.94 y muy bajo en la clase rápida que el valor es 0.6667. Los valores F1 y MCC también son muy altos para los estudiantes lentos de la clase en comparación con otras clases y valores bajos en la clase Rápida. El valor de brier es igual a todas las clases que el valor es 0.131. El análisis de desempeño detallado descrito en la tabla XII.

TABLA XI. C4.5 ANÁLISIS DE LA MATRIZ DE CONFUSIÓN DE CLASIFICADORES

C 4.5 Clasificador	Clase	Valores verdaderos				Total
		rápido	Muy-Rápido	promedio	lento	
	rápido	2			0	115
	Muy rápido	14	5	3		21
	promedio	0	0	119		127
	lento				47	50
	Total	0 120	0 16	3 127	50	313



TABLA XIII. C4.5 ANÁLISIS DEL RENDIMIENTO DEL MODELO

Clase	CA		ABC	ES	F1		Recuerdo	Brier MCC	
Muy rápido	0,9201	Sentido.	0,9893	1.4674	0,9191		0,9391	0,131	0,8711
Rápido	0,9201	0,9391 0,6667	0,9893	1.4674	0,7568	Anterior 0,9	0,875667	0,131	0,7495
Promedio	0,9201	0,937	0,9893	1.4674	0,937	0,937	0,937	0,131	0,894
Lento	0,9201	0,94	0,9893	1.4674	0,94	0,94	0,94	0,131	0,9286
Promedio	0,9201	0,8707	Especificaciones 0,93932 0,957	0,4674 0,9695	0,888225	0,913	0,8707	0,131	0,8608

#### 4.2.5 Análisis del modelo de máquina de vectores de soporte para el aprendizaje del estudiante:

En el modelo SVM, el total de instancias correctamente clasificadas es 267 de 313 en 0,29 segundos. La Tabla XIII muestra la Matriz de confusión del clasificador de CTs. Los valores se asignan a través de valores predichos y verdaderos. La clase de aprendizaje rápido clasificó correctamente 94 de 115, las instancias restantes (2 + 19 + 0) se clasificaron incorrectamente y la clase muy rápida clasificó 13 instancias de 21, las instancias clasificadas perdidas 8 se fueron rápidamente a la clase más cercana. Las instancias clasificadas correctamente de la clase promedio (moderada) son 113 y las instancias clasificadas incorrectamente son 9, 0 y 5 en rápido, muy rápido y lento, respectivamente. De las 50 instancias de aprendizaje lento, 47 se clasifican correctamente y 3 instancias de clasificación incorrecta se han ido al promedio. La Tabla XIII muestra un análisis detallado de la clasificación de SVM con la matriz de confusión para calcular los parámetros de rendimiento de SVM.

TABLA XIII. ANÁLISIS DE LA MATRIZ DE CONFUSIÓN DEL CLASIFICADOR SVM

Clasificador SVM		Valores verdaderos				
	Clase	rápido	Muy-Rápido	promedio	lento	Total
	rápido		2	19	0	115
	Muy rápido		13	0		21
	promedio	9	0	113		127
	lento	0	0	3	5	50
	Total	111	15	135	47	313

La Tabla XIV muestra el análisis de los parámetros de rendimiento de SVM. El valor de precisión promedio de todas las clases es 0,853 por encima de 0,8, por lo que este algoritmo se usa moderadamente para predecir las medidas de aprendizaje del estudiante. La clase lenta se clasifica correctamente en comparación con otras clases con este algoritmo donde los valores de sensibilidad o recuperación son 0,94 y los valores de MCC son 0,9066. El valor de especificidad es muy alto en una clase muy rápida que el valor es 0,9932. El valor de brier es igual a todas las clases que el valor es 0.1995.

El análisis de rendimiento detallado descrito en la tabla.

CUADRO XIV. ANÁLISIS DEL RENDIMIENTO DEL MODELO SVM

Clase	CA		ABC	ES	F1		Recordar	zarza	MCC
Muy rápido	0,853	Significado 0,619	0,9761	1.2827	0,7222	Anterior 0,8667	0,619	0,1995	0,717
Rápido	0,853	0,8174	0,9761	1.2827	0,8319	0,8468	0,8174	0,1995	0,7372
Promedio	0,853	0,8898	0,9761	1.2827	0,8626	0,837	0,8898	0,1995	0,7649
Lento	0,853	0,94	0,9761	1.2827	0,9216	0,9038	0,94	0,1995	0,9066
Promedio	0,853	0,8166	Especificaciones 0,9932 0,9141	0,8817 0,9271	0,9425	0,83458	0,86358	0,8166	0,1995

#### 4.2.6 Análisis comparativo de modelos de ML

Según la comparación, todos los rendimientos de los algoritmos son buenos, excepto el modelo Naïve Bayes, donde la precisión es solo del 36 %. Según el momento de construir el modelo, NBC es el mejor. Según el análisis, el k-NN es el mejor modelo donde la precisión para predecir la clase objetivo es

100% con 0,11 segundos de construcción del modelo. Los siguientes predictores son CT, C4.5 y SVM en orden donde las precisiones de rendimiento son del 94 %, 92 % y 85 %, respectivamente.

La Tabla XV muestra la descripción detallada de cada modelo de ML especificado.

TABLA XV. ANÁLISIS COMPARATIVO DE RENDIMIENTO DE TODOS LOS MODELOS ML ESPECIFICADOS

ML Algoritmo	Tiempo en Segundos	AC	Sens.		Especificaciones	F1	Rueg.		Brier MCC
NBC	0,06	0,3642	0,5819		AUC 0,81898	0,9391	0,534975		Recuperar 0,5819447
k-NN	0,11	1	1	1	1	1	1	1	0,0047
C4.5	0,13	0,9201	0,8707	0,96955	0,9893	0,888225	0,913	0,8707	0,131
CTS	0,15	0,9425	0,89785	0,853	0,9772	0,9937	0,91993	0,949975	0,89785
MVS	0,29	0,8166		0,9425	0,9761	0,83458	0,8166		0,1995

La figura 5 muestra la clase pronosticada rápida utilizando las curvas ROC. En este análisis, la curva ROC construida con medidas de especificidad (tasa de FP) y sensibilidad (tasa de TP) con valor de 0 a 1. En orden, el k-NN

El valor rápido de clase pronosticada es uno especificado con la línea roja. La línea amarilla indica que el modelo ROC de CT tiene un valor de AUC de 0,99. El valor C4.5 AUC es 0,98 indicado con la línea verde.

La Figura 6 muestra la clase Predicted muy rápido utilizando las curvas ROC. En este análisis, la curva ROC construida con medidas de especificidad (tasa de FP) y sensibilidad (tasa de TP) con valor de 0 a 1. En orden, el k-NN

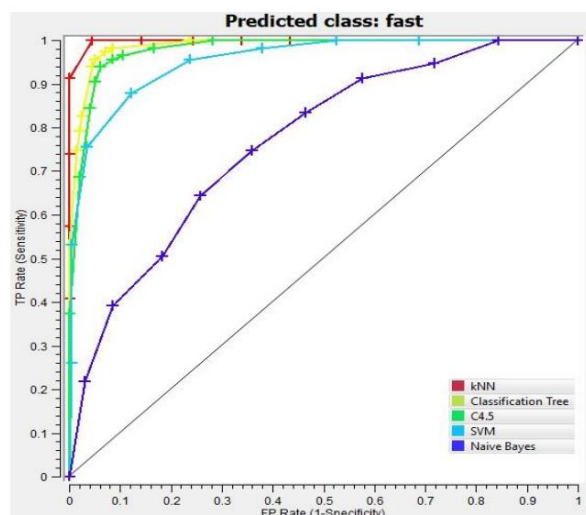


Figura 5. Análisis de ROCs ML especificados para clase pronosticada rápida

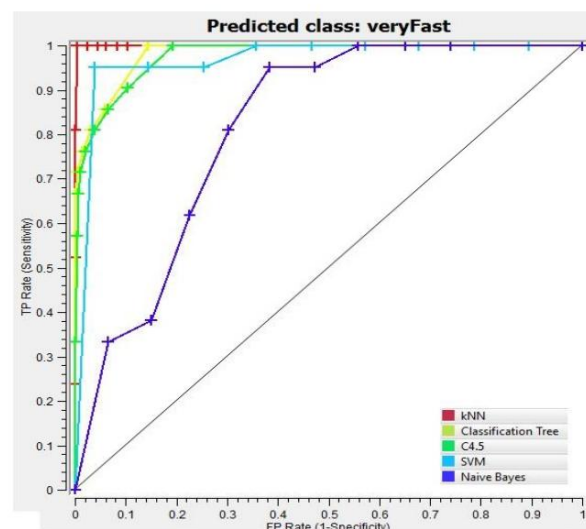


Figura 6. Análisis de ROCs ML especificados para la clase prevista muy rápido

La Figura 7 muestra el alumno moderado o promedio de la clase prevista utilizando las curvas ROC. En este análisis, la curva ROC construida con medidas de especificidad (tasa de FP) y sensibilidad (tasa de TP) con valor de 0 a 1. En orden, el valor moderado o promedio de la clase predicha de k-NN es uno especificado con la línea roja. La línea amarilla indica que el modelo ROC de CT tiene un valor de AUC de 0,99. El valor C4.5 AUC es 0.98 indicado con la línea verde

La figura 8 muestra la clase predicha, de aprendizaje lento, utilizando las curvas ROC. En este análisis, la curva ROC construida con especificidad (FP Rate) y Sensibilidad (TP

El valor rápido de clase pronosticado es uno especificado con la línea roja. La línea amarilla indica que el modelo ROC de CT tiene un valor de AUC de 0,99. El valor C4.5 AUC es 0.98 indicado con la línea verde.

Tasa) medidas con valores de 0 a 1. Todos los MLS son de clase casi predicha, de aprendizaje lento, más cerca de 1.

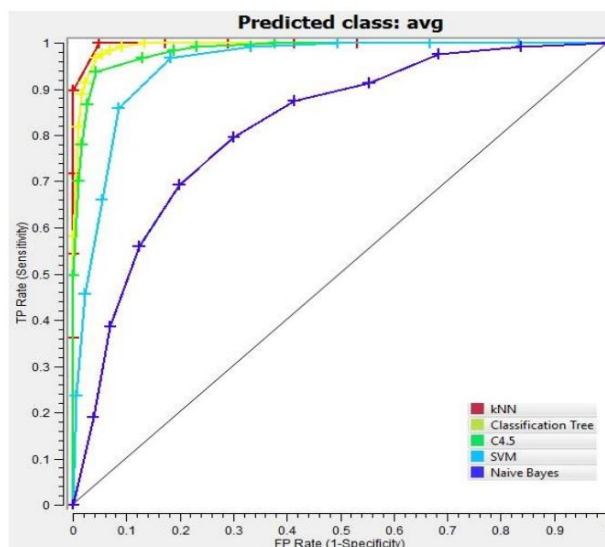


Figura 7. Análisis de ROC ML especificados para el promedio de clase pronosticado

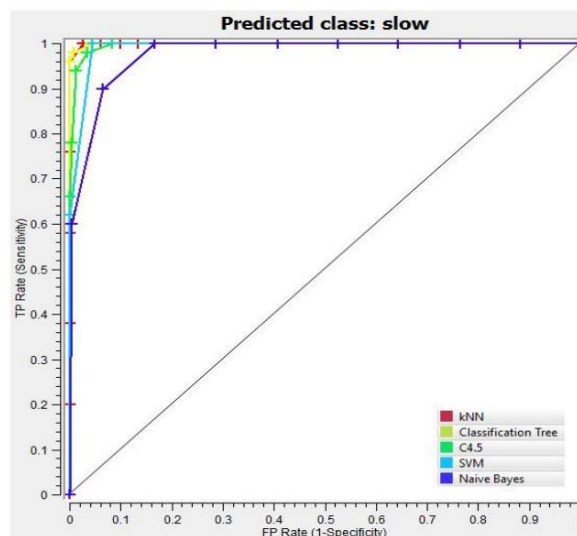


Figura 8. Análisis de ROCs ML especificados para clase prevista lenta

La figura 9 muestra el tiempo de construcción de los modelos ML. En este análisis, el modelo SVM toma el tiempo más alto de 0,29 segundos en comparación con otros modelos. La NBC toma el menor tiempo de 0,06 segundos. Los k-NN, C4.5 y CTs toman los tiempos 0.11, 0.13 y 0.15 segundos respectivamente.

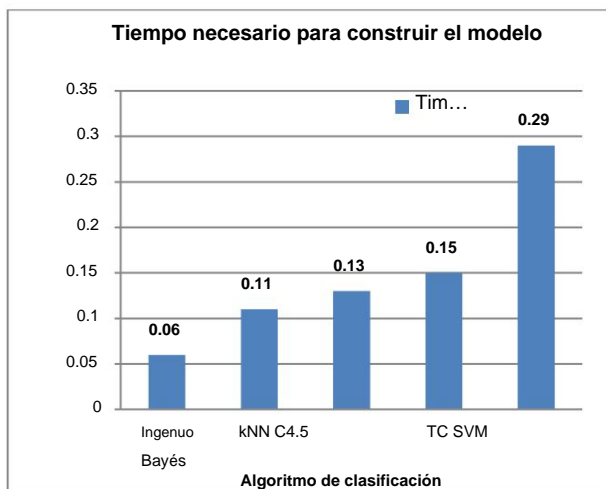


Figura 9. Tiempo necesario para construir los modelos ML

La Figura 10 muestra el análisis comparativo de los modelos ML de precisión de clasificación. En este análisis, el valor CA del modelo k NN es 1 con una precisión del 100 %. El siguiente modelo preciso son los TC con un valor de precisión de 0,9425. Los valores de precisión de los algoritmos ML C4.5 y SVM son 0,9201 y 0,853 respectivamente. El algoritmo NBC no se usa para predecir la clase objetivo donde solo realiza 0.3642 (36%).

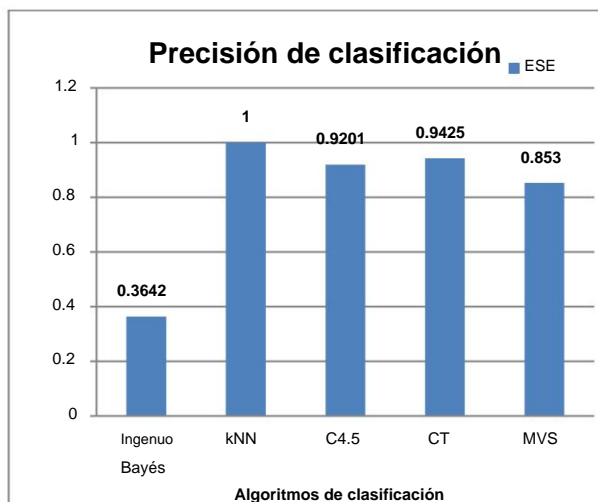


Figura 10. Análisis comparativo con respecto a la precisión de clasificación

## 5. CONCLUSIÓN

La educación es un factor crucial para la mejora personal, económica y social. El desempeño de los estudiantes y el desarrollo de habilidades dependen de sus tasas de aprendizaje relacionadas con los métodos de enseñanza, las capacidades del entrenador y las metodologías. En este empírico investigación educativa en estudiantes de ingeniería, estudiamos casi dos años en profundidad de las capacidades de aprendizaje de los estudiantes. Según los datos y el análisis estadístico, nosotros

observó que la menor cantidad de miembros de los estudiantes muy rápidos y lentos en las corrientes de ingeniería. Los estudiantes lentos también aumentan su tasa de aprendizaje con autoaprendizaje, métodos de enseñanza modernos y secciones de capacitación de recuperación. La minería de datos juega un papel muy importante en esta investigación en la que predecimos alumnos con modelos ML. El modelo k-NN es muy preciso en comparación con otros algoritmos especificados. Además, profundizaremos en este estudio con big data para todas las ramas de los estudiantes de ingeniería y lo compararemos con las capacidades de aprendizaje de los estudiantes de grado en ciencias y artes, y las predicciones con redes neuronales y métodos de aprendizaje profundo.

## RECONOCIMIENTO

Nos gustaría agradecer al Director Prof. VV Nageswara Rao y al Dr. A. Srinivas Rao Principal, AITAM College, Tekkali, HOD y al personal del Departamento CSE, por animarnos y apoyarnos. Estamos pensando en los estudiantes de AITAM para la cooperación para recopilar sus datos de ellos. Los datos que respaldan los hallazgos de este estudio están disponibles en la universidad de AITAM, Tekkali, Andhra Pradesh, India.

## REFERENCIAS

- [1] Fok, WW, He, YS, Yeung, HA, Law, KY, Cheung, KH, Ai, YY y Ho, P. (mayo de 2018). Modelo de predicción para el desarrollo futuro de los estudiantes mediante aprendizaje profundo y motor de inteligencia artificial tensorflow. En 2018 IV Congreso Internacional de Gestión de la Información (MICI) (pp. 103-106). IEEE.
- [2] Ma, X. y Zhou, Z. (2018, enero). Predicción de las tasas de aprobación de los estudiantes utilizando una máquina de vectores de soporte optimizada y un árbol de decisiones. En 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC) (págs. 209-215). IEEE.
- [3] Sekeroglu, B., Dimililer, K. y Tuncal, K. (marzo de 2019). Predicción y clasificación del rendimiento de los estudiantes mediante algoritmos de aprendizaje automático. En Actas de la 8.ª Conferencia Internacional sobre Tecnologías de la Información y la Educación de 2019 (págs. 7-11). ACM.
- [4] Cortez, P. y Silva, AMG (2008). Uso de la minería de datos para predecir el rendimiento de los estudiantes de secundaria.
- [5] Hussain, S., Dahan, NA, Ba-Alwib, FM y Ribata, N. (2018). Minería de datos educativos y análisis del rendimiento académico de los estudiantes utilizando WEKA. Revista indonesia de ingeniería eléctrica y ciencias de la computación, 9(2), 447-459.
- [6] Sivakumar, S. y Selvaraj, R. (2018). Modelado predictivo del rendimiento de los estudiantes a través del árbol de decisiones mejorado. En Avances en Electrónica, Comunicación y Computación (pp. 21-36). Springer, Singapur.
- [7] Yu, LC, Lee, CW, Pan, HI, Chou, CY, Chao, PY, Chen, ZH, ... y Lai, KR (2018). Mejora de la predicción temprana del fracaso académico mediante el análisis de sentimientos en comentarios autoevaluados. Revista de aprendizaje asistido por computadora, 34 (4), 358-365.
- [8] Sorour, SE, Mine, T., Goda, K. y Hirokawa, S. (2015). Un modelo predictivo para evaluar el rendimiento de los estudiantes. Revista de procesamiento de información, 23(2), 192-201.
- [9] Aziz, AA, Ismail, NH y Ahmad, F. (2013). MINERÍA RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES. Diario de Tecnología de la información teórica y aplicada, 53(3).

- [10] Zhang, Y., Oussena, S., Clark, T. y Kim, H. (junio de 2010). Utilice la minería de datos para mejorar la retención de estudiantes en la educación superior: un estudio de caso. En ICEIS (1) (págs. 190-197).
- [11] Mohsin, MFM, Wahab, MHA, Zaiyadi, MF, Noruega, N. M. y Hibadullah, CF (2010, junio). Una investigación sobre el factor de influencia de la calificación de programación de los estudiantes utilizando la minería de reglas de asociación. En Revista Internacional de Avances en Ciencias de la Información y Ciencias de Servicios. vol.
- [12] Kabakchieva, D. (2013). Predecir el rendimiento de los estudiantes mediante el uso de métodos de minería de datos para la clasificación. Cibernética y tecnologías de la información, 13(1), 61-72.
- [13] Huang, S. y Fang, N. (2011, octubre). Trabajo en progreso- Predicción del rendimiento académico de los estudiantes en un curso de introducción a la ingeniería. En 2011 Conferencia Fronteras en Educación (FIE) (pp. S4D-1). IEEE.
- [14] Parack, S., Zahid, Z. y Merchant, F. (2012, enero). Aplicación de minería de datos en bases de datos educativas para predecir tendencias y patrones académicos. En 2012 IEEE International Conference on Technology Enhanced Education (ICTEE) (págs. 1-4). IEEE.
- [15] García, EPI, & Mora, PM (2011, noviembre). Modelo de predicción del rendimiento académico para estudiantes de primer año. En 2011 X Congreso Internacional Mexicano de Inteligencia Artificial (pp. 169-174). IEEE.
- [16] Vital, TP, Lakshmi, BG, Rekha, HS y DhanaLakshmi, M. (2019). Análisis del desempeño de los estudiantes con el uso de estudios estadísticos y de conglomerados. En Soft Computing en Data Analytics (págs. 743-757). Springer, Singapur.
- [17] Ogunde, AO y Ajibade, DA (2014). Un sistema de minería de datos para predecir las calificaciones de graduación de los estudiantes universitarios utilizando el algoritmo de árbol de decisión ID3. Revista de informática y tecnología de la información, 2(1), 21-26.
- [18] Hamud, A. (2016). Selección del mejor algoritmo de árbol de decisión para la predicción y clasificación de la acción de los estudiantes.
- [19] Raihana, Z. y Farah Nabilah, AM, Clasificación de estudiantes basada en la calidad de vida y el rendimiento académico mediante el uso de una máquina de vectores de soporte. Revista de la Academia UiTM Negeri Sembilan 6(1):45-52, 2018.
- [20] Razaque, F., Soomro, N., Shaikh, SA, Soomro, S., Samo, J. A., Kumar, N., and Dharejo, H. Uso del algoritmo naïve bayes para el análisis del desempeño académico de los estudiantes de licenciatura. En: Ingeniería, Tecnologías y Ciencias Aplicadas (ICETAS), 2017 4th IEEE International Conference. págs. 1-5, 2017.
- [21] Parneet Kaur, Manpreet Singh, Gurpreet Singh Josan, Algoritmos de minería de datos basados en clasificación y predicción para predecir aprendices lentos en el sector educativo. 3ra Conferencia Internacional sobre Tendencias Recientes en Computación 2015 (ICRTC-2015)
- [22] Hasan, R., Palaniappan, S., Raziff, ARA, Mahmood, S. y Sarker, KU Student Academic Performance Prediction by using Decision Tree Algorithm, 4th International Conference on Computer and Information Sciences. 2018.
- [23] Thomas, CL, Cassady, JC y Heller, ML (2017). "La influencia de la inteligencia emocional, la ansiedad ante los exámenes cognitivos y las estrategias de afrontamiento en el rendimiento académico de pregrado" Aprendizaje y diferencias individuales, 55, 40 – 48.
- [24] Hamaideh, SH y Hamdan-Mansour, AM (2014). "Variables psicológicas, cognitivas y personales que predicen el rendimiento universitario entre los estudiantes de ciencias de la salud", Nurse Education Today, 34, 703 – 708.
- [25] Lin, Y., Clough, PJ, Welch, J. y Papageorgiou, KA (2017). "Las diferencias individuales en la fortaleza mental se asocian con el rendimiento académico y los ingresos", Personality and Individual Differences, 113, 178 - 183.
- [26] Una revisión sobre la predicción del rendimiento de los estudiantes utilizando técnicas de minería de datos. Amirah Mohamed Shahiria,\*, Wahidah Husaina, Nur'aini Abdul Rashida, a School of Computer Sciences Universiti Sains Malaysia 11800 USM, Penang, Malaysia
- [27] TM Christian, M. Ayub, Exploración de la clasificación usando nbtrees para predecir el rendimiento de los estudiantes, en: Ingeniería de datos y software (ICODSE), Conferencia internacional de 2014, IEEE, 2014, págs. 1 a 6.
- [28] Mayilvaganan .M., D. Kalpanadevi, Comparación de técnicas de clasificación para predecir el rendimiento del entorno académico de los estudiantes, en: Tecnologías de red y comunicación (ICCNT), Conferencia internacional de 2014, IEEE, 2014, págs. 113–118.



#### El Dr. PanduRanga Vital Terlapu obtuvo una

Licenciatura en Ciencias de la Computación de la Universidad Andhra de AP, India en 1995 y una Maestría en Aplicaciones Informáticas de la Universidad de Andhra en el año 1998. Completó su M.

Tech en Ciencias de la Computación e Ingeniería de la Universidad Acharya Nagarjuna de AP, India y completó su Ph.D. en Ciencias de la Computación e Ingeniería de la Universidad GITAM de AP, India. Tiene 19 años de

docencia y 13 años de experiencia investigadora. Actualmente trabaja como profesor asociado en el Departamento de Informática e Ingeniería del Instituto Aditya de Tecnología y Gestión (AITAM), India. Es miembro de ACM, miembro vitalicio de la Sociedad Internacional de Ingeniería y Ciencias de la Computación (ICSSES), EE. UU. y miembro vitalicio de la Sociedad India de Educación Técnica (ISTE), Nueva Delhi, India. Ha publicado más de 30 artículos de investigación en revistas internacionales de renombre, incluido SCOPUS indexado y una conferencia que incluye Springer, Elsevier y también está disponible en línea. Es revisor de revistas de renombre como Springer, Elsevier e IEEE. Su principal trabajo de investigación se centra en Machine Learning, Deep Learning y Data Mining, Data and Big Data Analytics, IoT e Inteligencia Computacional, Voice Analysis y Voice Processing, Bioinformatics.



**Smt. K. Sangeeta**, profesor asistente sénior, Departamento de CSE, AITAM, Tekkali. Ella hizo su MTech. en Informática e Ingeniería de IIT Madras. Su interés de investigación incluye el aprendizaje automático, la inteligencia artificial, el aprendizaje profundo y la criptografía. Ella tiene 13 años de

experiencia en el campo de la docencia y un año en el área industrial. Tiene 6 publicaciones en

Revistas internacionales, 2 en revistas nacionales y

presentó 2 artículos en congresos nacionales y 3 artículos en congresos internacionales



**La Dra. Kalyana Kiran Kumar** obtuvo los títulos de doctorado y maestría en 2015 y 2004, respectivamente, de la Universidad de Andhra, Visakhapatnam, India.

Actualmente trabaja como profesor en el Departamento de Ingeniería Eléctrica y Electrónica, Facultad de Ingeniería AITAM, TEKKALI, Andhra Pradesh. Tiene una vasta experiencia docente de 16 años. Ha publicado más de 20 artículos de investigación en revistas internacionales de renombre. Es miembro de IETE, ISTE e IAENG (HK).

Sus intereses de investigación incluyen sistemas a gran escala, diseño de controladores, sistemas de intervalos, sistemas fraccionarios y sistemas no lineales.