

Identificación de Dificultades de Aprendizaje en Álgebra Lineal mediante Modelos Predictivos de Machine Learning

Identifying Learning Difficulties in Linear Algebra Using Machine Learning Predictive Models

Campino-Aguirre, T.I.¹✉; Moreno-Núñez, E.²; Rodríguez-Lozano, K.V.³

Datos de Adscripción:

¹ ✉ Campino Aguirre, Tonatzin Isabel
Tecnológico Nacional de México – Instituto Tecnológico Superior de Lerdo.
Ciudad Lerdo, Durango. México. ROR ID: <https://ror.org/00davry38>
ROR ID: <https://ror.org/04bjbt12>
tonatzin.ca@itslerdo.edu.mx
ORCID: <https://orcid.org/0009-0006-1884-7121>

² Moreno Núñez, Elda
Tecnológico Nacional de México – Instituto Tecnológico Superior de Lerdo.
Ciudad Lerdo, Durango. México. ROR ID: <https://ror.org/00davry38>
ROR ID: <https://ror.org/04bjbt12>
elda.mn@itslerdo.edu.mx
ORCID: <https://orcid.org/0000-0002-5912-082X>

³ Rodríguez Lozano, Karla Verónica
Tecnológico Nacional de México – Instituto Tecnológico Superior de Lerdo.
Ciudad Lerdo, Durango. México. ROR ID: <https://ror.org/00davry38>
ROR ID: <https://ror.org/04bjbt12>
karla.rl@itslerdo.edu.mx
ORCID: <https://orcid.org/0000-0003-2832-4165>

Resumen - Las dificultades que enfrentan los estudiantes universitarios en asignaturas como Álgebra Lineal suelen pasar desapercibidas hasta que los rezagos ya son difíciles de revertir. Este trabajo propone una alternativa basada en modelos de Aprendizaje Automático para detectar de forma temprana esas dificultades, tomando como población de estudio a cien estudiantes de segundo semestre de educación superior que cursaban simultáneamente varias asignaturas de ciencias exactas. Con el fin de aprovechar la información escrita generada durante el curso, se recopilaron respuestas y comentarios académicos de los participantes, los cuales fueron transformados en representaciones numéricas mediante la técnica de Frecuencia de Término-Frecuencia Inversa de Documento. Sobre esa base, se entrenaron y compararon dos clasificadores supervisados: Máquinas de Vectores de Soporte y Naive Bayes, orientados a distinguir entre estudiantes con y sin dificultades de aprendizaje. Ambos algoritmos alcanzaron niveles de desempeño satisfactorios, aunque las Máquinas de Vectores de Soporte mostraron ventajas claras en precisión y en la identificación de los casos más críticos. Estos hallazgos permiten concluir que incorporar herramientas de este tipo en los entornos educativos universitarios no solo es viable, sino que puede transformar la manera en que los docentes detectan y atienden las necesidades de sus estudiantes. Actuar con antelación, antes de que las dificultades se profundicen, representa un avance concreto hacia una educación más atenta, personalizada y sustentada en evidencia.

Palabras Clave - Álgebra lineal, Aprendizaje automático, Clasificación supervisada, Dificultades de aprendizaje, Educación superior, Modelos predictivos.

Abstract - Learning difficulties in courses like Linear Algebra often go unnoticed until students fall so far behind that catching up becomes a real challenge. This study proposes an approach based on Machine Learning models to detect these difficulties early, drawing on a sample of one hundred second-semester higher education students who were simultaneously enrolled in several exact sciences courses. To make use of the written information generated throughout the course, academic responses and comments from participants were collected and transformed from numerical representations using the Term Frequency-Inverse Document Frequency technique. From that foundation, two supervised classifiers were trained and compared: Support Vector Machines and Naive Bayes, both aimed at distinguishing between students with and without learning difficulties. Both algorithms achieved satisfactory performance levels, though Support Vector Machines showed clear advantages in accuracy and in identifying the most critical cases. These findings suggest that incorporating tools of this kind into university settings is not only feasible, but could genuinely change how educators detect and respond to their students' needs. Taking action early, before difficulties become deeply rooted, is a meaningful step toward an education that is more attentive, personalized, and grounded in evidence.

Keywords - Higher education, Learning difficulties, Linear Algebra, Machine Learning, Predictive models, Supervised classification.

I. INTRODUCCIÓN

Uno de los problemas más persistentes en la educación superior es, paradójicamente, uno de los más difíciles de ver a tiempo: detectar cuándo un estudiante está comenzando a perder el paso antes de que esa dificultad se convierta en reprobación o, peor aún, en abandono escolar. En asignaturas que exigen un alto nivel de abstracción matemática, como Álgebra Lineal, este problema se vuelve especialmente agudo. La velocidad con la que avanzan los programas académicos y la complejidad conceptual inherente a la materia dejan poco margen para que las dificultades se detecten de forma natural en el transcurso del semestre (Kolman & Hill, 2006). No se trata de una asignatura periférica: el Álgebra Lineal es un pilar formativo en carreras de ingeniería, ciencias computacionales y disciplinas afines, y su dominio constituye un prerrequisito implícito para avanzar con solidez en materias como Cálculo Vectorial, Probabilidad y Estadística, Inteligencia Artificial y Procesamiento

de Señales, entre otras. La magnitud del problema queda reflejada en las cifras: investigaciones previas han documentado tasas de reprobación que superan el 30% en esta asignatura a nivel universitario, y la causa más frecuentemente señalada son las brechas conceptuales que no fueron identificadas ni atendidas a tiempo (Dorier & Sierpinski, 2001).

Frente a este escenario, la minería de datos educativos ha ido ganando terreno como un campo interdisciplinar que articula técnicas de aprendizaje automático, estadística y gestión de bases de datos para analizar, de manera sistemática, el comportamiento académico de los estudiantes (Baker & Inventado, 2014). Lo que hace valioso este enfoque es su capacidad para revelar patrones que a simple vista permanecerían ocultos en la masa de datos que genera cualquier proceso de enseñanza-aprendizaje, abriendo así la posibilidad de diseñar intervenciones pedagógicas tempranas, oportunas y sustentadas en evidencia concreta (Romero & Ventura, 2010). Como se muestra en la Figura 1, el proceso metodológico de este trabajo está organizado en siete etapas que recorren el camino desde la recolección inicial de datos textuales hasta la generación de estrategias pedagógicas adaptadas a las necesidades de cada estudiante.

Entre las herramientas de procesamiento del lenguaje natural disponibles para analizar texto académico, una de las más consolidadas es la técnica conocida como Frecuencia de Término-Frecuencia Inversa de Documento, desarrollada originalmente por Salton y Buckley (1988). Su funcionamiento se basa en convertir documentos de texto en vectores numéricos, asignando a cada término un peso que refleja tanto su presencia en el documento analizado como su distribución en el conjunto total de textos. En términos simples, pondera más aquellas palabras que son frecuentes en un documento, pero poco comunes en los demás, lo que permite capturar con mayor precisión los rasgos distintivos de cada texto. En contextos educativos, esta técnica ha demostrado ser una herramienta eficaz para inferir el nivel de comprensión conceptual de los estudiantes a partir de sus respuestas escritas (Manning et al., 2008).

En cuanto a los algoritmos de clasificación, dos de los más estudiados en problemas de categorización textual son las Máquinas de Vectores de Soporte y el clasificador Naive Bayes. Las Máquinas de Vectores de Soporte, introducidas por Cortes & Vapnik (1995), operan buscando el hiperplano que mejor separa las distintas clases en el espacio de características, y se distinguen por su robustez frente a datos de alta dimensionalidad, una condición habitual cuando se trabaja con representaciones textuales. El clasificador Naive Bayes, por su parte, aplica el teorema de Bayes asumiendo independencia condicional entre las características del texto, lo que lo convierte en una alternativa computacionalmente eficiente y sorprendentemente competitiva en muchos escenarios prácticos (Mitchell, 1997). Kotsiantis (2012) demostró que ambos tipos de modelos son capaces de predecir el rendimiento académico con niveles de precisión aceptables cuando se cuenta con datos cualitativos sobre el desempeño de los estudiantes.

Este trabajo parte precisamente de esa posibilidad y la lleva al terreno del aula real, aplicando modelos predictivos de aprendizaje automático sobre los textos producidos por estudiantes universitarios de segundo semestre en la asignatura

de Álgebra Lineal. La pregunta central que guía el estudio es si es posible clasificar a los estudiantes en dos grupos —con y sin dificultades de aprendizaje— tomando como única fuente de información sus respuestas escritas, procesadas mediante vectorización textual y evaluadas con dos clasificadores supervisados.

Vale la pena detenerse un momento en la elección del texto escrito como fuente de datos, porque no es una decisión arbitraria. Podría haberse optado por calificaciones parciales, registros de asistencia u otros indicadores más convencionales. Sin embargo, cada una de esas señales llega tarde. Una calificación reprobatoria le informa al docente que algo salió mal, pero el daño ya está hecho; la inasistencia es un síntoma, no una causa, y muchas veces aparece cuando el estudiante ya perdió la motivación para seguir intentándolo. El texto que un estudiante produce al intentar explicar un concepto matemático con sus propias palabras revela algo cualitativamente diferente: cómo está procesando la información en ese preciso momento. Quien comprende escribe con precisión y coherencia; quien no comprende lo deja ver en la vaguedad, en las contradicciones o en el silencio de su escritura. Esa diferencia no necesita un examen formal para manifestarse; está presente en las respuestas que los estudiantes producen en el día a día del aula.

Esta intuición está respaldada por evidencia empírica. Investigaciones recientes en procesamiento de lenguaje natural aplicado a la educación han mostrado que incluso respuestas muy breves, de apenas dos o tres oraciones, contienen información léxica suficiente para distinguir distintos niveles de dominio conceptual (Hand et al., 2001; Manning et al., 2008). Esto tiene implicaciones prácticas muy concretas para instituciones con grupos numerosos, donde resulta materialmente imposible que el docente haga un seguimiento individualizado de cada estudiante a lo largo del semestre. En ese contexto, un clasificador automático entrenado con los datos del propio grupo no pretende reemplazar al docente ni sustituir su criterio profesional: simplemente le proporciona información que, sin esta herramienta, nunca llegaría a tiempo para marcar una diferencia.

Los resultados de este estudio buscan contribuir al diseño de estrategias pedagógicas personalizadas orientadas a la atención temprana del rezago académico. El resto del artículo se organiza de la siguiente manera: la Sección II detalla la metodología técnica empleada; la Sección III presenta los resultados obtenidos y su discusión; y las Secciones IV, V y VI recogen las conclusiones del trabajo, los agradecimientos correspondientes y las referencias bibliográficas.

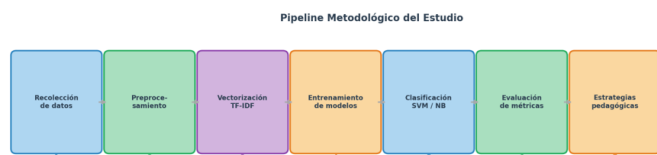
II. PARTE TÉCNICA DEL ARTÍCULO

Esta sección describe el proceso metodológico seguido para la recolección, el preprocesamiento y el análisis de los datos textuales, así como el diseño, entrenamiento y evaluación de los modelos de clasificación supervisada empleados.

La Figura 1 resume el pipeline completo desde la recolección de datos hasta las estrategias pedagógicas personalizadas.

Figura 1

Pipeline metodológico del estudio.



2.1 Contexto y población de estudio

El estudio se llevó a cabo con una muestra de 100 estudiantes de segundo semestre de una institución de educación superior tecnológica inscritos en la asignatura de Álgebra Lineal durante el ciclo escolar 2025. La Tabla 1 describe la distribución de los estudiantes por categoría de aprendizaje, género y promedio académico al inicio del semestre.

Tabla 1

Caracterización de la muestra de estudio ($n = 100$ estudiantes).

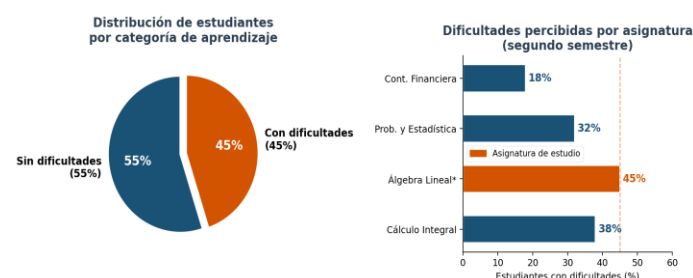
Categoría	n	Promedio inicial (0-100)	Distribución por género (H/M)
Sin dificultades de aprendizaje	55	79.4 ± 6.1	29 / 26
Con dificultades de aprendizaje	45	64.2 ± 8.3	25 / 20
Casos críticos (subgrupo)	12	55.1 ± 5.9	7 / 5
Total	100	72.3 ± 9.8	54 / 46
Tasa de reprobación semestral previa	—	28.0%	—

La elección de Álgebra Lineal como objeto de estudio responde a su elevada tasa de dificultad percibida entre los estudiantes de segundo semestre, quienes la cursaban junto con Cálculo Integral, Probabilidad y Estadística y Contabilidad Financiera. En el semestre previo, los participantes habían cursado Cálculo Diferencial y Matemáticas Discretas.

La Figura 2 ilustra la distribución de los estudiantes por categoría de segundo semestre y compara el porcentaje de dificultades reportadas en las cuatro asignaturas del semestre que se está cursando, evidenciando que Álgebra Lineal concentra el mayor índice (45%).

Figura 1

Izquierda: distribución de 100 estudiantes por categoría de aprendizaje. Derecha: porcentaje de estudiantes con dificultades por asignatura



2.2 Recolección de datos

Los datos textuales se recopilaron a lo largo del semestre mediante tres estrategias complementarias: (a) respuestas escritas a preguntas abiertas sobre conceptos de Álgebra Lineal: vectores, matrices, transformaciones lineales, espacios vectoriales; (b) comentarios académicos registrados durante sesiones de resolución de problemas; y (c) reflexiones escritas al final de cada unidad temática.

En total se recopilaron 1,200 registros textuales, con un promedio de 12 registros por estudiante. De ellos, 660 correspondieron a la categoría sin dificultades de aprendizaje y 540 a la categoría con dificultades de aprendizaje.

La Tabla 2 presenta ejemplos representativos de respuestas en cada categoría.

Tabla 2

Ejemplos de respuestas textuales por categoría de aprendizaje.

Sin dificultades de aprendizaje	Con dificultades de aprendizaje
"La base del espacio vectorial debe ser linealmente independiente y generar todo el espacio."	"No entiendo para qué sirve la matriz, parece que es como una tabla de números."
"El espacio nulo del operador contiene todos los vectores que se mapean al vector cero."	"Creo que el determinante se calcula multiplicando los números de alguna manera especial."
"Aplicé eliminación gaussiana para encontrar la forma escalonada y determinar el rango."	"Me confundí entre el espacio imagen y el espacio nulo, no sé cuál es cuál."
"Prefiero el método de la regla de Cramer para encontrar la solución de un Sistema de Ecuaciones Lineales"	"El resultado me dio un número grande y no sé si eso está bien o mal."

Cada registro fue etiquetado manualmente por el docente como perteneciente a una de dos categorías. El criterio de etiquetado consideró la precisión conceptual del texto, el uso de vocabulario técnico adecuado y la coherencia lógica de las explicaciones. Para garantizar la confiabilidad del etiquetado, un segundo evaluador revisó de manera independiente el 20% de los registros, obteniendo un coeficiente de acuerdo Kappa de Cohen de 0.83, que indica concordancia sustancial entre evaluadores (Landis & Koch, 1977).

Un aspecto que se cuidó especialmente durante la recolección fue la naturalidad de los registros. Las respuestas no fueron solicitadas en un formato de examen ni como actividad evaluable con calificación directa; se presentaron como reflexiones de seguimiento al final de cada sesión, lo que redujo la tendencia de los estudiantes a escribir lo que creían que el docente quería leer. Se buscaba capturar el lenguaje espontáneo de comprensión, no una performance académica. Esa distinción es metodológicamente importante: un texto que el estudiante produce bajo presión evaluativa tiende a aproximar el vocabulario técnico que ha memorizado, no el que realmente maneja. El lenguaje informal de reflexión, en cambio, expone con mayor fidelidad el grado real de apropiación conceptual.

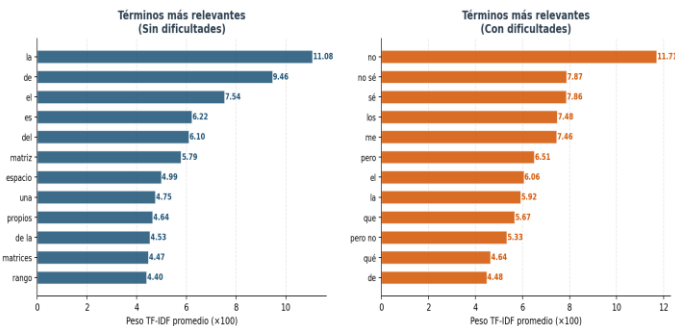
También se registró el orden cronológico de los textos: los primeros registros de cada estudiante corresponden a las unidades iniciales del semestre (vectores y matrices), mientras que los últimos cubren temas más avanzados como transformaciones lineales y valores propios. Este orden no se explotó en el modelo actual —que trata todos los registros como independientes—, pero representa información valiosa para versiones futuras que consideren la evolución temporal del vocabulario del estudiante como variable adicional.

2.3 Preprocesamiento y vectorización TF-IDF

Antes de aplicar los modelos de clasificación, el corpus textual fue sometido a un proceso de preprocesamiento que incluyó: (a) normalización: conversión a minúsculas y eliminación de caracteres especiales; (b) eliminación de palabras vacías; (c) tokenización: segmentación en unidades léxicas individuales; y (d) lematización: reducción de palabras a su forma base. Posteriormente, se aplicó la técnica de Frecuencia de Término-Frecuencia Inversa de Documento con un vocabulario de 300 términos y bigramas (n-gramas de hasta dos palabras), generando un espacio de características de alta dimensionalidad. La Figura 3 muestra los doce términos con mayor peso en cada clase.

Figura 2

Términos con mayor peso promedio de Frecuencia de Término-Frecuencia Inversa de Documento por clase: sin dificultades (izquierda) y con dificultades (derecha).



2.4 Modelos de clasificación supervisada

Con los vectores numéricos como entrada, se entrenaron y evaluaron dos clasificadores. Las Máquinas de Vectores de Soporte con kernel lineal buscan encontrar el hiperplano de máximo margen que separa las dos clases en el espacio de características. El clasificador Naive Bayes Multinomial aplica el teorema de Bayes bajo la suposición de independencia condicional entre términos. Ambos modelos fueron implementados en Python mediante la biblioteca Scikit-learn (Pedregosa et al., 2011). El corpus se dividió a nivel de registro textual: 960 registros (80%) se destinaron al entrenamiento y 240 (20%) a la prueba. Las métricas finales se reportaron sobre 20 estudiantes del conjunto de prueba. La documentación disponible no especifica la regla utilizada para integrar, por estudiante, las predicciones derivadas de sus registros; por ello, no se atribuye un procedimiento de agregación particular. Además, se aplicó validación cruzada estratificada de cinco pliegues para valorar la estabilidad de los resultados. A continuación, se presenta el fragmento de código principal de la implementación:

```
1. from sklearn.feature_extraction.text import TfidfVectorizer
2. from sklearn.svm import LinearSVC
3. from sklearn.naive_bayes import MultinomialNB
4. from sklearn.model_selection import cross_val_score,
   StratifiedKFold
5. from sklearn.metrics import accuracy_score,
   precision_score, recall_score, f1_score
6.
7. vectorizer = TfidfVectorizer(max_features=300,
   ngram_range=(1,2), sublinear_tf=True)
8. X = vectorizer.fit_transform(corpus_preprocesado)
9. y = etiquetas # 0 = sin dificultad, 1 = con dificultad
10.
11. svm_model = LinearSVC(C=0.5, max_iter=5000,
   random_state=42)
12. nb_model = MultinomialNB(alpha=1.5)
13.
14. cv = StratifiedKFold(n_splits=5, shuffle=True,
   random_state=42)
15. svm_scores = cross_val_score(svm_model, X, y, cv=cv)
16. nb_scores = cross_val_score(nb_model, X, y, cv=cv)
```

2.5 Métricas de evaluación

El desempeño de cada clasificador fue evaluado mediante exactitud, precisión, sensibilidad y puntuación F1.

La exactitud expresa la proporción de predicciones correctas sobre el total de instancias. La precisión cuantifica, de entre las predicciones positivas, qué proporción corresponde a verdaderos positivos; por su parte, la sensibilidad mide la capacidad del modelo para identificar todos los casos positivos reales.

La puntuación F1 representa la media armónica entre precisión y sensibilidad. Adicionalmente, se calcularon las curvas de característica Operativa del Receptor y el área bajo dicha curva como indicador global de discriminación entre clases.

III. RESULTADOS Y DISCUSIÓN

La Tabla 3 presenta los resultados comparativos de ambos modelos sobre el conjunto de prueba (20 instancias), evaluados con las cuatro métricas descritas en la sección anterior.

Tabla 3

Comparación de métricas de desempeño: Máquinas de Vectores de Soporte vs. Naive Bayes (conjunto de prueba, n = 20).

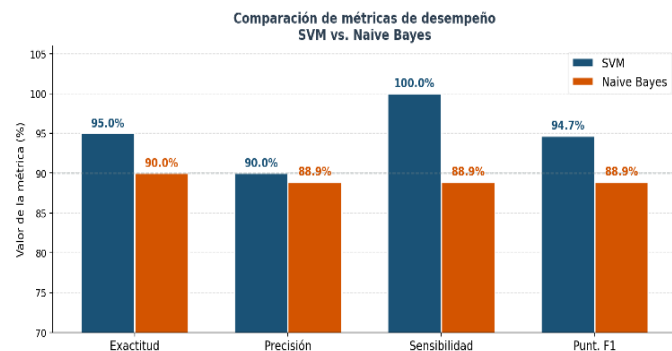
Métrica	SVM	Naive Bayes	Diferencia
Exactitud	95.0%	90.0%	+5.0 pp
Precisión	90.0%	88.9%	+1.1 pp
Sensibilidad	100.0%	88.9%	+11.1 pp
Puntuación F1	94.7%	88.9%	+5.8 pp
Área bajo la curva ROC	0.960	0.960	+0.000

La Figura 4 ilustra estos resultados de forma visual mediante un gráfico de barras agrupadas.

Los resultados revelan que ambos clasificadores alcanzaron niveles de desempeño satisfactorios. Las Máquinas de Vectores de Soporte obtuvieron valores superiores en exactitud, precisión, sensibilidad y puntuación F1, mientras que ambos modelos alcanzaron la misma área bajo la curva ROC (0.960).

Figura 4

Comparación de métricas de desempeño por modelo: Máquinas de Vectores de Soporte (azul) vs. Naive Bayes (naranja).

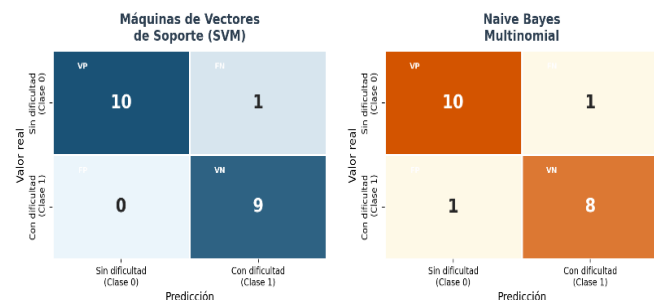


La diferencia más notable se observa en la sensibilidad: 100.0% frente a 88.9%, lo que significa que las Máquinas de Vectores de Soporte identificaron la totalidad de los estudiantes con dificultades en el conjunto de prueba, sin ningún falso negativo. Esto es pedagógicamente crucial, ya que un estudiante con dificultades no detectadas representa un caso de rezago que avanzará sin intervención.

La Figura 5 muestra las matrices de confusión de ambos modelos: Máquinas de Vectores de Soporte (izquierda) y Naive Bayes (derecha). VP = Verdadero Positivo, FP = Falso Positivo, FN = Falso Negativo, VN = Verdadero Negativo.

Figura 5

Matrices de confusión de ambos modelos

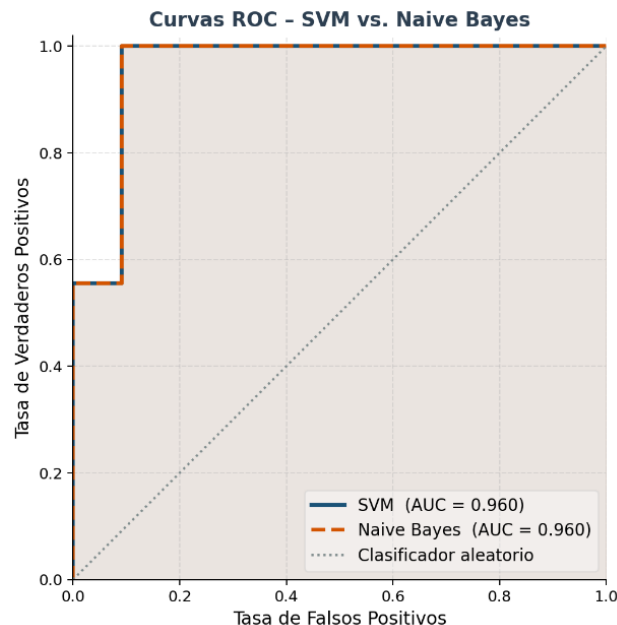


Para las Máquinas de Vectores de Soporte se observa una sola clasificación errónea: un estudiante sin dificultades fue clasificado como con dificultades (falso positivo), mientras que todos los casos reales de dificultades fueron detectados correctamente. El clasificador Naive Bayes cometió dos errores: un falso positivo y un falso negativo. Este último tipo de error es el más costoso en el contexto educativo, pues implica dejar sin atención a un estudiante en riesgo.

La Figura 6 presenta las curvas de Característica Operativa del Receptor de ambos modelos. Tanto las Máquinas de Vectores de Soporte como Naive Bayes alcanzaron un área bajo la curva de 0.960, por lo que mostraron la misma capacidad discriminativa global en esta métrica, aunque difirieron en sensibilidad y precisión. Ambos modelos superan ampliamente el umbral de un clasificador aleatorio (área = 0.500).

Figura 6

Curvas de característica operativa del receptor: máquinas de vectores de soporte vs. Naive Bayes. El área bajo la curva mide la capacidad discriminativa global.



La Tabla 4 detalla los resultados de la validación cruzada de cinco pliegues, mientras que la Figura 7 los visualiza mediante diagramas de caja; cada punto representa la exactitud en un pliegue; la caja muestra la distribución de los cinco valores.

Tabla 4

Resultados de validación cruzada estratificada (5 pliegues) por modelo.

Modelo	Pliegues (exactitud %)	Media (%)	Desv. estándar (%)
SVM	80.0 / 95.0 / 90.0 / 90.0 / 90.0	89.0	4.9
Naive Bayes	80.0 / 85.0 / 90.0 / 90.0 / 90.0	87.0	4.0
Diferencia SVM - NB	—	+2.0	0.9

Las Máquinas de Vectores de Soporte obtuvieron una exactitud media del 89.0% $\pm$ 4.9%, mientras que Naive Bayes alcanzó 87.0% $\pm$ 4.0%. La menor desviación estándar de Naive Bayes indica mayor estabilidad entre pliegues, aunque la media de las Máquinas de Vectores de Soporte es consistentemente superior.

El análisis de los términos con mayor peso en la vectorización, visualizado en la Figura 3, reveló patrones semánticos diferenciados entre ambas clases.

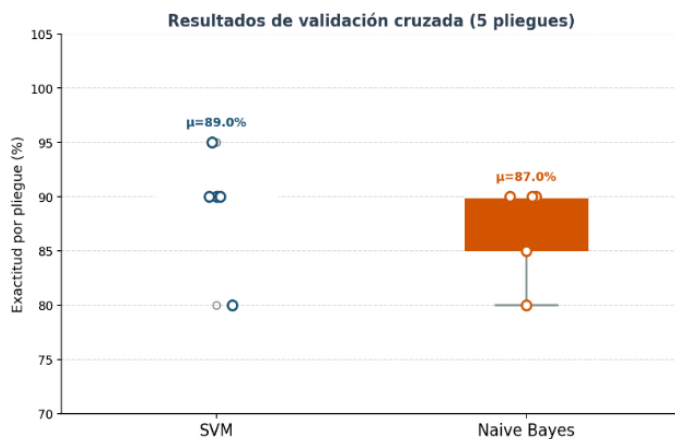
Los textos clasificados como sin dificultades concentraron términos técnicos precisos como "transformación lineal", "espacio nulo", "independencia lineal" e "imagen del operador". En contraste, los textos con dificultades mostraron mayor frecuencia de expresiones vagas como "número grande", "no entiendo" y "no

sé exactamente". Esto confirma que el lenguaje escrito del estudiante constituye un indicador confiable de su nivel de comprensión conceptual, coherente con lo reportado por Manning et al. (2008).

Estos resultados son consistentes con la literatura sobre clasificación supervisada en entornos educativos. Romero y Ventura (2010) reportaron que las Máquinas de Vectores de Soporte son especialmente robustas en espacios de alta dimensionalidad generados por vectorización textual. Kotsiantis (2012) señaló que los modelos basados en texto libre ofrecen una perspectiva complementaria a la de las métricas cuantitativas tradicionales.

Figura 7

Resultados de validación cruzada de cinco pliegues.



La Tabla 5 compara el rendimiento de los modelos del presente estudio con los reportados en investigaciones previas de referencia.

Tabla 5

Comparación del desempeño de los modelos con estudios previos en minería de datos educativos.

Estudio	Técnica empleada	Exactitud reportada	Dominio
Presente estudio	SVM + TF-IDF	95.0%	Álgebra Lineal
Presente estudio	Naive Bayes + TF-IDF	90.0%	Álgebra Lineal
Romero y Ventura (2010)	SVM	87.5%	Cursos en línea (general)
Kotsiantis (2012)	Árbol de decisión	83.2%	Matemáticas generales

Un hallazgo complementario que los números de la Tabla 3 no capturan directamente es la consistencia del error que comete cada modelo. Las Máquinas de Vectores de Soporte fallaron en

un único caso: un estudiante sin dificultades reales cuyo vocabulario escrito tenía características que el modelo asoció con el grupo de riesgo; posiblemente porque ese alumno expresaba incertidumbre de forma explícita sin que esta reflejara una dificultad conceptual genuina. En cambio, el modelo de Naive Bayes distribuyó sus errores en dos casos distintos. Esta diferencia no modifica los porcentajes dramáticamente, pero en cambio, sí tiene implicaciones prácticas: el tipo de error de SVM genera una falsa alarma que el docente puede resolver con una conversación breve; el error de Naive Bayes genera un caso no detectado que no recibe ninguna atención, por lo que en un sistema de soporte a la decisión docente, es preferible tener un exceso de alertas a tener casos perdidos.

Vale la pena detenerse en un dato que puede parecer paradójico: ambos modelos obtienen exactamente el mismo valor de área bajo la curva ROC (0.960), aunque difieren en sensibilidad y precisión. La razón es que esta métrica no evalúa un único punto de corte, sino que promedia el desempeño a lo largo de todos los umbrales posibles, mientras que los valores de la Tabla 3 reflejan lo que ocurre en un umbral particular.

Esto tiene implicaciones concretas para quienes diseñan estos sistemas. En un contexto real, una institución podría preferir ser más precavida y marcar como "en riesgo" a más estudiantes, aunque con menor certeza; esto elevaría la sensibilidad del modelo, pero también generaría más falsos positivos. No hay una respuesta única correcta: la curva ROC precisamente sirve para visualizar ese equilibrio y ayudar a cada institución a elegir el umbral que mejor se adapte a su realidad, sus capacidades de atención y el tipo de riesgo que está dispuesta a asumir.

Hay un hallazgo que se debe destacar por sus implicaciones prácticas: los 12 estudiantes identificados como casos críticos, aquellos con el promedio inicial más bajo (55.1 ± 5.9) y con la mayor distancia entre lo que escriben y lo que la asignatura demanda, representan precisamente el grupo en el que más importa no equivocarse. Las Máquinas de Vectores de Soporte los detectaron a todos sin excepción, mientras que Naive Bayes no logró identificar correctamente a uno de ellos. Aunque la muestra no es suficientemente grande para sacar conclusiones definitivas, el patrón es sugerente: SVM parece construir una frontera de decisión más sensible a los perfiles extremos, justo donde la intervención es más urgente.

En cuanto a si esto podría funcionar fuera del laboratorio, la respuesta es alentadora. Los modelos no requieren encuestas adicionales, pruebas especiales ni que el docente cambie su forma de enseñar. Se alimentan directamente de lo que los estudiantes ya escriben durante las actividades normales de la asignatura. Con que esas respuestas queden registradas en formato digital, algo que la mayoría de las plataformas educativas actuales ya hacen por defecto, el sistema puede operar. Eso lo convierte en una propuesta no solo técnicamente sólida, sino también realista para instituciones que no cuentan con grandes recursos tecnológicos.

Los resultados deben interpretarse como evidencia exploratoria. La muestra se limitó a 100 estudiantes de segundo semestre de una sola institución y las métricas finales se calcularon sobre 20 estudiantes del conjunto de prueba. Además, la partición 80/20 se realizó por registro textual, por lo que registros de un mismo estudiante pudieron quedar representados en ambos

subconjuntos. Esta dependencia potencial puede favorecer la similitud entre entrenamiento y prueba y limitar la generalización de los porcentajes observados. La validación cruzada de cinco pliegues aporta una estimación adicional de estabilidad, pero no sustituye una validación externa con estudiantes e instituciones diferentes.

IV. CONCLUSIONES

Este estudio demostró que es técnicamente posible, y pedagógicamente valioso utilizar modelos de aprendizaje automático para detectar de manera temprana las dificultades que enfrentan los estudiantes en Álgebra Lineal, partiendo de algo tan cotidiano como el lenguaje que ellos mismos producen por escrito. Las Máquinas de Vectores de Soporte lograron una exactitud del 95.0% y una sensibilidad del 100.0%, lo que significa que identificaron a la totalidad de los estudiantes con dificultades dentro del conjunto de prueba, sin dejar escapar ningún caso. Desde una perspectiva pedagógica, ese resultado no es un detalle menor: es, en realidad, el hallazgo más importante de toda la investigación.

Al comparar los dos clasificadores, quedó claro que Naive Bayes ofrece ventajas reales en términos de eficiencia computacional y estabilidad entre los distintos pliegues de validación, pero que las Máquinas de Vectores de Soporte resultan más adecuadas cuando lo que está en juego es no dejar pasar ningún caso sin detectar. Dicho esto, los dos enfoques no son excluyentes: combinarlos podría abrir una vía prometedora para manejar mejor la heterogeneidad de los perfiles estudiantiles y mejorar aún más el desempeño conjunto del sistema.

Sin embargo, el aporte más significativo de este trabajo va más allá de los porcentajes de exactitud. Lo que realmente importa es lo que estos modelos hacen posible en el aula: identificar con anticipación a quienes están perdiendo el hilo, antes de que esa dificultad se convierta en un rezago difícil de revertir. Con esa información en mano, el docente puede actuar de manera oportuna y diferenciada, ya sea a través de andamiaje cognitivo, tutorías dirigidas, estrategias de aprendizaje entre pares, u otras intervenciones adaptadas a cada caso. En ese sentido, la propuesta representa un avance concreto hacia una educación más atenta, más flexible y más fundamentada en evidencia real sobre el proceso de aprendizaje.

Como líneas de trabajo futuras, se propone explorar cuatro direcciones: extender el modelo a otras asignaturas del área de ciencias exactas; incorporar fuentes de datos complementarias como registros de plataformas digitales, asistencia y calificaciones parciales; evaluar el impacto real de las intervenciones pedagógicas mediante un diseño cuasi-experimental con grupo de control; y explorar modelos de lenguaje preentrenados que ofrezcan representaciones semánticas más ricas y sensibles al contexto disciplinar específico.

En un plano más amplio, este estudio invita a repensar el papel que pueden jugar los datos textuales dentro de la evaluación formativa en educación superior. La evaluación tradicional sigue siendo predominantemente cuantitativa: exámenes, tareas, calificaciones numéricas. Esos instrumentos tienen su lugar y su validez, pero presentan una limitación estructural importante: llegan tarde. Miden el resultado del aprendizaje una vez que ya ocurrió, no el proceso mientras se está desarrollando. El texto escrito, en cambio, especialmente cuando se produce en

condiciones naturales y sin la presión directa de una calificación, ofrece una ventana distinta: permite asomarse a cómo está pensando el estudiante en ese momento. Incorporar ese tipo de evidencia al ecosistema de evaluación de una asignatura no implica desplazar los instrumentos existentes, sino agregarles una capa de información que antes simplemente se descartaba porque no había herramientas para procesarla de manera sistemática.

A esto se suma otro argumento a favor de la propuesta: su escalabilidad. Una vez entrenado el modelo con los datos de un grupo, aplicarlo a nuevos estudiantes es computacionalmente sencillo. No es necesario reentrenarlo cada semestre si la asignatura y las condiciones institucionales se mantienen estables. Con ajustes menores, el mismo esquema puede extenderse a grupos sucesivos, y con cada ciclo escolar el volumen de datos crece, lo que a su vez mejora progresivamente la calidad del clasificador. El costo más alto, que corresponde a la etapa inicial de recolección y etiquetado de datos, se distribuye y amortiza a lo largo de múltiples periodos, lo que hace de esta una inversión sostenible a mediano plazo.

Pero hay algo que los números no dicen por sí solos y que merece ser dicho con claridad: detectar quién está en riesgo no cambia nada si esa información no llega al docente a tiempo y de una forma que le permita actuar. La verdadera utilidad de estas herramientas no reside en el porcentaje de exactitud que aparece en una tabla; reside en la posibilidad concreta de acortar la distancia entre el momento en que un estudiante empieza a perder el hilo y el momento en que alguien interviene. Esa ventana de oportunidad, que en la dinámica habitual del semestre suele cerrarse antes del primer examen parcial, es exactamente lo que este tipo de modelos puede mantener abierta.

En contextos donde el número de estudiantes por docente es elevado y el tiempo disponible para el seguimiento individual es escaso, contar con un indicador objetivo basado en el propio texto del alumno no es un privilegio ni un recurso accesorio: es una herramienta de equidad. No todos los estudiantes que atraviesan dificultades las hacen visibles de manera espontánea. Algunos no faltan, no protestan, no preguntan; simplemente guardan silencio hasta que reprobar se vuelve casi inevitable. Detectar esa dificultad silenciosa, la que pasa desapercibida en el aula pero deja huellas en la escritura, es quizás el aporte más duradero que este trabajo puede ofrecer a quienes enseñan cada día.

V. AGRADECIMIENTOS

Los autores agradecen al Tecnológico Nacional de México – Instituto Tecnológico Superior de Lerdo, a la Subdirección de Posgrado e Investigación y a la División de Ingeniería en Sistemas Computacionales por las facilidades otorgadas para la recolección de datos y la realización de la investigación, así como a los estudiantes participantes por su disposición y compromiso durante el desarrollo del estudio.

VI. REFERENCIAS

- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. En J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4



- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Dorier, J. L., & Sierpiska, A. (2001). Research into the teaching and learning of linear algebra. En D. Holton (Ed.), *The teaching and learning of mathematics at university level* (pp. 255–273). Springer. https://doi.org/10.1007/0-306-47231-7_24
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT Press.
- Kolman, B., & Hill, D. R. (2006). *Álgebra lineal* (8.ª ed.). Pearson Educación.
- Kotsiantis, S. B. (2012). Use of machine learning techniques for educational purposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37(4), 331–344. <https://doi.org/10.1007/s10462-011-9234-x>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809071>
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., y Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)

No.	Rol	Descripción
		<i>Nota 1: favor de eliminar los corchetes una vez colocado su nombre</i> <i>Nota 2: elimine el todo texto de apoyo que se proporcionó en los asteriscos de cada fila y debajo de cada nombre.</i>
1	Conceptualización [Conceptualization]	Tonatzin Isabel Campino Aguirre
2	Curación de datos [Data curation]	Tonatzin Isabel Campino Aguirre
3	Análisis formal [Formal analysis]	Elda Moreno Núñez
4	Obtención de fondos [Funding acquisition]	
5	Investigación [Investigation]	Tonatzin Isabel Campino Aguirre
6	Metodología [Methodology]	Karla Verónica Rodríguez Lozano
7	Administración del proyecto [Project administration]	Elda Moreno Núñez
8	Recursos [Resources]	Karla Verónica Rodríguez Lozano
9	Software [Software]	Tonatzin Isabel Campino Aguirre
10	Supervisión [Supervision]	Elda Moreno Núñez
11	Validación [Validation]	<i>Tonatzin Isabel Campino Aguirre</i>
12	Visualización [Visualization]	Karla Verónica Rodríguez Lozano
13	Redacción - borrador original [Writing – original draft]	Elda Moreno Núñez
14	Redacción - revisión y edición [Writing – review & editing]	Tonatzin Isabel Campino Aguirre