

ANÁLISIS DEL ÉXITO ACADÉMICO MEDIANTE APRENDIZAJE AUTOMÁTICO: ADICCIÓN Y CHATGPT

JUAN CARLOS TORRES-DIAZ

jctorres@utpl.edu.ec

<https://orcid.org/0000-0003-1503-0379>

Universidad Técnica Particular de Loja, Ecuador

RUTH MARÍA REÁTEGUI ROJAS

rmreategui@utpl.edu.ec

<https://orcid.org/0000-0002-7304-4413>

Universidad Técnica Particular de Loja, Ecuador

Recibido: 12 de setiembre del 2024 / Aceptado: 2 de octubre del 2024

doi: <https://doi.org/10.26439/interfases2024.n020.7390>

RESUMEN. En este trabajo, se analiza la incidencia de las variables adicción al teléfono, a la pornografía, número de veces que se desbloquea el teléfono a cada hora y nivel de confianza en ChatGPT sobre el éxito académico de un grupo de 4278 estudiantes de ocho universidades de Ecuador. Se emplean los siguientes métodos: árboles de decisión (DT), *random forest* (RF) y *support vector machine* (SVM). Los resultados obtenidos señalan niveles similares en la precisión alcanzada en los tres algoritmos, respecto a la exactitud, en caso de SMOTE, los DT son el algoritmo que presenta mayor exactitud (*accuracy* = 0,64); y, en el caso de RandomOverSampler, el algoritmo SVM muestra mayor exactitud (*accuracy* = 0,59).

PALABRAS CLAVE: ChatGPT / adicción / aprendizaje automático

ANALYSIS OF ACADEMIC SUCCESS USING MACHINE LEARNING: ADDICTION AND CHATGPT

ABSTRACT. This paper analyzes the impact of the variables phone addiction, pornography addiction, number of times the phone is unlocked per hour, and level of confidence in ChatGPT on the academic success of a group of 4278 students from eight universities in Ecuador. The decision trees (DT), random forest (RF), and support vector machine (SVM) methods are used. The results obtained indicate similar levels of precision achieved in the three algorithms; in terms of accuracy, in the case of SMOTE, DT is the algorithm that presents the highest accuracy (*accuracy* = 0,64); and, in the case of RandomOverSampler, the SVM algorithm had the highest accuracy (*accuracy* = 0,59).

KEYWORDS: ChatGPT / addiction / machine learning

INTRODUCCIÓN

La predicción del éxito o reprobación académica, así como del comportamiento general de los estudiantes dentro de los sistemas educativos, es un tema de gran relevancia que ha sido investigado de manera continua durante décadas. A lo largo de los años, se han identificado múltiples factores que influyen en los resultados académicos, lo que ha llevado a la implementación de diversos enfoques para tratar de comprender mejor este fenómeno. Entre los predictores más destacados se encuentran las variables demográficas, como el género, la edad, el nivel socioeconómico y la ubicación geográfica de los estudiantes, que han sido objeto de análisis en múltiples estudios recientes (Batool et al., 2023; Nayak et al., 2023). Estas variables proporcionan un contexto básico, pero necesario, para entender las diferencias en el rendimiento académico entre distintos grupos de estudiantes.

Otro conjunto clave de predictores incluye las calificaciones previas de los estudiantes y el contexto educativo en el que se desenvuelven, que abarca factores como la calidad de la enseñanza, los recursos disponibles y el ambiente escolar (Beaulac & Rosenthal, 2019; Cui et al., 2024; Nachouki et al., 2023). Estos elementos no solo reflejan el desempeño pasado, sino que también ofrecen un marco para predecir el éxito futuro, ya que un adecuado entorno educativo tiende a facilitar mejores resultados. Asimismo, los factores motivacionales y psicológicos juegan un papel importante en la predicción del rendimiento académico. Aspectos como la motivación intrínseca, el manejo del estrés y las habilidades de autorregulación influyen directamente en la capacidad de los estudiantes para cumplir con las exigencias académicas (Hellás et al., 2018). La combinación de estos factores puede variar significativamente entre individuos, lo que añade una capa de complejidad a la predicción del éxito académico.

Los predictores mencionados sirven como datos de entrada para diversos algoritmos de aprendizaje automático, los cuales se utilizan para analizar grandes cantidades de información y generar modelos predictivos. La precisión de estos modelos depende en gran medida del contexto en el que se aplican, así como de la naturaleza específica del fenómeno que se está estudiando. En cualquier caso, los resultados obtenidos permiten tanto a los profesores como a las instituciones educativas ajustar sus programas y enfoques de enseñanza, con el fin de mejorar la calidad del aprendizaje de sus estudiantes (Chen, 2006).

Otros predictores que usualmente no son considerados son la adicción al teléfono (Chaudhury & Tripathy, 2018; Hong et al., 2024) y el uso de ChatGPT, que muestra incidencia positiva sobre el rendimiento académico (Gutiérrez-Aguilar et al., 2024). Los estudiantes perciben esta herramienta como útil, puesto que mejoran su capacidad para buscar información, analizar datos (Elkhodr et al., 2023) y autorregular su ritmo y estilo de aprendizaje (Cai et al., 2023); sin embargo, estamos aún en etapas tempranas en la investigación del efecto de esta tecnología en el aprendizaje. Por ese motivo, en

este estudio se analiza el efecto de estas variables y, de manera específica, se aplican distintos algoritmos de aprendizaje automático en la predicción del éxito académico motivado por la adicción a la pornografía, la adicción al teléfono y la confianza en la herramienta ChatGPT; y, por último, se comparan sus resultados de ejecución.

1. ESTADO DEL ARTE

Estudios de revisión bibliográfica muestran de manera general variables, métodos y resultados que señalan el camino recorrido en la investigación en este ámbito. En el estudio de Hellas et al. (2018), se analizan investigaciones que utilizan una amplia gama de métodos para la predicción, incluidos DT, regresiones y en general técnicas de aprendizaje automático. En cuanto a los factores predictores, se identifican diversas categorías, como el historial educativo, datos demográficos y factores motivacionales y psicológicos. Sin embargo, se observa que muchos estudios tienden a usar un número limitado de factores a menudo relacionados con el rendimiento previo en asignaturas. La mayoría de estas investigaciones se enfocan en predecir variables, como las calificaciones finales de los cursos, el promedio de calificaciones (GPA) y la retención en los programas académicos. Al respecto, un trabajo acerca de la predicción del rendimiento de los estudiantes comparó aproximadamente 260 estudios en un rango de 20 años mediante un análisis de los principales factores que afectan la predicción del rendimiento de los estudiantes, los tipos de técnicas de minería de datos que incluyen algoritmos de predicción y selección de características, y las herramientas que más frecuentemente se utilizan. Los hallazgos muestran que ANN y RF son los algoritmos de minería de datos más utilizados, mientras que WEKA es la herramienta que cuenta con mayor aceptación para procesar este tipo de estudios. Este trabajo tiene una gran relevancia debido a que arrojó una clasificación de los predictores de la deserción, cuyo orden de importancia es el siguiente: calificaciones obtenidas, variables demográficas, actividades académicas, asistencia, inteligencia/personalidad y actividades sociales o extracurriculares (Batool et al., 2023).

El estudio de Chen y Zhai (2023) proporciona una visión exhaustiva de cómo las técnicas de aprendizaje automático pueden aplicarse eficazmente en la educación y destaca la importancia de adaptar los métodos a las características específicas de los datos. En ese sentido, se aplicaron siete técnicas de aprendizaje automático sobre tres conjuntos de datos. A través de cuatro métricas de evaluación, los resultados señalaron al algoritmo RF como superior.

En los algoritmos de aprendizaje automático empleados para predecir, su principal característica está en los altos niveles de precisión que alcanzan. En el estudio de Alghamdi y Rahman (2023), para predecir el rendimiento académico de estudiantes de secundaria se construyeron tres modelos predictivos con los algoritmos Naïve Bayes (NB), RF y J48. Además, se aplicó la técnica Synthetic Minority Over-sampling Technique (SMOTE) para equilibrar los datos. El rendimiento de los modelos se comprobó mediante

la validación cruzada de diez pasos y partición directa, los resultados mostraron que el modelo NB alcanzó una precisión de predicción del 99,34 %, seguido por el modelo RF con un 98,7 %.

En la misma línea, el estudio de Nachouki et al. (2023) busca predecir las calificaciones de los estudiantes en cursos universitarios con el algoritmo RF. El objetivo es identificar los factores que influyen en el rendimiento académico. Utilizan un *dataset* de 650 registros con datos como tipo de escuela secundaria, calificación de la escuela secundaria, género, categoría del curso, porcentaje de asistencia a clase, GPA y modo de entrega del curso (presencial, en línea o híbrido). Los resultados del modelo indicaron que el GPA y la escuela secundaria fueron los predictores más significativos. La categoría del curso y el porcentaje de asistencia a clase también mostraron una importancia considerable, mientras que el modo de entrega del curso y género no tuvieron un efecto significativo.

Nayak et al. (2023) realizaron un estudio para predecir el rendimiento académico en entornos de aprendizaje en línea mediante características demográficas, académicas y conductuales. Analizaron dos conjuntos de datos y emplearon modelos de clasificación como DT, NB, RF y perceptrón multicapa (MLP). El modelo MLP demostró ser superior, con una precisión del 90,74 % en el conjunto de datos GRIET y 97,08 % en Kalboard 360 tras optimización y selección de características. Sin embargo, al incluir las características conductuales, el modelo RF alcanzó una precisión perfecta del 100 %, resultado que fue superior a los otros métodos.

El estudio de Cui et al. (2024) se centra en la aplicación de modelos de regresión a datos de estudiantes universitarios, se aplicaron las regresiones: lineal, Ridge, Lasso, Huber, regresión de vectores de soporte (SVR), K-Nearest Neighbors (KNN), DT y redes neuronales (NN). Se utilizaron los puntajes de 21 cursos y la métrica de GPA. La regresión Lasso y las NN fueron los modelos más precisos en las tres configuraciones de predicción (con el uso de puntajes de cursos de los primeros dos o tres años, o GPA de semestre). La regresión Lasso, en particular, se destacó con valores de correlación (R) superiores a 0,99 y errores mínimos (RMSE y MAE).

Beaulac y Rosenthal (2019) estudiaron la capacidad de predecir si un estudiante completará su licenciatura y qué especialización elegirá mediante el uso de datos de los primeros semestres. De este modo, analizaron diez años de registros de la Universidad de Toronto, con más de 1,6 millones de calificaciones de 65 000 estudiantes. Emplearon el algoritmo RF para predecir el éxito académico y la elección de especialización. En consecuencia, este alcanzó un 91,19 % de precisión en predecir si un estudiante terminará su programa, lo que superó a la regresión lineal. Para la especialización, logró un 47,41 % de precisión, lo cual es relevante dado el número de opciones (71 especializaciones).

En el estudio de Sharma et al. (2023), se utilizaron modelos de aprendizaje automático para predecir el rendimiento académico de los estudiantes en función de la cantidad

de tiempo que dedicaron a participar en actividades extracurriculares. En este trabajo se implementaron tres algoritmos: RF, KNN y árboles de decisión. Con base en la precisión y las puntuaciones F1, se descubrió que los árboles de decisión con índice de Gini son el mejor algoritmo para predecir el éxito académico de los estudiantes logrando F1 0,84 y una precisión del 85 %.

El trabajo de Albaloooshi et al. (2019) extrae los logros de aprendizaje individualizados de los estudiantes a partir de datos de entrada (información del curso y resultados de las evaluaciones). De esta forma, predice el desempeño esperado en cursos futuros con base en los logros existentes en un conjunto de resultados estudiantiles. Se emplea regresión lineal simple y regresión lineal múltiple para determinar el desempeño esperado de los estudiantes en cursos futuros.

Musso et al. (2020) desarrollaron modelos NN para clasificar los promedios de calificaciones, retención académica y resultados de finalización del título. Las variables predictoras fueron las estrategias de aprendizaje y las estrategias para afrontar el proceso de estudios. La primera predijo mejor el GPA, mientras que la segunda fue determinante en la predicción de la finalización del grado. El mejor modelo para analizar estas variables resultó ser el de NN.

ElSharkawy et al. (2022) analizaron la empleabilidad en TI a partir de variables predictoras: la capacitación, habilidades blandas y habilidades técnicas. Asimismo, utilizaron algoritmos de árboles de decisión, NB, regresión logística, RF y SVM. Los árboles de decisión alcanzaron la mayor precisión con un 100 %, seguido de regresión logística y SVM, ambos con un 98 %. RF obtuvo una precisión del 97 %, mientras que NB obtuvo un 92 %. Los árboles de decisión también obtuvieron los mejores resultados en términos de *precision, recall* y *F1-score* (100 %).

Se han empleado técnicas de aprendizaje automático para analizar el tema de adicción en estudiantes universitarios mediante modelos predictivos (Hong et al., 2024). El estudio identifica el nivel de adicción del estudiante y, para ello, utiliza distintas variables predictoras: perfeccionismo, procrastinación, sentido de la vida, ansiedad, autocontrol, género y grado académico. Los métodos empleados son RF, SVM y la regresión logística. Los resultados señalan que el modelo RF alcanzó una precisión del 76,68 % en la predicción de la adicción al teléfono móvil en estudiantes universitarios. Perfeccionismo y procrastinación fueron los factores más influyentes. El modelo superó a otros como SVM y NB. En un estudio que utilizó técnicas de minería de datos, se analizó el impacto de la adicción al teléfono móvil sobre el desempeño académico, donde se emplearon métodos como SVM, NN y clasificadores NB (Chaudhury & Tripathy, 2018). El desempeño académico estuvo categorizado en tres niveles: bajo, medio y alto; se encontró correlación negativa significativa.

2. METODOLOGÍA

En este trabajo, se aplican tres métodos de aprendizaje automático para predecir el éxito académico de un estudiante que se matricula en un determinado número de asignaturas. Como predictores intervienen las siguientes variables: nivel de adicción a la pornografía, nivel de adicción al teléfono, la frecuencia con que se desbloquea el teléfono y el nivel de confianza en la herramienta ChatGPT.

2.1. Data

Se ha considerado un *dataset* de 4278 registros con información relacionada al uso de la tecnología y redes sociales en estudiantes de ocho universidades del Ecuador.

Para este experimento se seleccionaron las variables presentadas en la Tabla 1. Además, se creó una nueva variable que considera la diferencia entre la asignatura matriculado (P14_a) y asignatura aprobada (P14_b).

Tabla 1

Descripción de variables utilizadas

Código de la variable	Descripción	Valores
P06_a	Adicción al teléfono	0 a 9
P06_b	Pornografía	0 a 9
P08	Desbloqueo del teléfono	1 a 52
P12_b	Confianza en ChatGPT	0 a 9
Diff	Variable creada que resta la cantidad de asignaturas matriculadas (P14_a) y asignaturas aprobadas (P14_b)	0 a 7

Nota. Variables levantadas a través de encuesta.

En este trabajo se utilizó el lenguaje Python y Google Colab como entorno de programación.

2.2. Preprocesamiento

Luego de realizar una limpieza para eliminar registros que no presentaban algún valor o un valor negativo en las variables descritas en la parte superior, se obtuvo un total de 4223 registros.

Diff será la variable a predecir mediante el uso de clasificadores binarios. Se procedió a transformar los valores 1 a 7 (cantidad de materias reprobadas) a un valor de 1. De esta forma, se convierte en una variable dicotómica, donde el valor 0 representa a los alumnos que aprobaron todas las materias matriculadas y el valor 1 simboliza si ha reprobado una o más asignaturas.

Se realizó la división de la data y, para la etapa de entrenamiento, se utilizó el 70 % de la data y el 30 % se reservó para la evaluación.

Debido a que la variable a predecir Diff tiene 2437 registros para la clase 0 y 519 registros para la clase 1 (con respecto a la cantidad de datos de entrenamiento), es necesario balancear la data (ver la Tabla 2). Se ha considerado trabajar con la técnica *oversampling* que permite a la clase minoritaria igualarse en la cantidad de registros o de observaciones a la clase mayoritaria, de esta forma aumentaremos los datos en lugar de eliminar registros. En Python, el paquete *imbalanced-learn* contiene una serie de algoritmos para realizar *oversampling*, entre ellos SMOTE y RandomOverSampler.

Cabe recalcar que no hay un criterio certero que permita seleccionar si se trabajará con técnicas de *oversampling* o *undersampling* (Estabrooks et al., 2004), incluso la evaluación del desempeño de las técnicas varía de acuerdo con la métrica utilizada (Wainer, 2024). No obstante, se ha demostrado que las técnicas de *oversampling* son superiores con respecto a las de *undersampling* (García et al., 2020) e inclusive se ha demostrado que SMOTE trabaja mejor con datos que presentan una baja dimensionalidad (Blagus & Lusa, 2013), como es el caso del *dataset* utilizado en nuestro trabajo. Por lo antes expuesto, se ha decidido trabajar con SMOTE y considerar RandomOverSampler como otra posibilidad para la comparación. Este último se ha considerado, pues en el trabajo de Wongvorachan et al. (2023) se utilizó esta técnica con buenos resultados. Sobre estos algoritmos, por un lado, SMOTE es una técnica que genera observaciones sintéticas (Chawla et al., 2022), mientras que, por otro lado, RandomOverSampler forma nuevas muestras mediante una selección aleatoria con sustitución de las actuales disponibles (Imbalanced learn, 2014).

La Tabla 2 explica mejor el proceso:

Tabla 2

Registros por clase

Descripción de clases	Cantidad de registros
Variable Diff	Clase 0 = 3456 Clase 1 = 767
Variable Diff para entrenamiento	Clase 0 = 2437 Clase 1 = 519
Variable Diff para evaluación	Clase 0 = 1019 Clase 1 = 248
Variable Diff luego del balanceo con las dos técnicas de <i>oversampling</i>	Clase 0 = 2437 Clase 1 = 2437

Cabe recalcar que para el algoritmo *support vector machine* también se procedió a normalizar la data balanceada y de entrenamiento a través del uso de StandardScaler de Python.

2.3. Experimentación

Para el presente trabajo se consideraron tres algoritmos: DT, RF y SVM. Estos han sido ampliamente utilizados en trabajos que realizan clasificación, previo a realizar un balanceo de los datos (Kovács, 2019; Blagus & Lusa, 2013; Wongvorachan et al., 2023). Los árboles de decisión y RF fueron entrenados con la data original balanceada, mientras que SVM se entrenó con la data estandarizada luego del balanceo con StandardScaler de Python.

Para buscar los mejores parámetros se utilizó GridSearchCV de 5. La Tabla 3 muestra algunos de los parámetros considerados los mejores para cada uno de los algoritmos utilizados.

Tabla 3

Parámetros seleccionados con GridSearchCV

Algoritmo	Mejores parámetros
Balanceo con SMOTE	
Árbol de decisión	{'criterion': 'entropy', 'max_depth': 5}
Random forest	{'bootstrap': True, 'criterion': 'gini', 'max_depth': 5, 'max_features': 3, 'min_samples_leaf': 3, 'min_samples_split': 7, 'n_estimators': 100}
Support vector machine	{'C': 5, 'kernel': 'linear'}
Balanceo con RandomOverSampler	
Árbol de decisión	{'criterion': 'gini', 'max_depth': 5}
Random forest	{'bootstrap': True, 'criterion': 'gini', 'max_depth': 5, 'max_features': 3, 'min_samples_leaf': 3, 'min_samples_split': 5, 'n_estimators': 100}
Support vector machine	{'C': 5, 'kernel': 'linear'}

2.4. Evaluación

Para evaluar los modelos obtenidos con los tres algoritmos, se utilizaron las siguientes métricas: *precision*, *accuracy*, *recall* y *F1-score*; las fórmulas se muestran en la Tabla 4. Además, se crearon la matriz de confusión y el reporte de clasificación con ayuda de la librería sklearn.metrics.

Tabla 4

Métricas de evaluación

Métrica	Fórmula
<i>Precision</i>	$\frac{VP}{VP + FP}$
<i>Accuracy</i>	$\frac{VP + VN}{VP + FP + FN + VN}$
<i>Recall</i>	$\frac{VP}{VP + FN}$
<i>F1-score</i>	$2 * \frac{precision * recall}{precision + recall}$
VP: Verdaderos positivos, VN: Verdaderos negativos, FN: Falsos positivos, FN: Falsos negativos	

3. RESULTADOS

La Tabla 5 muestra los resultados obtenidos con los tres algoritmos anteriormente descritos.

Tabla 5

Métricas resultantes

Algoritmo	Precision	Accuracy	Recall	F1-score
SMOTE				
Árbol de decisión	0,52	0,64	0,53	0,52
Random forest	0,52	0,61	0,53	0,51
SVM	0,51	0,51	0,52	0,46
RandomOverSampler				
Árbol de decisión	0,53	0,54	0,55	0,49
Random forest	0,52	0,56	0,54	0,49
SVM	0,54	0,59	0,56	0,52

Como se puede observar, existen pequeñas diferencias entre los valores obtenidos en *precision*, *recall* y *F1-score* para los experimentos realizados con SMOTE y RandomOverSampler. Prácticamente, estas métricas se mantienen muy similares en todos los experimentos que se realizan con algoritmos basados en árboles. La gran diferencia se puede observar en los resultados obtenidos en *accuracy*, el cual muestra 0,64, el mejor resultado obtenido con SMOTE y con el algoritmo árbol de decisión. Cuando se aplica RandomOverSampler, SVM se destaca con un *accuracy* de 0,59.

Un punto importante a considerar con la métrica *accuracy* es que en datos no balanceados solo consideran esta métrica como desempeño de los algoritmos, lo cual puede

ser una estrategia errónea. Si el algoritmo predice correctamente la mayoría de los datos de la clase mayoritaria y se equivoca en la clase minoritaria, el valor de *accuracy* podría seguir siendo alto. Este tipo de métrica no considera los falsos negativos ni los falsos positivos; por el contrario, *recall* y *precision* trabajan con FN y FP, por lo que con mayor razón han sido utilizados.

Para aclarar la variedad en los resultados y considerando los aciertos (VP y VN) según la matriz de confusión, es posible observar que con SMOTE la clase mayoritaria consiguió mayores aciertos (71 %) con árboles de decisión, pero la clase minoritaria consiguió la mayoría de aciertos (54 %) con SVM. Con la técnica de RandomOverSampler, los resultados fueron contrarios, la clase mayoritaria obtuvo mayores aciertos (61 %) con SVM, pero la clase minoritaria (clase 1) consiguió la mayoría de aciertos (57 %) con árboles de decisión. Estos porcentajes se los puede obtener en base a las matrices de confusión que se presentan en la Tabla 6.

Tabla 6
Matriz de confusión e informe de clasificación

Algoritmo	Matriz de confusión	Informe de clasificación				
		precision	recall	F1-score	support	
Árbol de decisión	[728 291] [162 86]	SMOTE				
		0	0,82	0,71	0,76	1019
		1	0,23	0,35	0,28	248
		accuracy			0,64	1267
		macro avg.	0,52	0,53	0,52	1267
		weighted avg.	0,70	0,64	0,67	1267
		Random forest	[670 349] [148 100]	precision	recall	F1-score
0	0,82	0,66	0,73	1019		
1	0,22	0,40	0,29	248		
	accuracy			0,61	1267	
	macro avg.	0,52	0,53	0,51	1267	
	weighted avg.	0,70	0,61	0,64	1267	

SVM	[516 503] [114 134]	precision	recall	F1-score	support	
		0	0,82	0,51	0,63	1019
		1	0,21	0,54	0,30	248
		accuracy			0,51	1267
		macro avg.	0,51	0,52	0,46	1267
		weighted avg.	0,70	0,51	0,56	1267
RandomOverSampler						
Árbol de decisión	[545 474] [107 141]	precision	recall	F1-score	support	
		0	0,84	0,53	0,65	1019
		1	0,23	0,57	0,33	248
		accuracy			0,54	1267
		macro avg.	0,53	0,55	0,49	1267
		weighted avg.	0,72	0,54	0,59	1267
Random forest	[582 437] [123 125]	precision	recall	F1-score	support	
		0	0,83	0,57	0,68	1019
		1	0,22	0,50	0,31	248
		accuracy			0,56	1267
		macro avg.	0,52	0,54	0,49	1267
		weighted avg.	0,71	0,56	0,60	1267
SVM	[626 393] [124 124]	precision	recall	F1-score	support	
		0	0,83	0,61	0,71	1019
		1	0,24	0,50	0,32	248
		accuracy			0,59	1267
		macro avg.	0,54	0,56	0,52	1267
		weighted avg.	0,72	0,59	0,63	1267

Considerando las otras métricas, con RandomOverSampler se logran los valores más altos para *precision* y *recall*. Ahora, al igual que la métrica *accuracy*, es importante saber el comportamiento de los algoritmos con respecto a la clase minoritaria (ver la Tabla 6). Basados en el informe de clasificación, para la clase minoritaria *precision* tiene un valor de 0,24 con SVM y para *recall* esta misma clase presenta el valor más alto

de 0,57 para el árbol de decisión. Con este análisis se puede corroborar que los valores de las métricas varían dependiendo del algoritmo y de la técnica de balanceo, tal y como lo demuestra Wainer (2024).

4. CONCLUSIONES

Basados en los resultados obtenidos, se concluye que la elección del método de sobremuestreo puede tener un impacto significativo en el rendimiento de los algoritmos, especialmente en términos de *accuracy*. Los modelos basados en árboles parecen beneficiarse del uso de SMOTE, mientras que SVM muestra mejor desempeño con RandomOverSampler. En cuanto a las otras métricas, con RandomOverSampler se obtuvieron mejores resultados para *precision* y *recall*.

Con estos experimentos podemos corroborar lo que menciona Wainer (2024), que la evaluación del desempeño varía de acuerdo con la métrica utilizada. Además, hay que considerar que los resultados del balanceo dependen mucho del tipo de datos que se utiliza, del volumen de los datos, del porcentaje de desbalanceo de la data y de las técnicas de balanceo utilizadas (Newaz et al., 2022). En nuestro caso, hemos utilizado SMOTE y RandomOverSampler para balancear la data, pero hay trabajos cuyo objetivo ha sido demostrar el funcionamiento de varias técnicas (Estabrooks et al., 2004; Kovács, 2019) y que presentan una lista amplia de posibles técnicas a utilizar.

En investigaciones futuras, se plantea realizar más experimentos con otros algoritmos de clasificación, como aquellos basados en árboles. Además, se podría tratar con otras técnicas de balanceo como los híbridos (SMOTE-Tomek y SMOTE-ENN) o métodos basados en ensamblaje (UnderBagging, SMOTE-Bagging). También se propone la inclusión de otro tipo de variables que puedan ayudar a mejorar la clasificación.

REFERENCIAS

- Albalooshi, F., Alobaidy, H., & Ghanim, A. (2019). Mining students outcomes: An empirical study. *International Journal of Computing and Digital Systems*, 8(3), 229-241. <https://doi.org/10.12785/ijcnds/080303>
- Alghamdi, A. S., & Rahman, A. (2023). Data mining approach to predict success of secondary school students: A Saudi Arabian case study. *Education Sciences*, 13(3), 293. <https://doi.org/10.3390/educsci13030293>
- Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28, 905-971. <https://doi.org/10.1007/s10639-022-11152-y>

- Beaulac, C., & Rosenthal, J. S. (2019). Predicting university students' academic success and major using random forests. *Research in Higher Education*, 60, 1048-1064. <https://doi.org/10.1007/s11162-019-09546-y>
- Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14(106), 1-16. <https://doi.org/10.1186/1471-2105-14-106>
- Cai, Q., Lin, Y., & Yu, Z. (2023). Factors influencing learner attitudes towards ChatGPT-assisted language learning in higher education. *International Journal of Human-Computer Interaction*. Publicación anticipada en línea. <https://doi.org/10.1080/10447318.2023.2261725>
- Chaudhury, P., & Tripathy, H. K. (2018). A study on impact of smartphone addiction on academic performance. *International Journal of Engineering and Technology*, 7(2.6), 50-53. <https://doi.org/10.14419/ijet.v7i2.6.10066>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/jair.953>
- Chen, Y.-C. (2006). A study of comparing the use of augmented reality and physical models in chemistry education. *VRCIA'06: Virtual Reality Continuum and Its Applications 2006*, 1, 369-372. <https://doi.org/10.1145/1128923.1128990>
- Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, 28, 12039-12057. <https://doi.org/10.1007/s10639-023-11672-1>
- Cui, W., Sangsongfar, A., & Amdee, N. (2024). A comparative study of the applicability of regression models in predicting student academic performance. *Naresuan University Engineering Journal*, 19(1), 39-49. <https://ph01.tci-thaijo.org/index.php/nuej/article/view/255799>
- Elkhodr, M., Gide, E., Wu, R., & Darwish, O. (2023). ICT students' perceptions towards ChatGPT: An experimental reflective lab analysis. *STEM Education*, 3(2), 70-88. <https://doi.org/10.3934/steme.2023006>
- ElSharkawy, G., Helmy, Y., & Yehia, E. (2022). Employability prediction of information technology graduates using machine learning algorithms. *International Journal of Advanced Computer Science and Applications*, 13(10), 359-367. <https://doi.org/10.14569/IJACSA.2022.0131043>
- Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, 20(1), 18-36. <https://doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x>
- García, V., Sánchez, J. S., Marqués, A. I., Florencia, R., & Rivera, G. (2020). Understanding the apparent superiority of over-sampling through an analysis of local

- information for class-imbalanced data. *Expert Systems with Applications*, 158, artículo 113026. <https://doi.org/10.1016/j.eswa.2019.113026>
- Gutiérrez-Aguilar, O., Huarsaya-Rodríguez, E., & Duche-Pérez, A. (2024). The mediating effect of academic performance on ChatGPT satisfaction in university students. En G. F. Olmedo Cifuentes, D. G. Arcos Avilés y H. V. Lara Padilla (Eds.), *Emerging research in intelligent systems – Proceedings of the CIT 2023* (v. 2, pp. 353-365). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-52258-1_26
- Hellas, A., Ihtantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., & Liao, S. N. (2018). Predicting academic performance: A systematic literature review. En G. Rößling y B. Scharlau (Eds.), *ITiCSE 2018 companion: Proceedings companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (pp. 175-199). Association for Computing Machinery. <https://doi.org/10.1145/3293881.3295783>
- Hong, Y., Rong, X., & Liu, W. (2024). Construction of influencing factor segmentation and intelligent prediction model of college students' cell phone addiction model based on machine learning algorithm. *Heliyon*, 10(8), e29245. <https://doi.org/10.1016/j.heliyon.2024.e29245>
- Kovács, G. (2019). An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. *Applied Soft Computing*, 83, artículo 105662. <https://doi.org/10.1016/j.asoc.2019.105662>
- Musso, M. F., Rodríguez, C. F., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. *Higher Education*, 80, 875-894. <https://doi.org/10.1007/s10734-020-00520-7>
- Nachouki, M., Mohamed, E. A., Mehdi, R., & Abou Naaj, M. (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. *Trends in Neuroscience and Education*, 33, 100214. <https://doi.org/10.1016/j.tine.2023.100214>
- Nayak, P., Vaheed, S., Gupta, S., & Mohan, N. (2023). Predicting students' academic performance by mining the educational data through machine learning-based classification model. *Education and Information Technologies*, 28, 14611-14637. <https://doi.org/10.1007/s10639-023-11706-8>
- Newaz, A., Hassan, S., & Haq, F. S. (2022). An empirical analysis of the efficacy of different sampling techniques for imbalanced classification. *arXiv*. Publicación anticipada en línea. <https://doi.org/10.48550/arXiv.2208.11852>
- Sharma, N., Appukutti, S., Garg, U., Mukherjee, J., & Mishra, S. (2023). Analysis of student's academic performance based on their time spent on extra-curricular activities

using machine learning techniques. *International Journal of Modern Education and Computer Science*, 15(1), 46-57. <https://doi.org/10.5815/ijmecs.2023.01.04>

Imbalanced learn. (2014). *RandomOverSampler*. https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.RandomOverSampler.html#imblearn.over_sampling.RandomOverSampler

Wainer, J. (2024). An empirical evaluation of imbalanced data strategies from a practitioner's point of view. *Expert Systems with Applications*, 256, 124863. <https://doi.org/10.1016/j.eswa.2024.124863>

Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information*, 14(1), 54. <https://doi.org/10.3390/info14010054>