



INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

MNA - Maestría Inteligencia Artificial Aplicada

Proyecto.

A01796362 Eduardo Perez Carrillo

A01365006 Oscar Gerardo Álvarez Mejía

A01797139 Deménard Gardy Armand

A01796858 Luis Esteban Barranco Guida

A01796859 Fernando Abel Leal Villavicencio

Docente Titular:

Dr. Gerardo Rodríguez Hernández

Maestro Ricardo Valdez Hernández

Docente Asistente:

Maestra María Mylen Treviño Elizondo

Docente Tutor:

Livan Israel Valdes Esquivel

Operaciones de aprendizaje automático

12 octubre 2025

Índice

Índice	2
Índice de Ilustraciones	3
Índice de Tablas	4
Capítulo 1. Introducción	5
Capítulo 2. Planteamiento del problema.....	6
Capítulo 3. Marco teórico.....	8
Capítulo 4. Metodología (Parcial).....	13
Capítulo 5: Resultados	17
Capítulo 6. Conclusiones y Recomendaciones (parciales)	27
Referencias	29

Índice de Ilustraciones

Ilustración 1 Ejemplo de Matriz de Confusión	11
Ilustración 2 Grafico de Frecuencia de la Categoría de Performance	21
Ilustración 3 Curvas de Aprendizaje - XGBoost	22
Ilustración 4 Curvas de Aprendizaje - Regresión Logística	23
Ilustración 5 Curvas de Aprendizaje - SVM.....	24
Ilustración 6 Curvas de Aprendizaje - Regresión Logística Ordinal	25
Ilustración 7 Repositorio de GitHub.....	26

Índice de Tablas

Tabla 1. Roles y Tareas Fase 1	14
Tabla 2 Requerimientos ML Canvas Parte 1.....	18
Tabla 3 Requerimientos ML Canvas Parte 2.....	19
Tabla 4 Requerimientos ML Canvas Parte 3.....	19
Tabla 5 Frecuencia de la Categoría de Performance	21

Capítulo 1. Introducción

La evaluación escolar constituye uno de los pilares fundamentales del proceso educativo, ya que permite medir no solo el nivel de conocimiento adquirido por los estudiantes, sino también la efectividad de las estrategias pedagógicas implementadas por las instituciones [1]. A través de las calificaciones, los docentes y autoridades académicas obtienen información clave para identificar avances, rezagos y necesidades de apoyo específicas [2]. Sin embargo, el rendimiento académico no depende únicamente de las capacidades cognitivas de los estudiantes, sino que está influido por una amplia gama de factores sociales, económicos y personales que determinan su desempeño dentro y fuera del aula [3].

En este sentido, mantener un registro sistemático y analítico de estos factores se vuelve esencial para comprender la complejidad del aprendizaje [4]. Variables como el nivel educativo de los padres, el tipo de institución de procedencia, el idioma de instrucción o incluso el entorno socioeconómico pueden impactar de manera significativa en las calificaciones finales [4]. Analizar dichos elementos desde un enfoque cuantitativo y con herramientas basadas en datos permite no solo observar patrones históricos, sino también anticipar comportamientos futuros. Así, la adopción de medidas orientadas por evidencia, basadas en criterios objetivos y en el análisis de datos académicos, contribuye a diseñar estrategias de intervención más efectivas, equitativas y personalizadas para cada grupo de estudiantes [5].

En los últimos años, la integración de técnicas de *machine learning* en el ámbito educativo ha abierto nuevas oportunidades para el análisis predictivo del rendimiento estudiantil [6]. Sin embargo, el desarrollo de modelos de predicción efectivos requiere más que un entrenamiento algorítmico [7]: demanda procesos estructurados de gestión, monitoreo y mejora continua. En este contexto, el enfoque MLOps (*Machine Learning Operations*) se presenta como una metodología esencial para garantizar la trazabilidad, la reproducibilidad y la actualización constante de los modelos predictivos [8]. Al incorporar prácticas de ingeniería de software en el ciclo de vida del aprendizaje automático, MLOps permite automatizar la recolección de datos, el entrenamiento y la validación de modelos, asegurando que los resultados sean consistentes y útiles para la toma de decisiones educativas [9].

De esta manera, la combinación entre la analítica de datos educativos y la implementación de flujos MLOps ofrece una base sólida para transformar la gestión académica tradicional en un proceso inteligente y proactivo [10]. Este tipo de sistemas no solo predicen el rendimiento académico de los estudiantes, sino que además facilitan la identificación temprana de casos de riesgo, la optimización de recursos institucionales y la creación de políticas de apoyo fundamentadas en evidencia [11].

Finalmente, el presente documento se estructura de la siguiente manera: en el Capítulo dos se presenta el planteamiento del problema, donde se define la necesidad de implementar un modelo predictivo y se delimitan sus objetivos específicos. En el Capítulo tres se expone el marco teórico que sustenta la investigación, abarcando los conceptos relacionados con el aprendizaje automático, la analítica educativa y la metodología MLOps. El Capítulo cuatro describe la metodología empleada, incluyendo la definición del *dataset*, las variables analizadas y el flujo de trabajo propuesto. En el Capítulo cinco se presentan los resultados obtenidos durante la fase de análisis y modelado, mientras que el Capítulo seis expone las conclusiones generales y las recomendaciones derivadas del estudio.

Capítulo 2. Planteamiento del problema

Durante las últimas décadas, los sistemas educativos han experimentado una transformación significativa impulsada por la globalización, la digitalización y la creciente demanda de capital humano altamente calificado [12]. Este cambio ha puesto de manifiesto la necesidad de contar con mecanismos más precisos y objetivos para evaluar el rendimiento de los estudiantes y anticipar los factores que inciden en su éxito académico. En el contexto de la educación superior, donde los procesos de admisión y permanencia resultan determinantes para la trayectoria profesional, disponer de herramientas analíticas que permitan identificar patrones de desempeño se ha convertido en un componente esencial de la gestión educativa moderna [13].

Tradicionalmente, la evaluación del rendimiento estudiantil se ha limitado al análisis descriptivo de las calificaciones obtenidas, sin considerar de forma integral las variables que las condicionan. Sin embargo, la evidencia acumulada en diversos estudios ha demostrado que el desempeño académico está estrechamente relacionado con factores sociales, económicos y familiares, tales como el nivel educativo de los padres, la ocupación del hogar, el tipo de escuela de procedencia o el idioma de instrucción. Estas variables, aunque registradas en los sistemas institucionales, rara vez se aprovechan con fines analíticos o predictivos. El resultado es una pérdida de información valiosa que podría contribuir a mejorar la toma de decisiones en materia de becas, tutorías y programas de apoyo académico [14].

En particular, en los procesos de admisión universitaria, los responsables de las políticas educativas enfrentan el reto de evaluar grandes volúmenes de datos provenientes de aspirantes con contextos muy diversos. Identificar a los estudiantes con mayor probabilidad de éxito o de riesgo académico de forma temprana permitiría diseñar estrategias de acompañamiento más efectivas y equitativas. Sin embargo, esta tarea resulta compleja cuando se realiza de manera manual o utilizando herramientas

estadísticas tradicionales, pues éstas no siempre capturan las relaciones no lineales ni las interacciones entre las variables involucradas.

En este escenario, el aprendizaje automático (*machine learning*) surge como una alternativa sólida para el análisis y la predicción del rendimiento académico, al ofrecer modelos capaces de aprender de los datos históricos y generar estimaciones con alto nivel de precisión. No obstante, la implementación de estos modelos plantea desafíos relacionados con la gestión de datos, la reproducibilidad de resultados y la actualización continua del sistema.

Es aquí donde la metodología MLOps cobra relevancia, al proporcionar un marco estructurado que integra prácticas de ingeniería de software con los flujos de trabajo del aprendizaje automático. Gracias a MLOps, es posible establecer procesos estandarizados para la recolección, entrenamiento, validación, despliegue y monitoreo de modelos predictivos de forma automatizada y sostenible en el tiempo.

El problema que motiva esta investigación radica en la ausencia de un sistema predictivo que permita estimar, con base en información demográfica, académica y socioeconómica, el nivel de rendimiento esperado de los aspirantes a instituciones de educación superior. Actualmente, las decisiones en torno a la admisión o el apoyo académico se sustentan principalmente en criterios cuantitativos simples como el promedio de calificaciones o el resultado de exámenes sin aprovechar el potencial analítico de los datos disponibles. Esta limitación reduce la capacidad de las instituciones para detectar oportunamente a los estudiantes en riesgo de bajo desempeño y para diseñar políticas educativas más inclusivas y basadas en evidencia.

Para abordar esta problemática, se dispone de un conjunto de datos que integra información demográfica, académica y socioeconómica de aspirantes a instituciones de educación superior, el cual permitirá posteriormente analizar los factores asociados al desempeño académico mediante técnicas de clasificación supervisada.

Por tanto, la presente investigación propone el desarrollo de un modelo de clasificación supervisada que prediga el nivel de rendimiento académico de los estudiantes (*Excellent, Good, Average, Very Good*) a partir de variables demográficas, educativas y familiares.

Este planteamiento no busca únicamente desarrollar un modelo de predicción, sino también sentar las bases para un sistema sostenible, escalable y ético de analítica educativa, capaz de contribuir a la mejora de la calidad y la equidad en los procesos de enseñanza superior.

Capítulo 3. Marco teórico

3.1 Educación superior y evaluación de aspirantes

Las instituciones de educación superior constituyen un eje estratégico para el desarrollo social, científico y económico de cualquier país [15]. En ellas convergen los esfuerzos formativos orientados a la creación de capital humano capaz de responder a los desafíos tecnológicos y productivos contemporáneos [16]. Sin embargo, la creciente demanda de acceso a la universidad ha exigido la implementación de mecanismos de selección más precisos, transparentes y basados en criterios objetivos [17].

Los exámenes de admisión se han consolidado como una de las herramientas más utilizadas para valorar las competencias cognitivas y el potencial académico de los aspirantes [18]. Estos instrumentos buscan determinar, de manera estandarizada, el nivel de conocimientos y habilidades requeridas para cursar estudios universitarios. No obstante, la interpretación de sus resultados no siempre considera la diversidad de factores que influyen en el desempeño: las diferencias socioeconómicas, el tipo de formación previa o la disponibilidad de recursos educativos pueden impactar significativamente en los puntajes obtenidos [19].

En este contexto, la analítica educativa surge como un enfoque innovador que permite aprovechar los datos institucionales para comprender mejor los patrones de aprendizaje y desempeño de los estudiantes. Mediante la aplicación de técnicas estadísticas y de inteligencia artificial, es posible transformar la información acumulada en conocimiento útil para la gestión académica, la mejora de los procesos de admisión y la formulación de políticas de equidad educativa [20].

3.2 Aprendizaje automático

El aprendizaje automático (*Machine Learning*) se define como una rama de la inteligencia artificial que permite a los sistemas aprender patrones a partir de los datos sin ser explícitamente programados [21]. En lugar de depender de reglas fijas, los algoritmos de *machine learning* construyen modelos capaces de generalizar comportamientos y realizar predicciones basadas en ejemplos previos [22].

En el ámbito educativo, esta disciplina ha adquirido una relevancia creciente debido a su capacidad para identificar tendencias ocultas en los datos de los estudiantes, predecir su rendimiento y proponer estrategias personalizadas de apoyo académico [23]. Los modelos de *machine learning* pueden procesar grandes volúmenes de información y encontrar correlaciones entre variables que, de otra forma, pasarían desapercibidas [24].

Entre las aplicaciones más comunes destacan la predicción del desempeño académico, la detección de abandono escolar, la evaluación del aprendizaje adaptativo y la recomendación de cursos o tutorías personalizadas. Estas herramientas apoyan la toma

de decisiones institucionales y contribuyen a la mejora continua de los procesos formativos [25].

3.3 Aprendizaje supervisado

Dentro del campo del aprendizaje automático, el aprendizaje supervisado representa la categoría más utilizada para tareas predictivas. Este enfoque se basa en el uso de un conjunto de datos etiquetados, es decir, donde se conoce la variable de salida o *target*, para entrenar un modelo que aprenda la relación entre las entradas y los resultados esperados [26].

Los algoritmos supervisados se dividen principalmente en dos tipos: clasificación y regresión. En la clasificación, el objetivo es asignar una etiqueta a cada observación. En la regresión, en cambio, se busca estimar un valor continuo [27].

Algunos de los modelos más utilizados en clasificación incluyen los árboles de decisión, los bosques aleatorios (*Random Forest*), los métodos de ensamble como *Gradient Boosting* (*XGBoost* o *LightGBM*) y la regresión logística, cada uno con ventajas particulares en términos de interpretabilidad, velocidad de entrenamiento y precisión [28].

3.4 Datasets y analítica educativa

El concepto de *dataset* se refiere a un conjunto estructurado de datos que almacena información sobre entidades o eventos específicos, organizada en filas (instancias) y columnas (atributos) [29]. En el contexto del aprendizaje automático, la calidad, representatividad y consistencia del *dataset* son factores determinantes para la efectividad del modelo predictivo [30].

Los datasets educativos suelen contener información de tipo demográfico (género, edad, procedencia), académico (promedios, resultados de exámenes, tipo de institución) y socioeconómico (ocupación de los padres, nivel de ingreso, tipo de vivienda, etc.). La preparación del *dataset* implica diversas etapas: limpieza de datos, manejo de valores faltantes, normalización de variables y codificación de categorías [29][30].

Además, es fundamental garantizar la ética en el tratamiento de los datos, preservando la privacidad de los estudiantes y evitando sesgos en la interpretación de los resultados. Un modelo de predicción educativo no debe replicar desigualdades sociales, sino contribuir a mitigarlas mediante un análisis objetivo y transparente [31].

3.5 MLOps: Integración y operación de modelos de aprendizaje automático

El término MLOps combina las prácticas de *Machine Learning* y *DevOps* para gestionar el ciclo de vida completo de los modelos predictivos de forma automatizada, reproducible y escalable [32]. Su objetivo es facilitar la colaboración entre científicos de datos, ingenieros y analistas, garantizando que los modelos no solo se desarrollen

correctamente, sino que también se desplieguen, supervisen y actualicen de manera continua [33].

El flujo de trabajo típico de MLOps se compone de las siguientes etapas [34][35][36]:

1. Definición de requerimientos y recolección de datos: identificación de las fuentes, formatos y variables relevantes.
2. Preparación y limpieza de datos: normalización, imputación de valores faltantes y codificación de atributos categóricos.
3. Entrenamiento y validación del modelo: selección de algoritmos, ajuste de hiperparámetros y evaluación mediante métricas como *Accuracy*, *Precision*, *Recall* y *F1-Score*.
4. Registro y versionamiento: almacenamiento controlado de modelos, *datasets* y configuraciones mediante herramientas como MLflow o DVC.
5. Despliegue: implementación del modelo en un entorno de producción, ya sea en servidores locales o en la nube.
6. Automatización con contenedores: uso de tecnologías como Docker o Kubernetes para asegurar portabilidad, consistencia y escalabilidad en los entornos de ejecución.
7. Monitoreo y mantenimiento: seguimiento de métricas en tiempo real, detección de *data drift* y reentrenamiento periódico.
8. Gobernanza y auditoría: trazabilidad de los resultados, control de versiones y cumplimiento de criterios éticos y de privacidad.

En conjunto, el enfoque MLOps transforma el desarrollo de modelos en un proceso continuo, colaborativo y automatizado, garantizando que las predicciones se mantengan vigentes y confiables a lo largo del tiempo.

3.6 Evaluación de métricas de desempeño

La evaluación del desempeño de un modelo de aprendizaje automático constituye una etapa fundamental dentro del proceso de desarrollo, ya que permite determinar su capacidad para realizar predicciones correctas y generalizar el conocimiento aprendido a nuevos datos. En el caso de los modelos de clasificación supervisada las métricas más utilizadas para medir la calidad de las predicciones son *Accuracy*, *Precision*, *Recall* y *F1-Score*, además del uso de la matriz de confusión como herramienta de diagnóstico visual y cuantitativo [37][38].

La matriz de confusión es una representación tabular que muestra el número de aciertos y errores cometidos por el modelo en cada categoría. Cada fila de la matriz representa

las instancias reales, mientras que cada columna representa las predicciones realizadas. A partir de esta matriz es posible identificar con precisión en qué clases el modelo tiende a confundirse, proporcionando una visión más detallada que la que ofrecen las métricas agregadas [39].

Verdadera	TN	FN
	FP	TP
Predicción		

Ilustración 1 Ejemplo de Matriz de Confusión

El cuadrante TN (*True Negative*) corresponde a los casos en los que el modelo predijo correctamente la ausencia de la característica o clase de interés; es decir, identificó como “negativos” aquellos casos que realmente lo eran. Por su parte, el FN (*False Negative*) refleja los errores en los que el modelo no detectó una instancia positiva, clasificándola incorrectamente como negativa. En contraste, el FP (*False Positive*) muestra los casos en los que el modelo predijo erróneamente la presencia de la clase cuando en realidad no existía, mientras que el TP (*True Positive*) representa las predicciones acertadas en las que el modelo identificó correctamente una instancia positiva. En conjunto, estos cuatro valores permiten evaluar con precisión el equilibrio entre aciertos y errores, ofreciendo una visión integral del comportamiento del clasificador y de su capacidad para distinguir entre las distintas categorías analizadas [39].

La *accuracy* representa la proporción de predicciones correctas sobre el total de observaciones analizadas. Es una métrica global que indica, en términos generales, qué tan bien clasifica el modelo. Matemáticamente se expresa como [40]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Aunque es una medida intuitiva, su interpretación debe realizarse con precaución en *datasets* desbalanceados, ya que puede ocultar deficiencias en clases minoritarias.

La *precision* mide la proporción de predicciones positivas que fueron realmente correctas. Indica la confiabilidad del modelo cuando clasifica una instancia en una categoría específica. Se define como [40]:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Un valor alto de precisión implica que el modelo comete pocos falsos positivos. En contextos educativos, una alta precisión significa que la mayoría de los estudiantes identificados como de “bajo rendimiento”, efectivamente lo son.

El *recall* evalúa la capacidad del modelo para detectar correctamente todos los casos pertenecientes a una clase. Es decir, mide cuántos elementos relevantes fueron realmente recuperados. Un valor alto de *recall* indica que el modelo logra identificar la mayoría de los casos de interés [41]:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

El *F1-Score* combina *precision* y *recall* en una sola métrica, equilibrando ambos indicadores mediante su media armónica [41]:

$$F1 = 2 \frac{Precision * Recall}{Precision + recall} \quad (4)$$

Esta métrica resulta especialmente útil cuando el *dataset* está desbalanceado o cuando tanto los falsos positivos como los falsos negativos son relevantes. Un valor de *F1-Score* cercano a 1 indica un modelo equilibrado en su capacidad de identificar correctamente las clases.

Capítulo 4. Metodología (Parcial)

La presente investigación se enmarca como aplicada, cuantitativa, no experimental y de alcance exploratorio–descriptivo. El objetivo específico es: desarrollar un modelo de clasificación supervisada que prediga el nivel de rendimiento académico de los estudiantes (*Excellent, Good, Average, Very Good*) a partir de variables demográficas, educativas y familiares.

El *dataset* utilizado en esta investigación corresponde a una versión modificada del dataset “Student Performance on an Entrance Examination” [42] recopilado por el Prof. *Jiten Hazarika*. Este dataset contiene información de los aspirantes que participaron en el examen de admisión a instituciones de educación médica en el estado de Assam, India. La base de datos integra variables demográficas, académicas y socioeconómicas, entre las que destacan el género, el tipo de institución educativa de procedencia, el idioma de instrucción, las calificaciones obtenidas en los niveles de educación secundaria y media superior, así como la ocupación de los padres. La versión utilizada en este estudio fue adaptada y normalizada para su análisis mediante técnicas de aprendizaje supervisado, garantizando la consistencia en las etiquetas de clasificación y la integridad de las variables relevantes para la predicción del rendimiento académico.

El *dataset* empleado para el entrenamiento y análisis del modelo está conformado por un total de 679 registros y 13 columnas. Cada registro corresponde a un aspirante que participó en el examen de admisión a instituciones de educación médica del estado de Assam, India.

Está compuesto por trece variables, distribuidas entre atributos demográficos, académicos y socioeconómicos. Las variables incluidas son: *Performance*, que representa el nivel de rendimiento académico del estudiante; *Gender*, que indica el sexo del aspirante; *Caste*, asociada al grupo social de pertenencia; *coaching*, que especifica si el estudiante recibió preparación adicional y en qué modalidad; *time*, que refleja la duración o periodo de dicho apoyo; *Class_ten_education* y *twelve_education*, que describen el tipo de institución educativa cursada en los niveles de secundaria y preparatoria, respectivamente; *medium*, que señala el idioma de instrucción; *Class_X_Percentage* y *Class_XII_Percentage*, que registran los porcentajes de calificación obtenidos en cada nivel; *Father_occupation* y *Mother_occupation*, que detallan la ocupación de los padres; y finalmente *mixed_type_col*, una columna auxiliar que contiene valores heterogéneos derivados del proceso de recopilación original.

A continuación, se presenta la descripción detallada del proceso metodológico correspondiente a la Fase 1, en la cual se documentan los requerimientos del proyecto, se realiza la preparación y exploración del conjunto de datos, y se construye el modelo predictivo inicial que servirá como base para las siguientes etapas del ciclo de aprendizaje automático.

4.1 Fase 1: Análisis y diseño del modelo predictivo.

En esta etapa, el *Subject Matter Expert* (SME) asume un rol central al coordinar las actividades del equipo y distribuir la carga de trabajo conforme a las competencias técnicas y responsabilidades de cada integrante. Su participación asegura la correcta alineación entre los requerimientos del dominio educativo y el desarrollo técnico del modelo, promoviendo una ejecución ordenada, colaborativa y orientada a resultados verificables. A continuación, se presenta la tabla que detalla la asignación de funciones y tareas específicas dentro de esta fase.

Tabla 1. *Roles y Tareas Fase 1*

Rol	Nombre	Tareas
<i>Subject Matter Expert</i>	Eduardo Perez	Comprender el problema a resolver con el <i>dataset</i> . Colaborar en la creación y documentación de los requerimientos (ML Canvas). Definir junto al equipo cuál será el objetivo del modelo Proponer cómo una solución basada en ML puede aportar valor a estudiantes. Análisis de Resultados de Fase 1.
<i>Data Scientist</i>	Fernando Leal	Analizar los datos con técnicas de <i>Exploratory Data Analysis</i> (EDA). Detectar patrones, relaciones y valores atípicos. Preparar los datos aplicando transformaciones como normalización o codificación de categorías. Probar diferentes algoritmos de <i>Machine Learning</i> . Evaluar los modelos con métricas adecuadas.
<i>Data Engineer</i>	Oscar Álvarez	Encargarse de la limpieza de datos (eliminar valores nulos, duplicados o inconsistentes). Organizar y preparar los datos para que puedan usarse en los modelos. Documentar los cambios realizados para asegurar la trazabilidad.
<i>DevOps</i>	Deménard Gardy Armand	Apoyar en la organización del repositorio (estructura de carpetas, ramas en Git). Revisar que todo lo que se haga pueda repetirse de forma clara y ordenada. Colaborar en la documentación del proceso.
<i>Machine Learning Architect</i>	Esteban Barranco	Diseñar la ruta de trabajo del proyecto (desde los datos iniciales hasta la evaluación del modelo). Coordinar la colaboración entre el <i>Data Engineer</i> y el <i>Data Scientist</i> . Garantizar que el trabajo realizado en esta fase pueda crecer y adaptarse a etapas posteriores.

4.1.1 Análisis de Requerimientos

Una vez distribuida la carga de trabajo, el SME procede a la documentación de los requerimientos del proyecto utilizando el formato *Machine Learning Canvas* (ML Canvas). Esta herramienta permite estructurar de manera clara los elementos clave del sistema de aprendizaje automático, incluyendo la definición del problema, las fuentes de datos, las variables relevantes, los criterios de evaluación y la propuesta de valor. Su aplicación facilita una visión integral y coherente del proyecto, asegurando que cada decisión metodológica esté alineada con los objetivos del modelo predictivo y con las necesidades institucionales identificadas.

4.1.2 Manipulación y preparación de datos

Durante esta etapa se llevarán a cabo las tareas de limpieza, estandarización y control de versiones del conjunto de datos, con el propósito de preparar la información para su posterior análisis y modelado. En primer lugar, se implementará un proceso de depuración y validación de los registros, orientado a identificar y corregir inconsistencias en las variables, garantizando la uniformidad de los valores categóricos mediante operaciones de normalización textual y estandarización de etiquetas. Asimismo, se atenderá el tratamiento de valores nulos o ausentes, aplicando estrategias estadísticas que permitan sustituirlos por valores representativos de acuerdo con la naturaleza de cada variable, manteniendo la coherencia interna del conjunto de datos.

Adicionalmente, se contempla la eliminación de columnas o atributos no relevantes para los objetivos de la investigación, optimizando así la estructura del *dataset* y reduciendo el ruido que pudiera afectar el desempeño del modelo. Todo este proceso será documentado de forma sistemática, asegurando que las transformaciones realizadas puedan ser replicadas y auditadas en fases posteriores del proyecto. Finalmente, se gestionará el versionado de los artefactos de datos a través de herramientas especializadas, con el fin de garantizar la trazabilidad, transparencia y reproducibilidad dentro del ciclo de vida del proyecto MLOps.

Para el desarrollo de esta tarea se utilizará el lenguaje de programación Python, dado su amplio uso en proyectos de análisis de datos y su compatibilidad con herramientas de ciencia computacional.

En esta etapa se emplearán principalmente las librerías Pandas y NumPy. La primera permitirá estructurar y manipular los datos a través de marcos tipo *DataFrame*, facilitando la organización, depuración y normalización de las variables. Por su parte, NumPy se utilizará para realizar operaciones numéricas y manejar arreglos multidimensionales, optimizando el procesamiento interno de los datos.

4.1.3 Exploración y preprocesamiento de datos

Para esta fase se empleará el lenguaje de programación Python, apoyado principalmente en las bibliotecas *Pandas* y *Matplotlib*, que permiten realizar la exploración, descripción y visualización del conjunto de datos. A través de *Pandas*, se cargará la base de datos y se ejecutarán operaciones destinadas a examinar la estructura del *dataset*, incluyendo la cantidad de registros y variables, los nombres de las columnas, los tipos de datos y una vista preliminar de los primeros y últimos registros. Esto permitirá verificar la integridad de la información y detectar posibles inconsistencias o valores atípicos antes de proceder con el preprocesamiento.

Por su parte, mediante *Matplotlib* se generarán representaciones gráficas descriptivas que facilitarán la interpretación visual de las distribuciones y patrones presentes en las variables. Además, las visualizaciones permitirán analizar la distribución de la variable objetivo (*Performance*), así como la relación entre las características académicas y demográficas de los estudiantes.

4.1.4 Versionado de Datos.

Para el versionado de datos se empleará la herramienta *Weights & Biases*, debido a su capacidad para registrar, rastrear y gestionar de manera centralizada las distintas versiones de los conjuntos de datos utilizados durante el desarrollo del modelo. Su integración permitirá documentar cada modificación realizada en las etapas de limpieza, transformación y preparación, garantizando así la trazabilidad, reproducibilidad y transparencia del proceso metodológico. De esta forma, se asegurará que los resultados obtenidos puedan ser verificados y replicados en fases posteriores del ciclo MLOps.

4.1.5 Construcción, ajuste y evaluación de Modelos de Machine Learning.

En esta etapa se llevará a cabo la construcción, ajuste y evaluación del modelo de aprendizaje supervisado, con el propósito de predecir el nivel de rendimiento académico de los estudiantes a partir de sus características demográficas, educativas y familiares. Para ello, se empleará el lenguaje de programación Python, junto con las bibliotecas *Pandas*, *NumPy*, *Scikit-learn*, *XGBoost*, *MLflow* y *Weights & Biases* (W&B), las cuales conforman un entorno robusto para el procesamiento de datos, la creación de modelos predictivos y el control de versiones.

El proceso metodológico iniciará con la ingeniería de características (*Feature Engineering*), en la cual se aplicarán transformaciones destinadas a optimizar la representación de los datos antes del entrenamiento del modelo. Se utilizarán estructuras como *Pipeline* y *ColumnTransformer* para organizar las etapas de preprocesamiento y aplicar operaciones específicas según el tipo de variable. Las variables numéricas serán

tratadas mediante técnicas de imputación estadística (*SimpleImputer*) utilizando la mediana como valor representativo, mientras que las variables categóricas se codificarán mediante *OrdinalEncoder* y *OneHotEncoder*, con el fin de traducir la información cualitativa en formatos adecuados para el modelo. Este flujo permitirá estandarizar los datos y garantizar su compatibilidad con el algoritmo seleccionado.

Una vez finalizada la fase de preprocesamiento, se procederá a la construcción, entrenamiento y evaluación comparativa de cuatro modelos de clasificación supervisada, con el propósito de determinar cuál ofrece el mejor desempeño en la predicción del nivel de rendimiento académico de los estudiantes. Los modelos seleccionados para esta etapa serán: XGBoost, Regresión Logística (Logistic Regression), Máquinas de Vectores de Soporte (SVC) y un modelo de Regresión Logística Ordinal, elegidos por su efectividad en tareas de clasificación multiclase y su capacidad para manejar datos categóricos y ordinales.

El conjunto de datos será dividido en 80% para entrenamiento y 20% para validación, garantizando un proceso de evaluación equilibrado que reduzca el riesgo de sobreajuste. Para medir el rendimiento de cada modelo se emplearán las métricas Accuracy, Precision, Recall y F1-Score, con el fin de valorar tanto la exactitud global como el desempeño individual por categoría.

Adicionalmente, se generarán las curvas de aprendizaje para cada uno de los modelos con el objetivo de analizar el comportamiento del error de entrenamiento y validación a medida que aumenta el tamaño del conjunto de datos. Estas gráficas permitirán evaluar la capacidad de generalización de los algoritmos y detectar posibles indicios de sobreajuste o subajuste, proporcionando una comprensión más profunda sobre la estabilidad y el rendimiento de cada enfoque antes de su integración en el pipeline MLOps.

4.1.6 Control de Versiones

Con el propósito de asegurar la organización, trazabilidad y control de versiones de todos los componentes desarrollados, se creará un repositorio del proyecto en la plataforma GitHub. Este espacio permitirá centralizar el código fuente, los notebooks de análisis, los scripts de procesamiento y los archivos de configuración, garantizando una gestión ordenada del flujo de trabajo.

Capítulo 5: Resultados

Como punto inicial para la exposición de los resultados, se presenta nuevamente el ML Canvas formulado durante la Fase 1: Análisis y diseño del modelo predictivo, el cual

resume los componentes esenciales del proyecto y su correspondencia con los objetivos definidos. Esta herramienta facilitó la organización integral de los requerimientos técnicos, las fuentes de datos, los indicadores de evaluación y la propuesta de valor del sistema desarrollado. A continuación, se muestra la tabla que sintetiza este análisis, donde se establecen las necesidades detectadas, los elementos metodológicos considerados y los beneficios proyectados de la aplicación del modelo predictivo.

Tabla 2 Requerimientos ML Canvas Parte 1







TAREA DE PREDICCIÓN 	DECISIONES 	PROPUESTA DE VALOR 	RECOPIACIÓN DE DATOS 	FUENTES DE DATOS 
<p>Entrada: Variables demográficas, académicas y socioeconómicas.</p> <p>Salida (Target): Performance (Excellent, Good, Average, Poor en Nivel de desempeño esperado).</p> <p>Tipo de tarea: Clasificación supervisada.</p> <p>Algoritmos candidatos:</p> <ul style="list-style-type: none"> • Regresión Logística • SVM • Gradient Boosting (XGBoost, LightGBM) • Regresión Logística Ordinal 	<p>Identificar factores sociales, educativos y personales que afectan en el desempeño de los estudiantes.</p> <p>Diseñar tutorías específicas de acuerdo a las necesidades de cada estudiante.</p> <p>La predicción del rendimiento también pueden ser:</p> <ul style="list-style-type: none"> • Apoyar políticas de becas o refuerzos educativos. • Ayudar a identificar candidatos en riesgo académico. • Guiar estrategias de admisión o intervención temprana. 	<p>La propuesta de valor de este enfoque ofrece la posibilidad de generar información para diferentes actores del sistema educativo. Para las instituciones, el modelo permitirá identificar factores éxito académico y diseñar programas de apoyo. Para los estudiantes y sus familias, ofrecerá una visión más clara sobre el impacto real de invertir en programas de coaching, ayudando a una mejor toma de decisiones educativas. Finalmente, a nivel social, este análisis contribuirá a visibilizar posibles desigualdades en el acceso a recursos de formación, fomentando el diseño de políticas más justas y equitativas.</p>	<p>Los datos recolectados a partir de registros académicos y formularios de inscripción.</p> <p>Los datos incluyen variables demográficas (género, casta), educativas (porcentajes de clase X y XII, tipo de institución), y familiares (ocupación de los padres).</p>	<p>Los datos son obtenidos del dataset público "Student Performance on an Entrance Examination".</p> <p>Dataset recopilado por el Prof. Jiten Hazarika, que contiene información académica y socioeconómica de los aspirantes a universidades médicas de Assam.</p>

Tabla 3 Requerimientos ML Canvas Parte 2

SIMULACIÓN DE IMPACTO 	HACIENDO PREDICCIONES 	CONSTRUYENDO MODELOS 	CARACTERÍSTICAS 
<p>Mayor precisión de predicción y de transparencia</p> <p>Métricas de evaluación: Accuracy, Precision, Recall, F1-Score, Confusion Matrix</p>	<p>Las predicciones se hacen en bloque al finalizar las aplicaciones del examen de admisión.</p> <p>Se pueden hacer predicciones en tiempo real para calcular la probabilidad de aprobación de los estudiantes.</p>	<p>Construir un solo modelo de clasificación supervisada que prediga el nivel de rendimiento (Performance) de los estudiantes en categorías como <i>Excellent, Good, Poor, Average</i>.</p> <ul style="list-style-type: none"> Entrenamiento inicial con el dataset histórico. Actualización anual con nuevos datos de ingreso. Frecuencia: una vez por ciclo académico (anual). Tiempo estimado para el modelado: 1-2 semanas para limpieza, entrenamiento y validación. 	<p>Entradas que alimentan el modelo:</p> <ul style="list-style-type: none"> Gender Caste Coaching (ninguno, dentro o fuera de Assam) Class_ten_education (tipo de institución) Twelve_education (tipo de institución) Medium (idioma de instrucción) Class_X_Percentage Class_XII_Percentage Father_occupation Mother_occupation

Tabla 4 Requerimientos ML Canvas Parte 3

MONITOREO 		
Seguimiento de métricas técnicas (Accuracy, Precision, Recall, F1-Score, Confusion Matrix)	Monitoreo de las diferencias de desempeño entre grupos(género, casta, desempeño previo, etc.)	Detectar si hay cambios en los patrones de los datos.

Durante la etapa de manipulación y preparación de datos, se ejecutó un proceso de limpieza integral orientado a eliminar inconsistencias, estandarizar valores y garantizar la uniformidad semántica del conjunto de datos. El script implementado realizó una serie de transformaciones columna por columna, corrigiendo errores tipográficos, variaciones de formato y discrepancias en el uso de mayúsculas y minúsculas.

En la variable Performance, se unificaron las distintas formas de escritura de las categorías, reemplazando variantes como *gOOD*, *aVERAGE*, *vG* y *eXCELLENT* por las etiquetas correctas Good, Average, Vg y Excellent respectivamente. De manera similar, se normalizó la variable Gender estandarizando *MALE* y *FEMALE* como *male* y *female*,

mientras que en *Caste* se corrigieron las abreviaciones y errores de capitalización para mantener la coherencia entre categorías (ST, OBC, General, SC).

El mismo procedimiento se aplicó a las variables *coaching*, *time*, *Class_ten_education*, *twelve_education*, *medium*, *Father_occupation* y *Mother_occupation*, homogeneizando las entradas para eliminar inconsistencias de escritura y mejorar la legibilidad de los datos. Asimismo, se aplicó un tratamiento de valores nulos mediante la imputación de la moda (valor más frecuente) en cada variable, con el fin de preservar la integridad de las observaciones sin introducir sesgos estadísticos.

Finalmente, se eliminó la columna *mixed_type_col*, identificada como irrelevante para el análisis por contener valores de tipo mixto que afectaban la consistencia del *dataset*. Tras estas transformaciones, el conjunto de datos resultante presentó una estructura más limpia, coherente y preparada para las etapas de modelado.

Todas las operaciones fueron registradas y versionadas a través de W&B, permitiendo conservar un historial completo de los cambios realizados sobre el *dataset* y asegurando su trazabilidad en el flujo MLOps.

Durante la etapa de exploración y preprocesamiento de datos, se llevó a cabo un examen detallado con el propósito de conocer la composición y las propiedades generales del conjunto de información. En primera instancia, se procedió a cargar el archivo de datos modificado, comprobando su correcta lectura y la integridad de los registros mediante la visualización inicial y final de las filas contenidas en el *dataset*. Posteriormente, se identificó la cantidad total de observaciones y variables, además de los tipos de datos correspondientes a cada una, diferenciando entre aquellas de carácter numérico y las de tipo categórico.

De forma complementaria, se efectuó un análisis descriptivo inicial destinado a determinar el número de valores únicos, las categorías más frecuentes y la presencia de posibles inconsistencias en la distribución de los atributos. Para cada variable se calcularon las frecuencias absolutas, lo que permitió identificar la proporción de estudiantes en los distintos niveles de desempeño y grupos sociodemográficos.

Después de aplicar la normalización de etiquetas en la variable *Performance*, los valores fueron unificados para eliminar inconsistencias tipográficas y garantizar uniformidad semántica. Tras el proceso de limpieza, las categorías finales quedaron consolidadas en *Excellent*, *Very Good (Vg)*, *Good* y *Average*, con la siguiente distribución definitiva:

Tabla 5 Frecuencia de la Categoría de *Performance*

Categoría de rendimiento	Frecuencia	Porcentaje
Good	213	31.39 %
Very Good (Vg)	200	29.46 %
Average	158	23.27 %
Excellent	108	15.89 %
Total	679	100 %

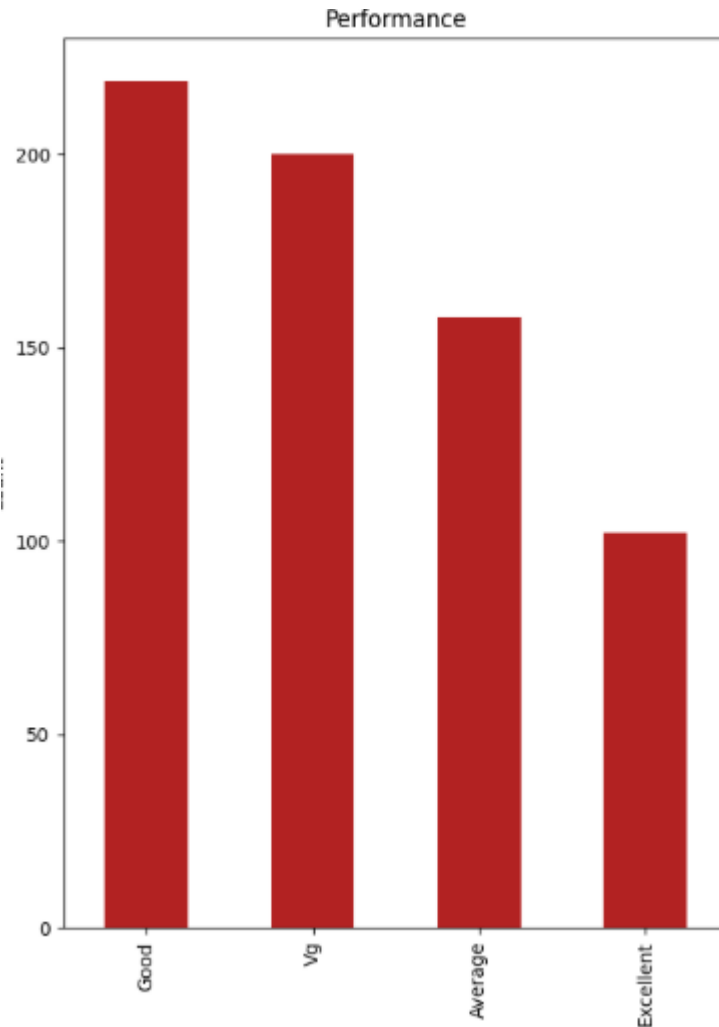


Ilustración 2 Grafico de Frecuencia de la Categoría de *Performance*

Después del proceso de corrección, el conjunto de datos adquirió una estructura uniforme y consistente, libre de discrepancias en la escritura y de valores duplicados que pudieran comprometer el desempeño del modelo predictivo. Se observa que la mayor proporción de estudiantes se agrupa en los niveles *Good* y *Very Good*, los cuales, en

conjunto, superan el 60% de la muestra analizada, reflejando una tendencia general hacia un rendimiento académico medio-alto.

El modelo XGBoost presentó un desempeño general moderado, alcanzando una *Accuracy* de 0.4741. El análisis por clase evidenció un comportamiento desigual entre categorías, donde la clase *Average* obtuvo los mejores resultados con una precisión de 0.86 y un *F1-Score* de 0.80, mientras que las categorías *Good*, *Very Good* y *Excellent* mostraron valores considerablemente menores, con *F1-Scores* de 0.47, 0.33 y 0.29, respectivamente. Esta tendencia sugiere que el modelo logró identificar con mayor efectividad a los estudiantes con desempeño promedio, pero tuvo dificultades para distinguir entre niveles de rendimiento superior, posiblemente debido al desbalance en la distribución de clases.

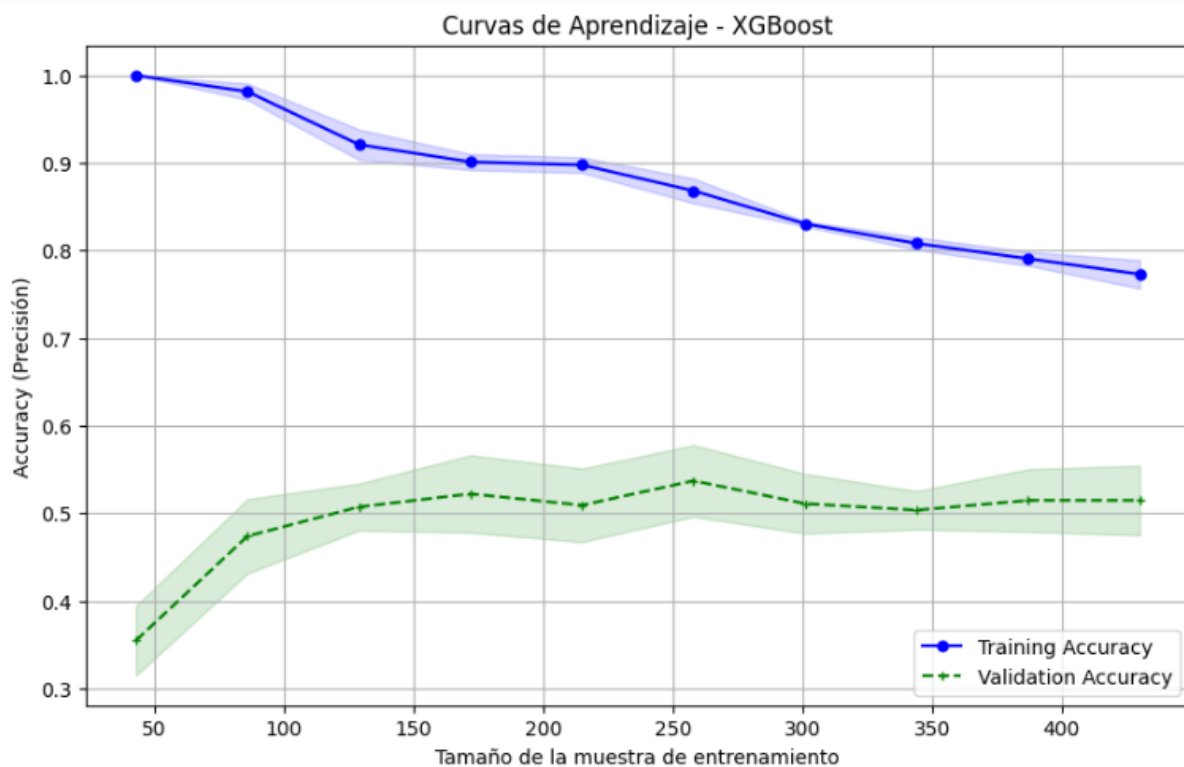


Ilustración 3 Curvas de Aprendizaje - XGBoost

En cuanto a las curvas de aprendizaje, se observa una clara brecha entre el desempeño de entrenamiento y el de validación, lo que indica un grado de sobreajuste (*overfitting*). El modelo alcanzó una alta precisión en el conjunto de entrenamiento, pero la *Accuracy* de validación permaneció cercana al 50%, sin mostrar convergencia significativa. Este comportamiento revela que, aunque XGBoost aprendió patrones específicos del conjunto de entrenamiento, su capacidad de generalización hacia nuevos datos fue limitada.

El modelo de Regresión Logística mostró un rendimiento ligeramente superior al de XGBoost, alcanzando una *Accuracy* de 0.5185. En el análisis por clase, se observa que la categoría *Average* continúa siendo la mejor representada, con una precisión de 0.82 y un *F1-Score* de 0.77, mientras que las categorías *Good* y *Very Good* presentan un comportamiento intermedio, con *F1-Scores* de 0.51 y 0.47, respectivamente. Por su parte, la clase *Excellent* mantiene un bajo desempeño (*F1-Score* = 0.20), lo que refleja la dificultad del modelo para distinguir correctamente los casos de mayor rendimiento académico, probablemente debido al desbalance de la muestra y a la similitud entre los atributos que caracterizan a los niveles altos de desempeño.

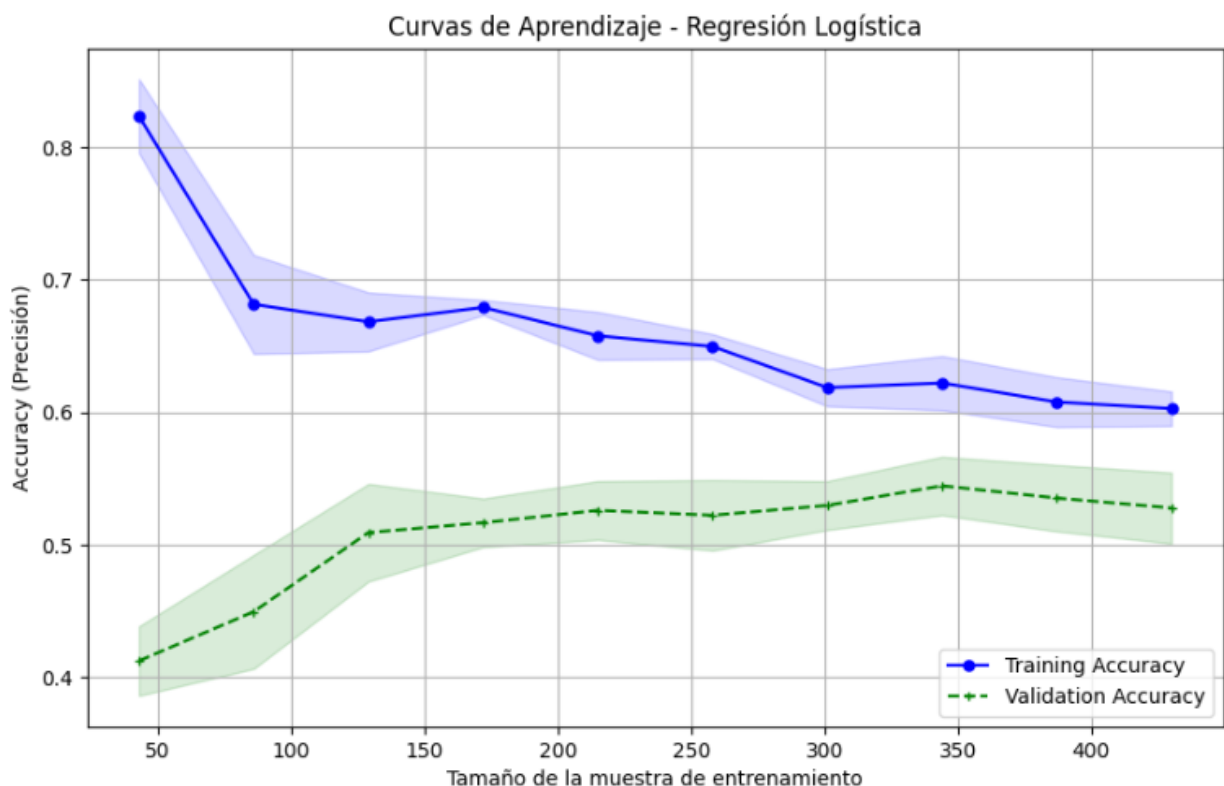


Ilustración 4 Curvas de Aprendizaje - Regresión Logística

Las curvas de aprendizaje evidencian un patrón de comportamiento consistente con un ligero sobreajuste. A medida que el tamaño del conjunto de entrenamiento aumenta, la precisión de entrenamiento desciende progresivamente, estabilizándose cerca de 0.6, mientras que la precisión de validación se mantiene alrededor de 0.5 sin converger plenamente. Esta diferencia indica que el modelo logra aprender patrones generales de los datos, pero aún enfrenta limitaciones para generalizar con mayor precisión.

El modelo SVM obtuvo una *Accuracy* de 0.4963, mostrando un desempeño similar al de los modelos previos, aunque con un comportamiento particular en las métricas por clase. La categoría *Average* nuevamente se posicionó como la mejor representada, con una

precisión de 0.95 y un $F1$ -Score de 0.78, mientras que las clases *Good* y *Very Good* mantuvieron niveles moderados de desempeño, con $F1$ -Scores de 0.48 y 0.45, respectivamente. Por el contrario, la clase *Excellent* mostró un $F1$ -Score de apenas 0.10, lo que evidencia una baja capacidad del modelo para identificar correctamente los casos de mayor rendimiento.

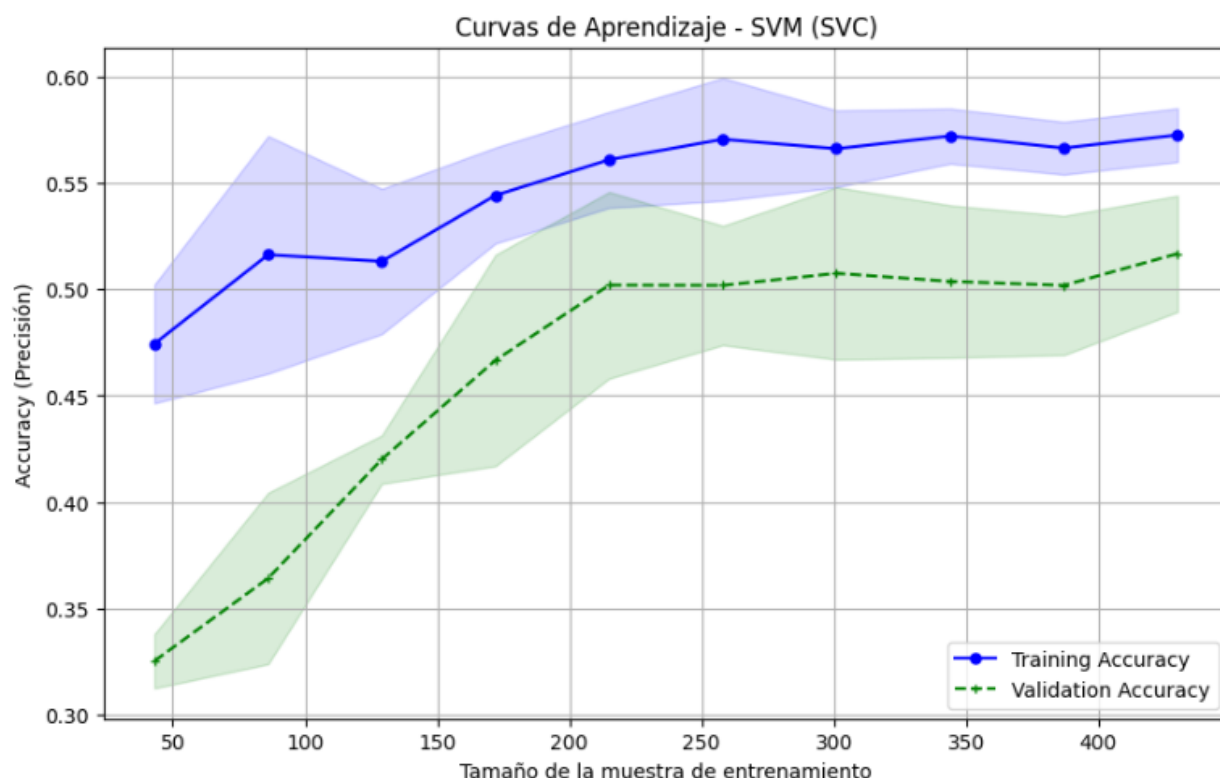


Ilustración 5 Curvas de Aprendizaje - SVM

En relación con las curvas de aprendizaje, se observa una tendencia ascendente tanto en la precisión de entrenamiento como en la de validación, lo que indica que el modelo continúa mejorando su ajuste conforme aumenta el tamaño de la muestra. Sin embargo, la brecha persistente entre ambas curvas refleja la existencia de ligero sobreajuste, aunque con una estabilidad superior respecto a los modelos anteriores. Este comportamiento sugiere que SVM presenta una buena capacidad de aprendizaje progresivo y una tendencia hacia la convergencia, lo que lo convierte en un modelo prometedor para etapas posteriores.

El modelo de Regresión Logística Ordinal obtuvo una *Accuracy* global de 0.5111, posicionándose como uno de los algoritmos con mejor equilibrio general entre las clases. La categoría *Average* alcanzó una precisión perfecta (1.00) y un $F1$ -Score de 0.77, consolidándose como la más estable del conjunto, mientras que *Very Good* presentó un $F1$ -Score de 0.52, evidenciando un avance en la identificación de niveles de desempeño medio-alto. En contraste, las categorías *Good* y *Excellent* registraron desempeños más

modestos (*F1-Scores* de 0.44 y 0.16, respectivamente), lo que indica que, si bien el modelo es capaz de capturar relaciones ordinales entre las clases, aún enfrenta limitaciones para diferenciar los extremos del espectro de rendimiento

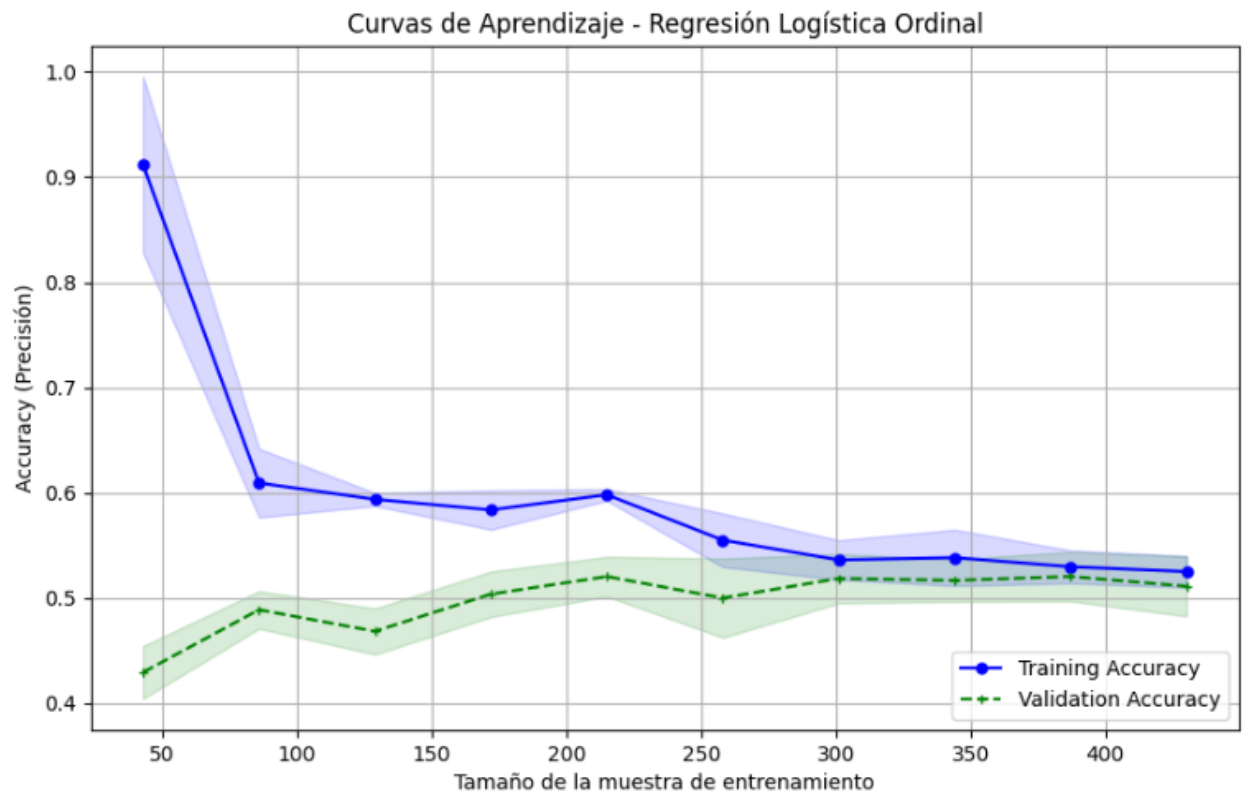
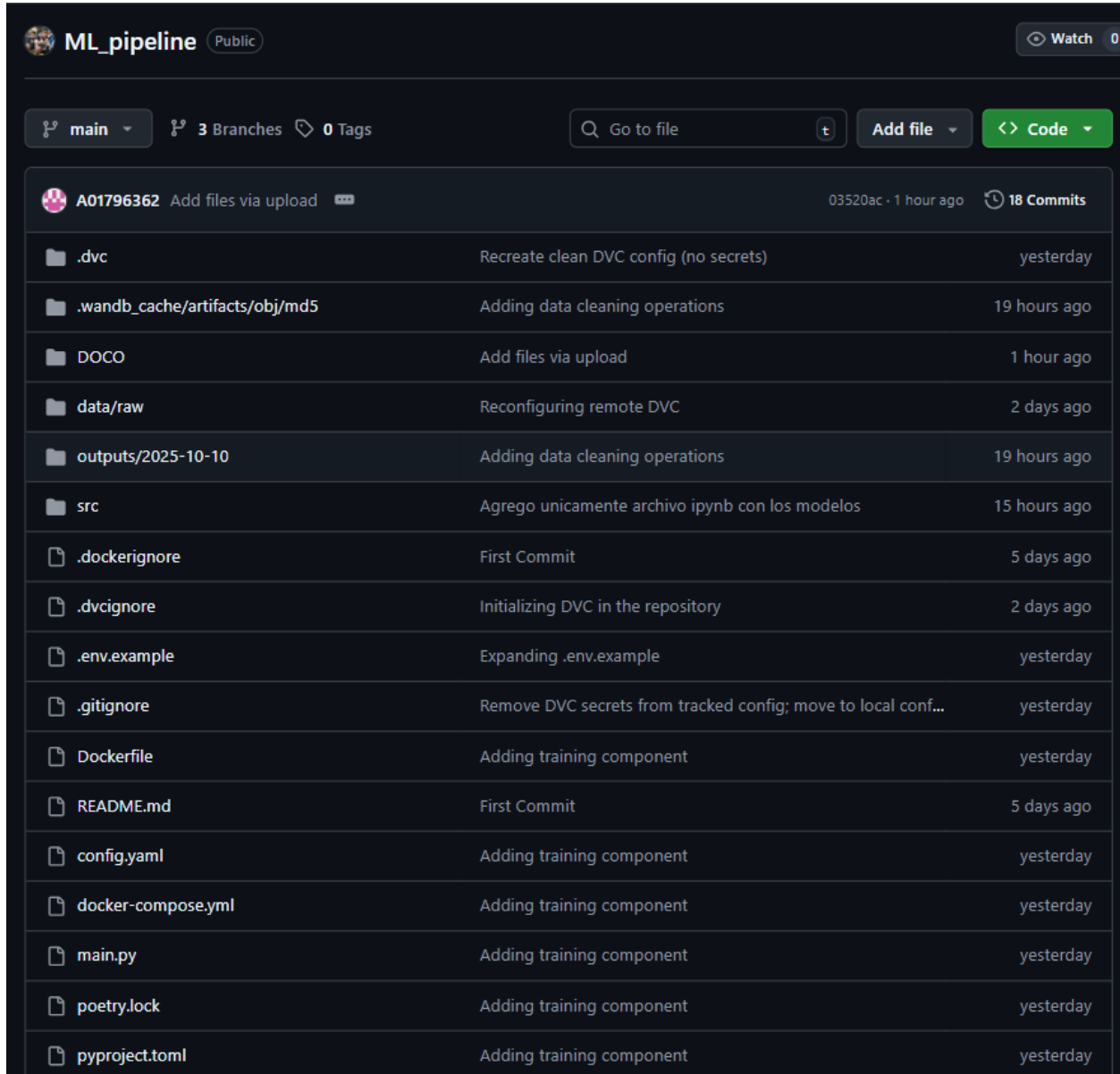


Ilustración 6 Curvas de Aprendizaje - Regresión Logística Ordinal

Las curvas de aprendizaje evidencian una tendencia progresiva hacia la estabilidad y la convergencia entre la precisión del entrenamiento y la de validación, lo que indica un comportamiento más equilibrado y con menor grado de sobreajuste en comparación con los modelos previos. Conforme se incrementa el tamaño de la muestra, la precisión de validación se mantiene constante en torno al 50%, mientras que la de entrenamiento desciende de manera gradual hasta aproximarse a ese mismo nivel, reflejando así una mejor capacidad de generalización y un ajuste más consistente del modelo frente a nuevos datos.

La creación del repositorio en GitHub permitió centralizar todos los elementos del proyecto y mantener un control eficiente sobre las versiones del código, los notebooks y los archivos de configuración. Esta herramienta facilitó la colaboración entre los integrantes del equipo, promoviendo un registro detallado de los cambios y una documentación continua de los avances durante cada etapa de la investigación. Su integración resultó fundamental para asegurar la reproducibilidad y coherencia del

desarrollo, consolidando un flujo de trabajo alineado con las buenas prácticas del ciclo de vida MLOps.



The screenshot shows the GitHub interface for a repository named 'ML_pipeline'. At the top, there's a header with the repository name, a 'Public' badge, and a 'Watch' button. Below this, a navigation bar shows 'main' as the selected branch, '3 Branches', and '0 Tags'. A search bar 'Go to file' and buttons for 'Add file' and 'Code' are also present. The main content area displays a list of files and folders, each with a commit message and a timestamp. The files include configuration files like '.dvc', '.wandb_cache/artifacts/obj/md5', '.DOCO', 'data/raw', and 'outputs/2025-10-10', as well as source code files like '.dockerignore', '.dvcignore', '.env.example', '.gitignore', 'Dockerfile', 'README.md', 'config.yaml', 'docker-compose.yml', 'main.py', 'poetry.lock', and 'pyproject.toml'. The commit messages range from 'Recreate clean DVC config (no secrets)' to 'Adding training component'.

File/Folder	Commit Message	Time
.dvc	Recreate clean DVC config (no secrets)	yesterday
.wandb_cache/artifacts/obj/md5	Adding data cleaning operations	19 hours ago
DOCO	Add files via upload	1 hour ago
data/raw	Reconfiguring remote DVC	2 days ago
outputs/2025-10-10	Adding data cleaning operations	19 hours ago
src	Agrego unicamente archivo ipynb con los modelos	15 hours ago
.dockerignore	First Commit	5 days ago
.dvcignore	Initializing DVC in the repository	2 days ago
.env.example	Expanding .env.example	yesterday
.gitignore	Remove DVC secrets from tracked config; move to local conf...	yesterday
Dockerfile	Adding training component	yesterday
README.md	First Commit	5 days ago
config.yaml	Adding training component	yesterday
docker-compose.yml	Adding training component	yesterday
main.py	Adding training component	yesterday
poetry.lock	Adding training component	yesterday
pyproject.toml	Adding training component	yesterday

Ilustración 7 Repositorio de GitHub

Capítulo 6. Conclusiones y Recomendaciones (parciales)

El desarrollo de la Fase 1: Análisis y diseño del modelo predictivo permitió establecer una base sólida para la construcción del sistema de clasificación supervisada orientado a predecir el rendimiento académico de los estudiantes. A través del proceso de limpieza y estandarización del conjunto de datos, se logró garantizar la coherencia interna del mismo, eliminando inconsistencias, errores tipográficos y duplicidades semánticas que afectaban la integridad de la información original.

El análisis exploratorio evidenció que la muestra cuenta con una distribución equilibrada entre las categorías de rendimiento, predominando los niveles *Good* y *Very Good*, que en conjunto representan más del 60% de los casos. Este hallazgo sugiere que el grupo estudiado presenta un desempeño académico medio-alto, con menor representación de los niveles *Average* y *Excellent*. Asimismo, se identificaron variables relevantes de tipo demográfico, educativo y familiar que servirán como insumo para el entrenamiento del modelo, permitiendo analizar de manera más precisa los factores asociados al éxito académico.

La incorporación del enfoque MLOps y el uso de herramientas como W&B y Python proporcionaron un marco metodológico reproducible, escalable y trazable, asegurando la transparencia de cada transformación aplicada. Esta integración facilitará el avance hacia las siguientes fases, donde se llevará a cabo el entrenamiento, ajuste y despliegue del modelo predictivo dentro de un entorno controlado y automatizado.

Entre los modelos evaluados, la Regresión Logística Ordinal se perfila como la opción más adecuada para integrarse en fases posteriores del proyecto. Aunque su *Accuracy* (0.5111) se mantiene en un rango similar al de los demás algoritmos, presenta un comportamiento más estable y equilibrado entre las métricas de entrenamiento y validación, además de una mejor capacidad de generalización. Su naturaleza ordinal le permite representar de manera coherente las relaciones jerárquicas entre las categorías de rendimiento académico (*Excellent*, *Very Good*, *Good*, *Average*), lo que otorga una ventaja conceptual frente a modelos puramente categóricos.

A diferencia de los demás enfoques, la Regresión Logística Ordinal mostró curvas de aprendizaje convergentes y una reducción visible del sobreajuste, lo que indica un modelo más consistente y menos dependiente del conjunto de entrenamiento. Por ello, se considera el punto de partida más apropiado para la siguiente fase del proyecto, donde se buscará optimizar su desempeño mediante técnicas de regularización, ajuste de hiperparámetros y potencial integración dentro del entorno automatizado MLOps.

Aunque el proceso de limpieza y estandarización del conjunto de datos permitió alcanzar una estructura coherente y adecuada para el modelado, resulta conveniente seguir perfeccionando esta fase en etapas futuras del proyecto. En especial, se sugiere profundizar en la detección de valores atípicos, incorporando métodos estadísticos y representaciones visuales que faciliten la identificación de registros anómalos o inconsistentes dentro de las variables numéricas. Esta mejora contribuiría a fortalecer la calidad del conjunto de datos y asegurar una representación más precisa y confiable de la población estudiada.

Se considera necesario llevar a cabo un ajuste más profundo de los hiperparámetros del modelo, con el fin de potenciar su capacidad predictiva y elevar su desempeño a un nivel superior. Este proceso implicará la exploración sistemática de configuraciones mediante técnicas como *Grid Search* o *Random Search*, analizando diferentes valores de regularización, tipos de penalización y parámetros de optimización. Del mismo modo, será pertinente ampliar el alcance de la validación cruzada e implementar mecanismos de selección automática de variables, orientados a mejorar la representación de las características más influyentes. Con estas acciones, se espera incrementar la precisión y robustez del modelo.

Referencias

- [1] B. C. Fulano-Vargas and I. N. Meneses-Runza, "La evaluación escolar desde la perspectiva de los estudiantes," *RECIE. Revista Caribeña de Investigación Educativa*, vol. 7, no. 1, pp. 163–182, 2023.
- [2] M. A. L. O. Guizado, "Influencia de la evaluación del desempeño directivo y su efectividad en la gestión pública de la educación escolar," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 6, no. 5, pp. 874–892, 2022.
- [3] G. Oyarzún Vargas and A. Falabella, "Indicadores de Desarrollo Personal y Social: La ilusión de la evaluación integral de la calidad," *Psicoperspectivas*, vol. 21, no. 1, pp. 149–162, 2022.
- [4] L. L. Blanco Figueredo and K. E. Arias Ortega, "Perspectiva decolonial de la evaluación escolar en contexto de diversidad social y cultural," *Perfiles Educativos*, vol. 44, no. 177, pp. 168–182, 2022.
- [5] A. Clément, É. Robinot, and L. Trespeuch, "The use of ESG scores in academic literature: A systematic literature review," *Journal of Enterprising Communities: People and Places in the Global Economy*, vol. 19, no. 1, pp. 92–110, 2025.
- [6] X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, and T. Y. J. Han, "Explainable machine learning in materials science," *npj Computational Materials*, vol. 8, no. 1, p. 204, 2022.
- [7] G. Nicora, M. Rios, A. Abu-Hanna, and R. Bellazzi, "Evaluating pointwise reliability of machine learning prediction," *Journal of Biomedical Informatics*, vol. 127, p. 103996, 2022.
- [8] M. M. John, H. H. Olsson, and J. Bosch, "Towards MLOps: A framework and maturity model," in *Proc. 2021 47th Euromicro Conf. on Software Engineering and Advanced Applications (SEAA)*, Palermo, Italy, Sept. 2021, pp. 1–8. IEEE.
- [9] B. P. R. Rella, "MLOps and DataOps integration for scalable machine learning deployment," *International Journal for Multidisciplinary Research*, vols. 1–3, 2022. [Online]. Available: <https://www.researchgate.net/publication/390554912>
- [10] F. Lanubile, S. Martínez-Fernández, and L. Quaranta, "Teaching MLOps in higher education through project-based learning," in *Proc. 2023 IEEE/ACM 45th Int. Conf. on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, Melbourne, Australia, May 2023, pp. 95–100. IEEE.
- [11] S. Mohan, G. Gujrati, and N. Senthamarai, "Measuring and predicting student's academic performance using ML," in *Proc. Congress on Smart Computing Technologies*, Singapore: Springer Nature Singapore, Dec. 2022, pp. 387–403.

- [12] C. F. E. Hoyos and J. C. A. Barilla, "El sistema educativo como sistema esencial para el desarrollo y la transformación social," *Revista Oratores*, no. 14, pp. 144–156, 2021.
- [13] L. M. Castro-Benavides, J. A. Tamayo-Arias, and D. Burgos, "Escenarios de la docencia frente a la transformación digital de las Instituciones de Educación Superior," *Education in the Knowledge Society (EKS)*, vol. 23, pp. e27866–e27866, 2022.
- [14] W. M. Romero Casalliglla, J. O. Chulca Abalco, G. G. Imbaquingo Guzmán, S. E. Pineda Anchaguano, E. C. Aules Aules, G. O. Tipán Sánchez, *et al.*, "Evaluación para el aprendizaje: más allá de las calificaciones," *Revista InveCom*, vol. 5, no. 1, 2025.
- [15] C. E. Mendoza Suárez and C. A. Bullón Romero, "Gestión del conocimiento en instituciones de educación superior: una revisión sistemática," *Horizontes. Revista de Investigación en Ciencias de la Educación*, vol. 6, no. 26, pp. 1992–2003, 2022.
- [16] H. Saravia Domínguez, P. Saavedra Villar, L. M. Felices Vizarreta, M. M. Campos Espinoza, and J. R. Janampa Urbano, "La aplicación del diseño curricular por competencias en la Educación Superior: Una revisión sistemática 2019–2023," *Comuni@cción*, vol. 15, no. 1, pp. 92–104, 2024.
- [17] Y. P. C. González, S. Z. J. Mora, and R. G. M. Morillo, "Tendencias y desafíos políticos y socioculturales de la educación superior contemporánea en Latinoamérica," *Revista Boletín Redipe*, vol. 11, no. 1, pp. 71–91, 2022.
- [18] E. J. C. Ortegón and R. R. Rey-González, "Una revisión crítica del modelo de admisión por estratos académicos en la Universidad Nacional de Colombia," *Revista Educación Superior y Sociedad (ESS)*, vol. 34, no. 2, pp. 52–75, 2022.
- [19] M. A. Rodríguez Medina, E. R. Poblano-Ojinaga, L. Alvarado Tarango, A. González Torres, and M. I. Rodríguez Borbón, "Validación por juicio de expertos de un instrumento de evaluación para evidencias de aprendizaje conceptual," *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, vol. 11, no. 22, 2021.
- [20] J. I. Pincay-Ponce, A. De Giusti, J. J. Reyes-Cárdenas, A. G. Franco-Pico, A. V. Macías-Espinales, and P. A. Quiroz-Palma, "Analítica de datos de factores socioeconómicos que inciden en el rendimiento escolar: Revisión sistemática," *Revista Ibérica de Sistemas e Tecnologías de Informação*, no. E54, pp. 531–545, 2022.
- [21] J. N. M. Álava, L. A. M. Bermeo, J. Morales-Carrillo, and L. Cedeño-Valarezo, "Modelos de aprendizaje automático: Aplicación y eficiencia," *Revista Científica de Informática ENCRYPTAR*, vol. 7, no. 14, pp. 87–114, 2024.
- [22] O. E. D. Rodríguez and M. G. P. Hernández, "Minería de intenciones a partir de una base del conocimiento y aprendizaje automático supervisado," *3C TIC: Cuadernos de Desarrollo Aplicados a las TIC*, vol. 10, no. 3, pp. 65–101, 2021.

- [23] A. Almufarreh, K. M. Noaman, and M. N. Saeed, "Academic teaching quality framework and performance evaluation using machine learning," *Applied Sciences*, vol. 13, no. 5, p. 3121, 2023.
- [24] K. Fahd, S. Venkatraman, S. J. Miah, and K. Ahmed, "Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature," *Education and Information Technologies*, vol. 27, no. 3, pp. 3743–3775, 2022.
- [25] A. Guillén Perales, F. Liébana-Cabanillas, J. Sánchez-Fernández, and L. J. Herrera, "Assessing university students' perception of academic quality using machine learning," *Applied Computing and Informatics*, vol. 20, no. 1/2, pp. 20–34, 2024.
- [26] J. M. A. Acosta, M. L. Ramirez, and R. G. Cabrera, "Impacto del preprocesamiento en la clasificación automática de textos usando aprendizaje supervisado y Reuters 21578," *Revista Colombiana de Tecnologías de Avanzada*, vol. 1, no. 43, pp. 110–118, 2024.
- [27] B. P. Cordero-Torres, "Algoritmos de aprendizaje supervisado para proyección de ventas de camarón ecuatoriano con lenguaje de programación Python," *Economía y Negocios*, vol. 13, no. 2, pp. 30–51, 2022.
- [28] E. M. García, C. C. López, J. A. M. Rivas, and D. L. A. Capistran, "Evaluación de algoritmos de aprendizaje supervisado usando modelos binarios para clasificación de análisis de sentimiento," *Tecnología Educativa Revista CONAIC*, vol. 11, no. 1, pp. 92–97, 2024.
- [29] M. C. Mihaescu and P. S. Popescu, "Review on publicly available datasets for educational data mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 3, p. e1403, 2021.
- [30] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*. Cambridge, MA: Morgan Kaufmann, 2022.
- [31] S. K. Dubey, S. Gupta, S. K. Pandey, and S. Mittal, "Ethical considerations in data mining and database research," in *Proc. 2024 11th Int. Conf. on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, Mar. 2024, pp. 1–4. IEEE.
- [32] C. I. Loeza-Mejía, E. Sánchez-DelaCruz, and I. O. José-Guzmán, "Machine learning methodologies: history and challenges," *Evolutionary Intelligence*, vol. 18, no. 3, p. 58, 2025.

- [33] B. Eken, S. Pallewatta, N. Tran, A. Tosun, and M. A. Babar, "A multivocal review of MLOps practices, challenges and open issues," *ACM Computing Surveys*, vol. 58, no. 2, pp. 1–35, 2025.
- [34] S. Mei, C. Liu, Q. Wang, and H. Su, "Model provenance management in MLOps pipeline," in *Proc. 2022 8th Int. Conf. on Computing and Data Engineering*, Tokyo, Japan, Jan. 2022, pp. 45–50.
- [35] S. Moreschini, D. Hästbacka, and D. Taibi, "MLOps pipeline development: The OSSARA use case," in *Proc. 2023 Int. Conf. on Research in Adaptive and Convergent Systems*, Barcelona, Spain, Aug. 2023, pp. 1–8.
- [36] F. J. Iriarte, M. E. Ortiz, L. Unzueta, J. Martínez, J. Zaldivar, and I. Arganda-Carreras, "Multi-level XAI-driven MLOps pipeline for the adjustment of fruit and vegetable classifiers," in *Proc. 2024 IEEE 12th Int. Conf. on Intelligent Systems (IS)*, Warsaw, Poland, Aug. 2024, pp. 1–6. IEEE.
- [37] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," *arXiv preprint arXiv:2008.05756*, 2020.
- [38] A. Tharwat, "Classification Assessment Methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2021, doi: 10.1016/j.aci.2018.08.003.
- [39] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 820–836, 2024, doi: 10.1162/tacl_a_00675.
- [40] O. Rainio, P. Kumpulainen, and T. Hämmäläinen, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, 2024, doi: 10.1038/s41598-024-56706-x.
- [41] M. Ghanem *et al.*, "Limitations in Evaluating Machine Learning Models for Imbalanced Binary Outcome Classification in Spine Surgery: A Systematic Review," *Brain Sciences*, vol. 13, no. 12, p. 1723, 2023, doi: 10.3390/brainsci13121723.
- [42] Student Performance on an Entrance Examination [Dataset]. (2018). UCI Machine Learning Repository. <https://doi.org/10.24432/C58D0H>.