



Ciencia Latina Revista Científica Multidisciplinaria, Ciudad de México, México.

ISSN 2707-2207 / ISSN 2707-2215 (en línea), enero-febrero 2026,

Volumen 10, Número 1.

https://doi.org/10.37811/cl_rcm.v10i1

MODELO PARA PREDECIR EL ÉXITO PROFESIONAL EN EGRESADOS DE LA UNS MEDIANTE APRENDIZAJE AUTOMÁTICO BASADO EN DATOS ACADÉMICOS Y SOCIOECONÓMICOS

MODEL FOR PREDICTING PROFESSIONAL SUCCESS IN UNS
GRADUATES THROUGH MACHINE LEARNING BASED ON
ACADEMIC AND SOCIOECONOMIC DATA

Johan Max Alexander López Heredia

Universidad Nacional del Santa, Perú



DOI: https://doi.org/10.37811/cl_rcm.v10i1.22531

Modelo para Predecir el Éxito Profesional en Egresados de la UNS mediante Aprendizaje Automático basado en Datos Académicos y Socioeconómicos

Johan Max Alexander López Heredia¹

jlopez@uns.edu.pe

<https://orcid.org/0009-0003-1653-5835>

Universidad Nacional del Santa

Perú

RESUMEN

El estudio que precede a este artículo tuvo como objetivo desarrollar un modelo predictivo basado en el aprendizaje automático para estimar el grado de correlación entre el éxito profesional de egresados de la Escuela Profesional de Ingeniería de Sistemas e Informática de la UNS y sus características académicas y socioeconómicas. Se recopilaron datos académicos, socioeconómicos y laborales relacionados a las variables de estudio. Tras el pre procesamiento de variables, se aplicaron técnicas de oversampling y se entrenaron algoritmos de aprendizaje automático (Random Forest, XGBoost y MLPClassifier). La población se conformó por 825 egresados de la carrera profesional y la muestra intencional, por 96 de ellos. Los resultados evidenciaron que Random Forest y XGBoost alcanzaron precisión y macro F1-score del 100% en validación y prueba, superando el umbral del 80% inicialmente propuesto. Además, la importancia de variables subraya que factores como estudios de posgrado y certificaciones, pesan más que la nota promedio en la predicción del éxito. Con ello se confirmó la hipótesis de que la conjunción de datos académicos y socioeconómicos permite anticipar el nivel de éxito, ofreciendo a la UNS una herramienta de diagnóstico y acción para fortalecer la inserción laboral y la formación continua de sus estudiantes.

Palabras clave: aprendizaje automático, éxito profesional, modelo predictivo, egresados universitarios, empleabilidad

¹ Autor principal

Correspondencia: jlopez@uns.edu.pe

Model for Predicting Professional Success in UNS Graduates through Machine Learning based on Academic and Socioeconomic Data

ABSTRACT

The study preceding this article aimed to develop a machine learning-based predictive model to estimate the correlation between professional success of graduates from the Professional School of Systems and Informatics Engineering at UNS and their academic and socioeconomic characteristics. Academic, socioeconomic, and employment data related to study variables were collected. After variable preprocessing, oversampling techniques were applied and machine learning algorithms (Random Forest, XGBoost, and MLPClassifier) were trained. The population comprised 825 graduates, with an intentional sample of 96. Results showed that Random Forest and XGBoost achieved 100% precision and macro F1-score in validation and testing, exceeding the initially proposed 80% threshold. Furthermore, feature importance analysis revealed that factors such as postgraduate studies and certifications weigh more than grade point average in predicting success. This confirmed the hypothesis that combining academic and socioeconomic data enables anticipating success levels, providing UNS with a diagnostic and action tool to strengthen labor market insertion and continuous education for its students.

Keywords: machine learning, professional success, predictive model, university graduates, employability

*Artículo recibido 15 diciembre 2025
Aceptado para publicación: 19 enero 2026*



INTRODUCCIÓN

El éxito profesional de egresados universitarios se entiende de acuerdo a lo versado por diversos autores como la confluencia de distintos factores, para Baquero Pérez y Ruesga (2019), en primer lugar, el éxito profesional se manifiesta a través de la obtención de un empleo estable, medido a través del tipo de contrato, siendo los indefinidos un indicador de mayor éxito. Asimismo, considera el nivel salarial, donde ubicarse en rangos superiores indica un mayor logro en el ámbito laboral. Otro factor relevante es la adecuación del empleo a la formación recibida, es decir, desempeñar un trabajo acorde al nivel educativo completado. Del mismo modo, el tipo de jornada laboral resulta significativo, siendo más exitoso obtener un trabajo a tiempo completo. Finalmente, la satisfacción con el trabajo realizado también constituye un determinante del éxito profesional. Por su parte, Lavado et al. (2014) enfatiza en aspectos menos tangibles del éxito profesional, como la capacidad de innovar, adaptarse a las tecnologías emergentes y contribuir al desarrollo tecnológico del país. Asimismo, el reconocimiento por parte de la comunidad profesional, la participación en proyectos significativos con impacto social positivo y la formación continua constituyen elementos que enriquecen la concepción del éxito en la vida profesional.

Los egresados universitarios que aspiran al éxito profesional enfrentan múltiples dificultades para insertarse al mundo laboral, tales como la falta de experiencia, escasa empleabilidad, inadecuación de competencias, bajos salarios iniciales, entre otras. Asimismo, existen persistentes brechas significativas entre egresados de universidades públicas y privadas en términos de oportunidades laborales y remuneración. Para mejorar la transición al mundo laboral se requieren reformas en el sistema educativo, mayor conexión entre universidades y centros laborales, y programas de inserción y capacitación laboral para egresados (Yamada et. al, 2013)

En la Universidad Nacional del Santa el seguimiento del desempeño profesional de los egresados se realiza de manera manual con procesos lentos y datos dispersos que no permiten identificar oportunamente los factores que influyen en éxito laboral. Esta situación limita la capacidad institucional de la UNS para implementar estrategias de apoyo temprano y programas personalizados. Si bien se han desarrollado modelos predictivos para estimar el rendimiento académico, son escasos los enfocados en el éxito profesional posterior al egreso (Hung, et. al., 2012), y no se disponía de un modelo predictivo



específico para los egresados de la UNS que integre variables académicas y socio económicas (López et. al., 2019).

Ante esta problemática, la presente investigación buscó desarrollar un modelo predictivo basado en aprendizaje automático que permita estimar el grado de correlación entre el éxito profesional de los egresados de la Escuela Profesional de Ingeniería de Sistemas e Informática de la UNS y sus características académicas y socioeconómicas. El propósito fue construir una herramienta que alcance alta precisión y que además permita identificar claramente qué factores académicos y socio económicos son determinantes en el éxito profesional superando las limitaciones de los procesos manuales y datos dispersos que caracterizaban el seguimiento institucional tradicional.

Asimismo, se pretendió que los hallazgos obtenidos contribuyan a la optimización de las estrategias educativas y programas de apoyo estudiantil de la institución. El modelo desarrollado aspira a servir como un sistema de diagnóstico y acción que permita a la UNS identificar tempranamente a estudiantes con alto potencial de éxito profesional brindándoles mentorías y oportunidades especializadas así como detectar aquellos que podrían enfrentar dificultades en su inserción laboral para diseñar intervenciones oportunas que fortalezcan su formación continua y vinculación en el mercado laboral.

La aplicación de técnicas de aprendizaje automático en el ámbito educativo representa una oportunidad para trascender el análisis descriptivo tradicional (fichas socioeconómicas, encuestas a egresados) hacia un enfoque predictivo que permita anticipar resultados y tomar decisiones informadas (Peña-Ayala, 2014). Este enfoque se alinea con Ley N° 30220, Ley Universitaria en el Perú, que insta a las universidades a evaluar el desempeño de sus egresados para elevar la calidad educativa, así también lo dice el Plan Nacional de Educación Superior y Técnico-Productiva 2021-2025 que promueve una educación superior inclusiva y equitativa (MINEDU, 2021).

Diversos estudios han aplicado técnicas de aprendizaje automático para predecir resultados académicos y laborales de estudiante suniversitarios. Baffa et. al. (2023) desarrollaron modelos predictivos basados en regresión logística, árboles de decisión y bosques aleatorios para estimar la empleabilidad de estudiantes en Nigeria, alcanzando una presición de 98% con Ramdom Forest. En India, Bhagavan et. al. (2020) desarrollaron un algoritmo híbrido HLVQ que superó en precisión a métodos tradicionales para predecir la probabilidad de graduación y empleo. En Latinoamérica, Bedoya et. al (2019)



identificaron que factores socioculturales como el nivel educativo de la madre y el rendimiento académico influyen significativamente en el tiempo para conseguir el primer empleo. En el contexto peruano, Caselli (2021) desarrolló un modelo predictivo aplicando Machine Learning y Deep Learning para el seguimiento a estudiantes universitarios en la UNS obteniendo un 98.97% de precisión en la predicción de deserción. Estos antecedentes fundamentan la viabilidad metodológica del presente estudio y evidencian la efectividad del aprendizaje automático en contextos educativos.

El principal aporte de esta investigación radica en el desarrollo de un modelo predictivo específico para estimar el éxito profesional de egresados universitarios en el contexto peruano, un enfoque escasamente abordado en la literatura mientras que Baffa et. al (2023) se enfocaron en predecir empleabilidad pre-graduación y Caselli (2021) en deserción estudiantil, el presente estudio aborda la predicción del éxito profesional posterior al egreso. Además a diferencia de Bedoya et. al. (2019), quienes identificaron el nivel educativo de la madre como factor determinante, los hallazgos de esta investigación revelan que factores como estudios de posgrado y las certificaciones profesionales tiene mayor peso predictivo que las variables familiares o el promedio de notas, lo que sugiere una evolución en el mercado laboral hacia la valoración de la formación continua. Asimismo, los resultados obtenidos (100% de precisión y F1-Score) superan las métricas reportadas en estudios previos, ofreciendo a la UNS una herramienta de diagnóstico inexistente hasta el momento para fortalecer la inserción laboral de sus egresados.

El estudio contempló como variables independientes dos categorías de características predictoras: las académicas (promedio de notas, ciclos cursados, obtención del grado de bachiller y título profesional) y las socioeconómicas (tipo de convivencia, dependencia económica, condición laboral del estudiante y del responsable económico). La variable dependiente fue el éxito profesional categorizado en niveles de 0 al 5 basado en la autoevaluación del egresado respecto a su inserción laboral, estabilidad, satisfacción, logros y reconocimientos profesionales. Las variables categóricas fueron codificadas mediante técnicas de One-Hot Encoding y Label Encoding, mientras que las numéricas fueron normalizadas para su procesamiento mediante los algoritmos de aprendizaje automático seleccionados.

Los procesos de licenciamiento de la Superintendencia Nacional de Educación Universitaria (SUNEDU) y acreditación del Sistema Nacional de Evaluación, Acreditación y Certificación de la Calidad Educativa (SINEACE) han posicionado el seguimiento a egresados como un indicador



fundamental de calidad universitaria generando la necesidad de que las instituciones demuestren la pertinencia de su formación mediante el desempeño profesional de sus graduados. No obstante, este seguimiento se realiza mayoritariamente de forma manual y descriptiva, sin herramientas que permitan anticipar resultados. Ante ello, esta investigación planteó como hipótesis que un modelo de aprendizaje automático predice el éxito profesional con un F1-Score superior a 0.80 para comprobarlo se establecieron como objetivos: recolectar y analizar datos identificando variables determinantes, diseñar e implementar un modelo predictivo, validar su eficacia mediante técnicas estadísticas, y derivar recomendaciones para la mejora educativa institucional

METODOLOGÍA

El presente estudio se desarrolló bajo un enfoque cuantitativo, ya que se realizó la recolección, procesamiento y análisis de datos numéricos y categóricos mediante técnicas estadísticas y algoritmos de aprendizaje automático. Este enfoque permitió establecer relaciones entre variables académicas, socioeconómicas y el éxito profesional de los egresados, con el objetivo de desarrollar un modelo predictivo basado en evidencia empírica.

Por otro lado, se trató de una investigación aplicada en tanto tiene como finalidad no solo describir teóricamente o establecer relaciones entre el éxito profesional de los egresados y otras variables, sino que se busca aportar con un modelo predictivo, una herramienta funcional orientada a “comprender y resolver problemas prácticos”, lo cual según Bunge (2014) es la esencia de la investigación aplicada. Asimismo, Hernández-Sampieri et al. (2014), quienes afirman que una investigación aplicada tiene como objeto un problema que se perfila a la acción, en el caso específico de la investigación realizada se trata de la optimización del seguimiento y predicción del éxito de los egresados.

El alcance del estudio realizado fue correlacional predictivo, de acuerdo con Según Hernández-Sampieri y Mendoza-Torres (2018) las investigaciones de alcance correlacional “tienen como propósito conocer la relación o grado de asociación que existe entre dos o más conceptos, variables, categorías o fenómenos en un contexto en particular. Permiten cierto grado de predicción” (p. 105). Asimismo, Hernández-Sampieri et al. (2014), afirma que el conocimiento de la relación entre variables permite predecir el comportamiento futuro de los sujetos. Al emplear técnicas de Machine Learning, se busca no solo identificar asociaciones socioculturales, sino alcanzar el nivel de pronóstico descrito



por Kerlinger y Lee (2002), anticipando el éxito profesional de los egresados a partir de su perfil académico previo.

La población del estudio estuvo conformada por los 825 egresados de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional del Santa (UNS), registrados en el sistema de gestión académica institucional hasta el año 2024. Esta población abarcó egresados de diferentes cohortes y representa la totalidad de profesionales formados en esta especialidad que han completado sus estudios universitarios en la UNS. Por otro lado, la muestra estuvo conformada por 98 egresados de la Escuela Profesional de Ingeniería de Sistemas e Informática de la UNS, quienes fueron seleccionados de acuerdo a la disponibilidad y completitud de sus datos académicos, socioeconómicos y laborales, en el registro del sistema de gestión académica institucional, asimismo se tuvo en cuenta la representatividad de diferentes años de egreso, lo cual garantizó que el modelo predictivo capture la variabilidad temporal en las trayectorias profesionales de los egresados. Se tuvo como criterios de exclusión las inconsistencias o falta de datos de los egresados en más del 90% de valores faltantes en las variables de interés para la investigación.

Las técnicas de recolección fueron el análisis documental, la encuesta y la entrevista, los respectivos instrumentos fueron los registros (tras la revisión de los mismos), los cuestionarios y las guías de entrevista. En una primera etapa la recolección de datos se realizó a través del análisis documental de los registros sistemas de gestión académica y de bienestar de la UNS para extraer registros históricos y actualizados relacionados con el desempeño académico, las condiciones socioeconómicas y la información de egreso. De este análisis documental se obtuvo un dataset final obtenido en formato CSV, que integra información proveniente de diversas fuentes internas de la institución y de instrumentos de recolección primaria (cuestionarios y guías de entrevista). Por razones de ética y confidencialidad, todos los datos fueron anonimizados para proteger la identidad de los participantes.

El preprocesamiento de la información se inició con la limpieza de datos, luego se realizó la transformación y codificación, seguida de la reducción de la dimensionalidad y la división del dataset donde el conjunto de datos fue dividido en subconjuntos de entrenamiento (70%), validación (15%) y prueba (15%), lo que permitió ajustar y evaluar la capacidad de generalización del modelo.



En cuanto al software y plataformas, para el procesamiento de datos en general, se utilizó Phyton y R, haciendo uso de librerías como pandas, NumPy, scikit-learn y matplotlib (Müller & Guido, 2016; Murphy, 2022); también se utilizó Google Colab y Kaggle; OpenML y Visual Studio Code y PyCharm versión Comunitaria.

Finalmente, para el desarrollo del modelo predictivo se siguió una secuencia lógica de cuatro fases, la cual inició con la fase de selección de algoritmos, lo cual permitió la evaluación de diversas técnicas de aprendizaje automático; posteriormente se realizó la fase del entrenamiento del modelo, a través del ajuste de hiperparámetros mediante el uso del subconjunto de validación, aplicando técnicas de validación cruzada (k-fold cross-validation) para evitar el sobreajuste; la tercera fase consistió en evaluar el desempeño de cada modelo, utilizando, en el conjunto de prueba, métricas apropiadas para clasificación multiclas: precisión (accuracy), recall, F1-score macro y matriz de confusión. En la cuarta fase del desarrollo del modelo predictivo se analizó las variables predictoras para identificar cuáles tenían mayor incidencia en el éxito profesional, proporcionando información clave para la optimización de las estrategias educativas y de apoyo institucional.

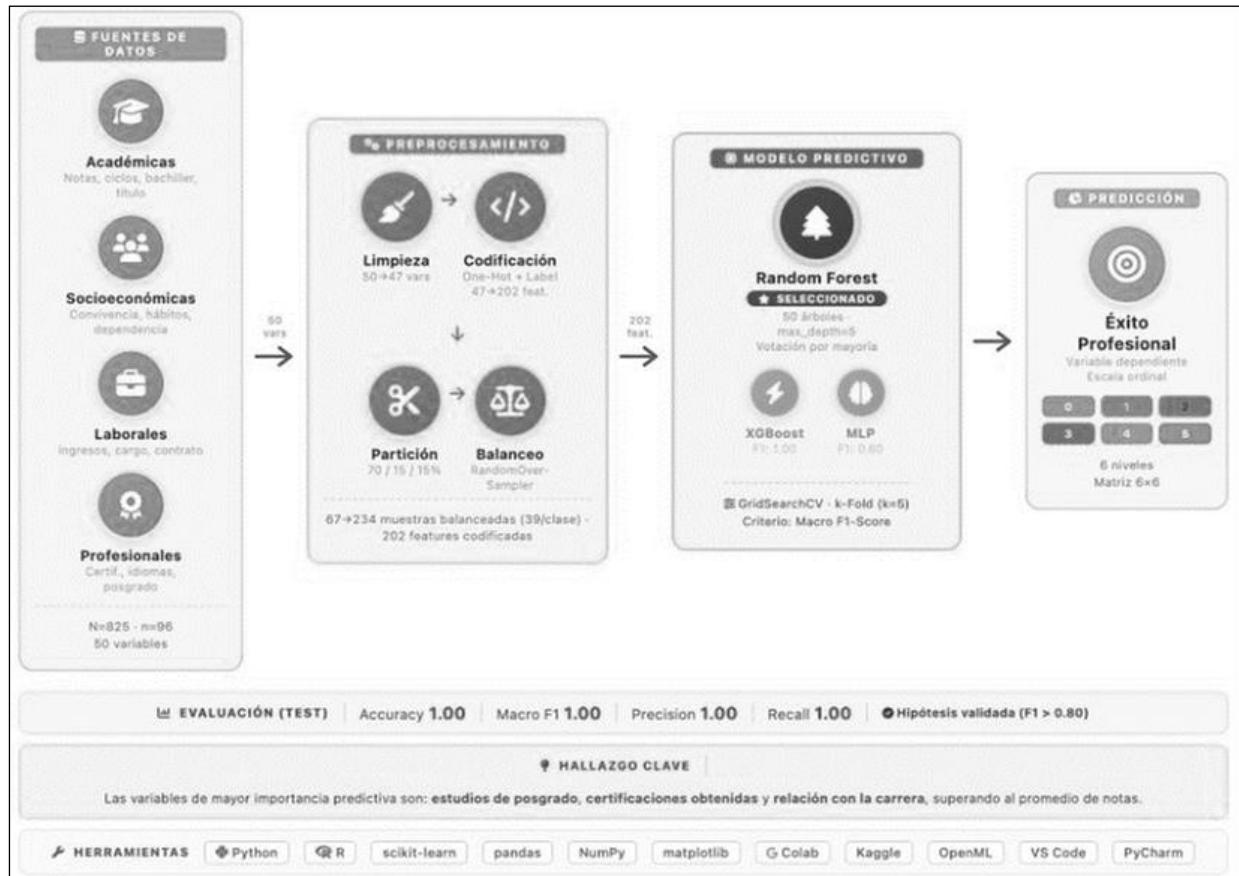
RESULTADOS Y DISCUSIÓN

El desarrollo de la investigación permitió lograr los resultados que se presentan a continuación en función de los objetivos del estudio, de los cuales los objetivos específicos contribuyeron al logro del objetivo general, el cual comprendió el desarrollo del modelo predictivo basado en Aprendizaje Automático para estimar el grado de correlación entre el éxito profesional de los egresados de la Escuela Profesional de Ingeniería de Sistemas e Informática de la UNS y sus características académicas y socioeconómicas, con el propósito de contribuir a la optimización de las estrategias educativas y de apoyo estudiantil.

En cuanto a este objetivo general, el desarrollo del modelo predictivo siguió una secuencia lógica en su construcción, la cual se aprecia en la siguiente figura, detallándose cada una de sus cuatro fases.



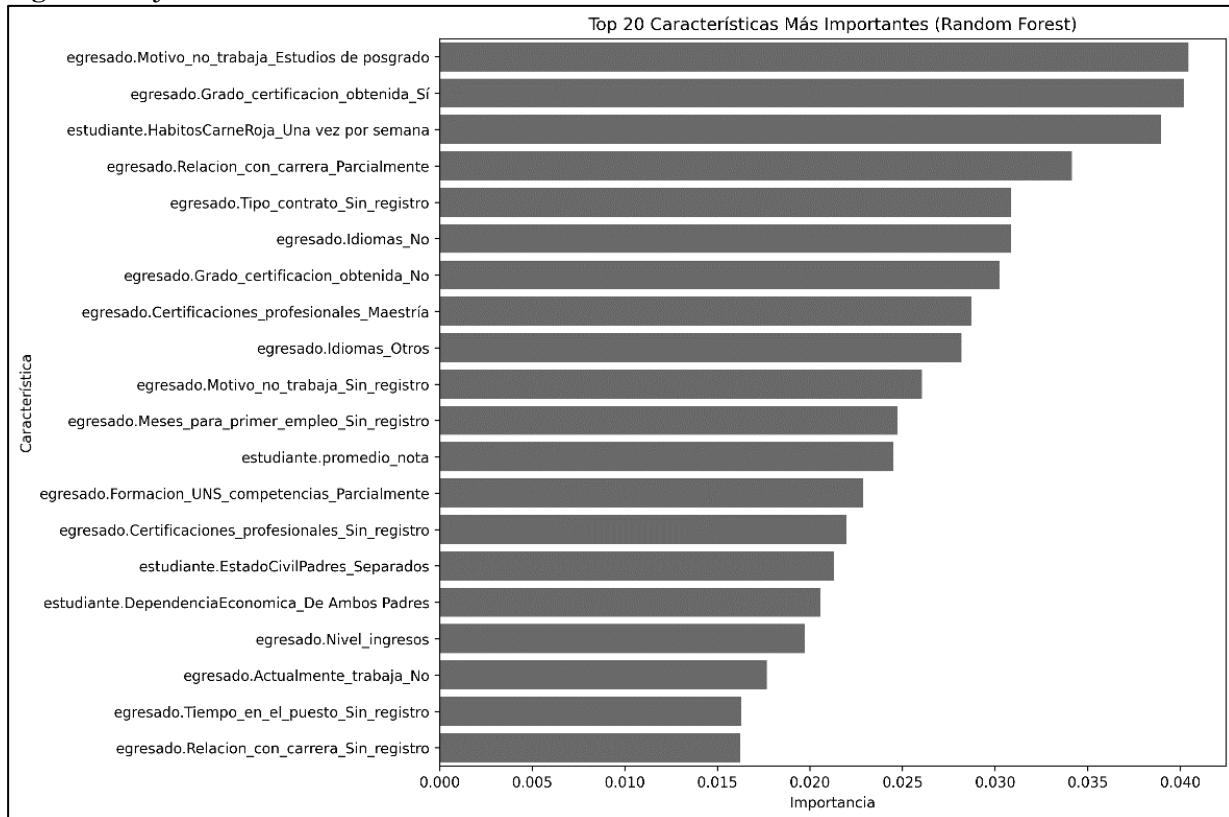
Figura 1 Representación del modelo predictivo.



Nota. La figura 1 presenta la representación general del modelo predictivo diseñado en esta investigación, integrando las fuentes de datos, las etapas de preprocesamiento, el clasificador seleccionado y la variable de salida. Fuente: Elaboración propia utilizando la herramienta diagrams.net.

Para el cumplimiento del primer objetivo específico, el cual se enfocó en el análisis de los datos de los egresados con la finalidad de determinar las variables que incidieron en su éxito profesional se utilizó el script de interpretación (*6_3_model_interpretation.py*) llegándose a mostrar las veinte variables más influyentes en el éxito profesional.

Figura 2 Objetivo 1: Análisis de datos.

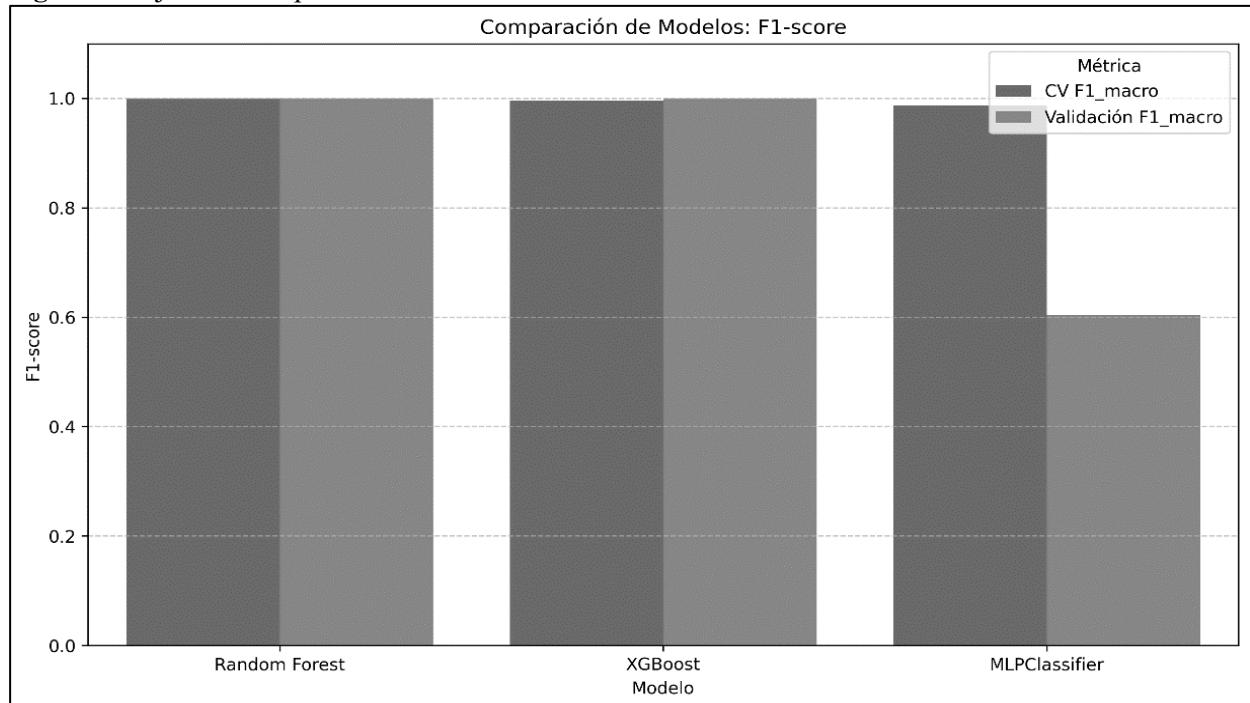


Nota. Feature Importance - Característica más importante top 20 (6_3_model_interpretation.py). Fuente: Elaboración propia mediante la biblioteca matplotlib en el entorno Google Colab.

El análisis de importancia de variables (Figura 2) reveló que factores como estudios de posgrado (egresado.Motivo_no_trabaja_Estudios de posgrado), obtención de certificaciones (egresado.Grado_certificacion_obtenida_Sí) y la relación del empleo con la carrera presentaron mayor peso predictivo que el promedio de notas académico. Esto evidencia la multidimensionalidad del éxito profesional, donde la formación continua supera al rendimiento académico tradicional.

Respecto al segundo objetivo de la investigación, se buscó la implementación del modelo predictivo basado en algoritmos de aprendizaje automático que integre variables identificadas para estimar el éxito profesional. Se entrenaron tres algoritmos de aprendizaje automático tras aplicar oversampling para balancear las clases minoritarias. Se compararon los modelos Random Forest, XGBoost y MLPClassifier por macro F1-score tanto en validación cruzada (*CV F1_macro*) como en la evaluación sobre el conjunto de validación.

Figura 3 Objetivo 2: Implementación del modelo.

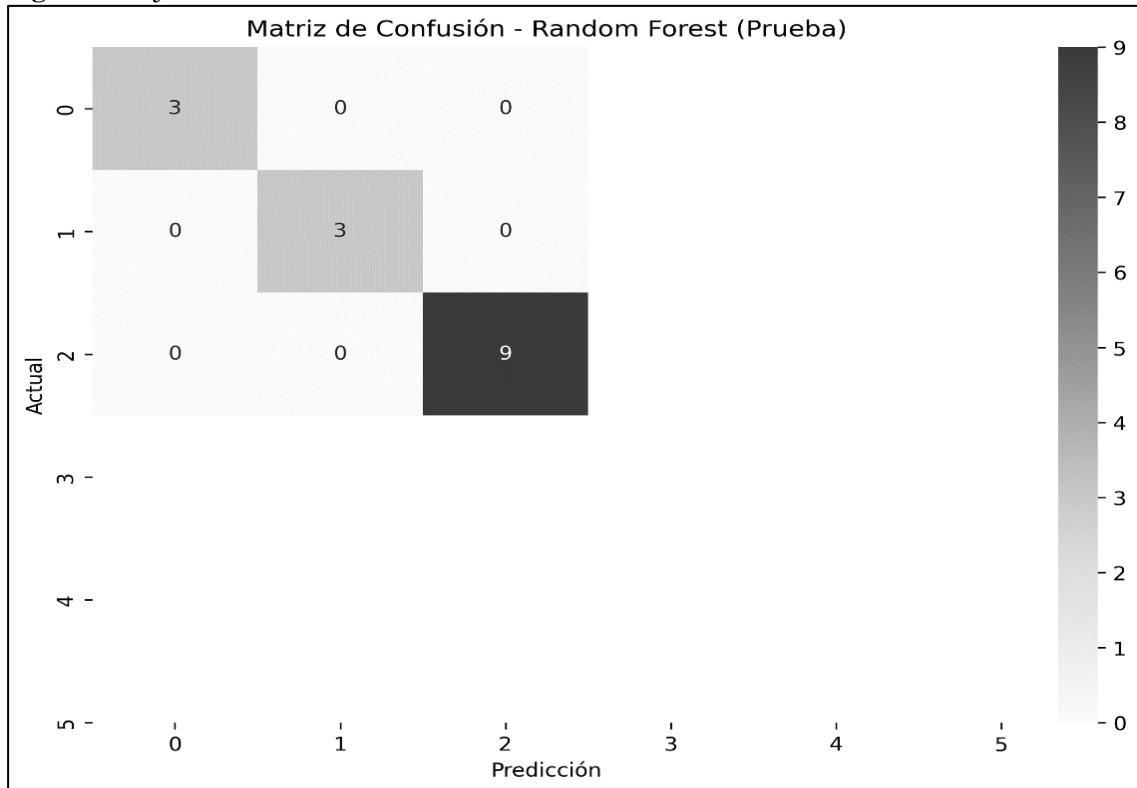


Nota. Comparación de modelos por marco F1-Score (con 6_modeling.py). Fuente: Elaboración propia mediante la biblioteca seaborn en el entorno Google Colab.

Los resultados (Figura 3) mostraron que Random Forest y XGBoost alcanzaron 100% de precisión y macro F1-score de 1.00 tanto en validación como en prueba, superando ampliamente el umbral del 80% establecido en la hipótesis. MLPClassifier obtuvo 85.71% de exactitud y 0.60 de macro F1-score.

Respecto a la validación de la eficacia y de la precisión del modelo, establecidas en el tercer objetivo específico, se realizaron comparando las predicciones generadas con los datos reales de desempeño profesional, aplicando técnicas de validación cruzada y análisis estadístico.

Figura 4 Objetivo 3: Validación.



Nota. Matriz de Confusión empleando Random Forest – test (6_modeling.py). Elaboración propia mediante la biblioteca seaborn en el entorno Google Colab.

La matriz de confusión del conjunto de prueba (Figura 4) confirmó la clasificación perfecta de las clases presentes (0, 4 y 5) en las 15 muestras evaluadas, validando la eficacia del modelo Random Forest seleccionado.

El objetivo específico cuatro buscó derivar implicaciones y proponer recomendaciones para la mejora en la estrategia educativa y de apoyo institucional, a partir de los hallazgos obtenidos, permitiendo optimizar la intervención en la inserción laboral de los egresados.

En base al análisis de la figura 1, tenemos como resultado las recomendaciones que se realizan en base a los hallazgos. En tal sentido la UNS debería fortalecer programas de mentoría sobre posgrados y certificaciones profesionales, dado su alto peso en la predicción del éxito, así como implementar este modelo como sistema de alerta temprana para identificar estudiantes que requieran intervención.

Los resultados obtenidos (100% de precisión) superan ligeramente a los reportados por Baffa et al. (2023), quienes alcanzaron 98% con Random Forest para predecir empleabilidad en Nigeria. Sin embargo, debe considerarse que el tamaño muestral reducido (96 registros) en comparación con estudios internacionales podría explicar esta precisión perfecta, sugiriendo posible sobreajuste. Talero (2023)

trabajó con cientos de registros de egresados en Colombia. Esta limitación se mitigó parcialmente con RandomOverSampler, expandiendo a 234 muestras de entrenamiento, pero representa una diferencia metodológica importante que debe considerarse al interpretar la generalización de los resultados. En cuanto a similitudes, también con un resultado de 100% de exactitud, ElSharkawy et al. (2022), reportaron esta totalidad con Decision Tree para egresados de TI en Egipto, lo cual coincide con lo obtenido en esta investigación mediante Random Forest y XGBoost. Esto sugiere que los algoritmos basados en árboles son particularmente efectivos en contextos educativos con variables categóricas predominantes, aunque el tamaño muestral reducido de ambos estudios obliga a interpretar estos resultados con cautela.

Además del tamaño muestral y los resultados finales, desarrollados en líneas previas, es preciso señalar que en estudios que antecedieron al presente, la forma en que se organizó el recabo de datos puede generar coincidencias o marcar diferencias en cuanto a los resultados. Tal es el caso de Casuat y Festijo (2020), con 27,000 registros de egresados filipinos, lograron 91.22% de exactitud mediante SVM con SMOTE. Si bien el presente estudio alcanza valores superiores, cabe señalar que aquellos autores emplearon clasificación binaria (empleable/no empleable), mientras que aquí se utilizó una escala multinivel (0-5), lo cual dificulta una comparación directa, pero aporta mayor granularidad diagnóstica. Respecto a los diversos factores que pueden generar predicción de éxito en los egresados, a diferencia de Bedoya et al. (2019), quienes identificaron el nivel educativo de la madre como predictor principal, esta investigación encontró que los estudios de posgrado y certificaciones profesionales tienen mayor peso que las variables familiares, lo cual podría reflejar diferencias contextuales o la evolución del mercado laboral hacia la valoración de la formación continua. Se han encontrado también similitudes entre los resultados de esta investigación con la de Haque et al. (2024), quienes habiendo trabajado con datos del Ministerio de Educación de Malasia, alcanzaron 80% de exactitud con redes neuronales artificiales y coinciden en un punto clave: los factores socioeconómicos y de bienestar (satisfacción con servicios universitarios, hábitos personales) influyen significativamente en la empleabilidad, más allá de las calificaciones académicas. Asimismo, se coincidió con Jayachandran y Joshi (2024), quienes utilizando XGBoost optimizado con TLBO en una muestra de egresados de ingeniería en India, alcanzaron 87.8% de exactitud y su hallazgo más relevante fue que los factores socioeconómicos tienen



peso comparable a los académicos, esto refuerza lo observado en el modelo de la UNS respecto a la importancia de la formación continua (posgrados, certificaciones) frente al promedio de notas tradicional.

Finalmente, en cuanto al algoritmo utilizado, mientras Bhagavan et al. (2020) desarrollaron un algoritmo híbrido HLVQ específico, nuestro enfoque priorizó algoritmos estándar interpretables. Esta decisión se alinea con Caselli (2021) en la misma UNS, quien también utilizó modelos interpretables para el seguimiento estudiantil, aunque él alcanzó 98.97% en predicción de deserción versus nuestro 100% en éxito profesional.

Todos estos hallazgos ofrecen a la UNS una herramienta diagnóstica para fortalecer la inserción laboral de sus egresados, representando una transición del seguimiento descriptivo tradicional hacia un enfoque predictivo alineado con las exigencias de SUNEDU y SINEACE.

CONCLUSIONES

Se recolectaron y examinaron datos académicos, socioeconómicos y laborales de 96 egresados, identificando que variables como la realización de estudios de posgrado, las certificaciones profesionales y la relación del empleo con la carrera muestran una influencia notable en el éxito profesional.

Se desarrolló un modelo predictivo fundamentado en algoritmos de aprendizaje automático (Random Forest, XGBoost y MLPClassifier). El proceso de codificación (One-Hot, Label Encoding) y la aplicación de oversampling permitieron manejar la presencia de clases minoritarias (por ejemplo, las categorías 1 y 2 del éxito profesional).

Con base en la partición de datos (70%-15%-15%) y la comparación de predicciones con observaciones reales, se validó la eficacia del modelo. Los dos algoritmos líderes (Random Forest y XGBoost) alcanzaron exactitud y macro F1-score del 100% en validación y prueba, superando así la meta mínima ($\geq 80\%$). Estos resultados confirman la capacidad del modelo para representar adecuadamente el fenómeno, aunque también podrían reflejar un posible sobreajuste dada la muestra reducida.

El análisis de la importancia de variables reveló que factores socioeconómicos y de formación continua (certificaciones, posgrados) tienen un peso clave, a menudo por encima de la nota promedio. Con ello



se evidenció la necesidad de estrategias institucionales centradas no solo en el rendimiento académico, sino también en la vinculación laboral y el apoyo a la formación posgraduada de los estudiantes.

El modelo predictivo confirmó la hipótesis de que la combinación de **datos académicos y socioeconómicos** puede pronosticar con alta precisión el éxito profesional (macro F1-score de 1.00 en el conjunto de prueba).

La excelencia en la clasificación de las clases más comunes (“0” y “5”) y el rescate de clases minoritarias mediante *oversampling* sugieren que el sistema maneja la diversidad de perfiles, aunque la escasa representación original de algunas categorías (1, 2, 3) amerita un seguimiento cuidadoso de la generalización.

El **promedio de notas** mantiene un rol relevante, pero factores como “*egresado.Motivo_no_trabaja_Estudios de posgrado*” y la obtención de “*egresado.Grado_certificacion obtenida_Sí*” destacan con un peso superior al 0.04 en la escala de importancia relativa, reflejando la visión multidimensional del éxito (Murphy, 2022; Müller & Guido, 2016).

Los resultados superan los benchmarks internacionales reportados en la literatura (Baffa et al., 98%; Bhagavan et al., con HLVQ), validando la efectividad del enfoque, aunque con la salvedad del tamaño muestral limitado comparado con estos estudios.

Al predecir el éxito profesional con base en información diversa, la investigación aporta un instrumento de **diagnóstico proactivo** para la Escuela Profesional de Ingeniería de Sistemas e Informática, con posible aplicación y escalamiento a otras carreras de la Universidad Nacional del Santa.

El modelo, junto a las gráficas de interpretación, respalda la premisa de que la formación complementaria (posgrados, certificaciones) y la alineación del empleo con la carrera profesional son aspectos sustanciales en la satisfacción y la inserción laboral de los egresados.



REFERENCIAS BIBLIOGRÁFICAS

- Baffa, M. H., Miyim, M. A., & Dauda, A. S. (2023). Machine learning for predicting students' employability. *UMYU Scientifica*, 2(1), 241-253. http://doi.org/10.56919/usci.2123_001
- Baquero Pérez, J., & Ruesga Benito, S. M. (2019). Factores determinantes del éxito en la inserción laboral de los estudiantes universitarios: El caso de España. *Atlantic Review of Economics*, 2(2), 1-24. <https://hdl.handle.net/10419/213800>
- Bedoya Herrera, O.M., López Trujillo, M. & Marulanda Echeverry, C.E. (2019). Modelo predictivo para la identificación de factores socioculturales asociados al tiempo de búsqueda del primer empleo en egresados universitarios. *Revista Virtual Universidad Católica del Norte*, 58, 3-18. <https://www.redalyc.org/articulo.oa?id=194260979002>
- Bhagavan, K. S., Thangakumar, J., & Subramanian, D. V. (2020). Predictive analysis of student academic performance and employability chances using HLVQ algorithm. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-019-01674-8>
- Bunge, M. (2014). *La ciencia, su método y su filosofía*. Siglo XXI Editores. <https://archive.org/details/bunge-mario.-ciencia-su-metodo-y-su-filosofia-2014>
- Caselli Gismondi, H. E. (2021). *Modelo predictivo basado en Machine Learning como soporte para el seguimiento académico del estudiante universitario* [Tesis doctoral, Universidad Nacional del Santa]. <https://repositorio.uns.edu.pe/handle/20.500.14278/3804>
- Casuat, C. D., Festijo, E. D., y Alon, A. S. (2020). Predicting students' employability using support vector machine: A SMOTE-optimized machine learning system. *International Journal of Emerging Trends in Engineering Research*, 8(5), 2101-2106. <https://doi.org/10.30534/ijeter/2020/102852020>
- ElSharkawy, G., Helmy, Y., y Yehia, E. (2022). Employability prediction of information technology graduates using machine learning algorithms. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(10), 359-367. <https://doi.org/10.14569/IJACSA.2022.0131043>



Haque, R., Quek, A., Ting, C.-Y., Goh, H.-N., y Hasan, M. R. (2024). Classification techniques using machine learning for graduate student employability predictions. *International Journal on Advanced Science, Engineering and Information Technology*, 14(1), 45-56.

<https://doi.org/10.18517/ijaseit.14.1.19549>

Hernández-Sampieri, R., Fernández-Collado, C., y Baptista-Lucio, P. (2014). *Metodología de la investigación (6.ª ed.)*. McGraw-Hill Education.

https://apiperiodico.jalisco.gob.mx/api/sites/periodicooficial.jalisco.gob.mx/files/metodologia_de_la_investigacion_-roberto_hernandez_sampieri.pdf

Hernández-Sampieri, R., y Mendoza-Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.

http://www.biblioteca.cij.gob.mx/Archivos/Materiales_de_consulta/Drogas_de_Abuso/Articulos/MetodologiaInvestigacionRutas.pdf

Hung, J. L., Hsu, Y. C., & Rice, K. (2012). Integrating data mining in program evaluation of K-12 online education. *Journal of Educational Technology & Society*, 15(3), 27-41.

<https://www.jstor.org/stable/jeductechsoci.15.3.27>

Jayachandran, S., y Joshi, B. (2024). Customized support vector machine for predicting the employability of students pursuing engineering. *International Journal of Information Technology*, 16, 3193-3201. <https://doi.org/10.1007/s41870-024-01818-w>

Kerlinger, F. N., y Lee, H. B. (2002). *Investigación del comportamiento: Métodos de investigación en ciencias sociales (4.ª ed.)*. McGraw-Hill Interamericana.

<https://archive.org/details/investigaciondel0000kerl>

Lavado, P., Martínez, J. J., & Yamada, G. (2014). *¿Una promesa incumplida? La calidad de la educación superior universitaria y el subempleo profesional en el Perú* (Documento de Trabajo No. 23). Asociación Peruana de Economía. <https://perueconomics.org/wp-content/uploads/2014/01/WP-23.pdf>

Ministerio de Educación. (2020). *Plan Nacional de Educación Superior y Técnico-Productiva: Decreto Supremo N° 012-2020-MINEDU*. <https://repositorio.minedu.gob.pe/handle/20.500.12799/6921>



Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press.

<https://probml.github.io/pml-book/book1.html>.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4), 1432-1462.

<https://doi.org/10.1016/j.eswa.2013.08.042>

SUNEDU. (2023). *Universidades licenciadas*. SUNEDU. <https://www.sunedu.gob.pe/lista-de-universidades-licenciadas/#:~:text=Lista%20de%20universidades%20licenciadas,se%20han%20otorgado%2096%20licenciamientos>

Talero Távara, S. (2023). *Herramienta para la predicción de retiro de afiliados de la Asociación de Egresados de la Universidad de los Andes* [Trabajo de grado, Universidad de los Andes].

<https://repositorio.uniandes.edu.co/entities/publication/15216dcb-95b2-4de6-823c-6c53625973a9>

Universidad Nacional del Santa. (2017). *Plan Estratégico Institucional 2017-2019*. Dirección de Planificación.

Universidad Nacional del Santa. (2022). *Plan Estratégico Institucional 2019-2025*.

[https://www.uns.edu.pe/transparencia/recursos/410e309afdbc1875cac33870a50e4fcc.%20\(1\).pdf](https://www.uns.edu.pe/transparencia/recursos/410e309afdbc1875cac33870a50e4fcc.%20(1).pdf)

Universidad Nacional del Santa. (2023). *Misión de la Universidad Nacional del Santa*.

<https://www.uns.edu.pe/#/universidad/mision>

Yamada, G., Castro, J. F., & Medina, S. (2019). *Cuando la educación no cumple su promesa: Brechas persistentes en habilidades básicas de peruanos del milenio* (Documento de Trabajo No. 157).

Asociación Peruana de Economía. <https://perueconomics.org/wp-content/uploads/2019/12/WP-157.pdf>

