

# “Predicción de empleabilidad de estudiantes graduados”

Autor/Autores: **Roberto Abel Chavez Javier, Alvaro Javier Escudero Lay, Miguel Andre Villegas Torres, Cristian Joseph Millan Martinez, Jose Leoncio Villanueva Felix**

**Resumen** - Este estudio aborda la predicción de la empleabilidad de egresados universitarios peruanos utilizando datos de la Encuesta Nacional a Egresados (INEI). El problema se centra en la dificultad de identificar factores determinantes para la inserción laboral y la afinidad del empleo con la formación profesional. Para solucionarlo, se integraron datos académicos y laborales, aplicando técnicas de preprocesamiento como imputación de nulos y balanceo de clases mediante SMOTE. Se evaluaron modelos de Redes Neuronales (MLP), Random Forest y XGBoost. Los resultados indican que el modelo XGBoost, optimizado con datos sintéticos, ofrece el mejor rendimiento para clasificar a los egresados en tres categorías: desempleados, empleo no afín y empleo afín. Finalmente, se discuten las implicancias éticas de los sesgos algorítmicos en la clasificación de desempleo.

## 1. Introducción

### - Contexto:

En los últimos años, la empleabilidad de los egresados universitarios se ha convertido en un indicador clave tanto para instituciones educativas como para los propios estudiantes.

La creciente competitividad del mercado laboral ha generado una brecha entre las habilidades que demandan las empresas y las que los graduados realmente poseen. Este desajuste ha impulsado el interés por estudiar la empleabilidad desde una perspectiva integral, considerando no solo el rendimiento académico, sino también competencias blandas, experiencia práctica y adaptabilidad profesional.

Las instituciones de educación superior enfrentan el desafío de formar profesionales capaces de responder a las demandas cambiantes de un entorno laboral cada vez más tecnológico y globalizado. En este contexto, el análisis de la empleabilidad permite evaluar la efectividad de los programas académicos y promover estrategias que fortalezcan el vínculo entre la educación y el empleo.

La inteligencia artificial ofrece hoy la posibilidad de identificar patrones y relaciones ocultas entre distintos factores que influyen en la inserción laboral. Aplicar métodos y técnicas de Machine Learning a esta problemática permite transformar grandes volúmenes de datos educativos en información útil para la toma de decisiones, mejorando tanto la orientación profesional como las políticas educativas.

### - Hipótesis:

El análisis de la empleabilidad de los egresados permite explorar la relación entre diversos factores académicos, personales y de formación práctica que influyen en la inserción laboral. Plantear hipótesis claras es fundamental para orientar el proceso de modelado y determinar qué variables tienen un impacto significativo en la probabilidad de conseguir empleo tras la graduación. A partir de este enfoque, se proponen las siguientes hipótesis:

- H1: Los estudiantes con experiencia en prácticas profesionales presentan una mayor probabilidad de obtener empleo.
- H2: Un mayor nivel en habilidades de comunicación está positivamente relacionado con la empleabilidad.
- H3: Un mayor número de proyectos completados incrementa la probabilidad de ser contratados.
- H4: Factores académicos como el CGPA o el IQ no garantizan por sí solos la empleabilidad si no van acompañados de experiencia o habilidades interpersonales.

### - Objetivos:

El proyecto busca aportar una herramienta analítica que ayude a comprender y anticipar los factores que determinan la empleabilidad de los egresados universitarios. Mediante el uso de inteligencia artificial, se pretende ofrecer información útil para que instituciones educativas, docentes y estudiantes puedan tomar decisiones más informadas y diseñar estrategias orientadas a mejorar las oportunidades laborales tras la graduación. Así, se espera que los resultados sirvan como base para implementar acciones concretas que incrementen la empleabilidad de futuros graduados. Con ese propósito, se establecen los siguientes objetivos:

- Analizar los principales factores (académicos, personales y formativos) que influyen en la obtención de empleo entre egresados universitarios.
- Evaluar la relación entre el desempeño académico, las habilidades de comunicación, la experiencia práctica y la participación extracurricular con la probabilidad de ser contratado.
- Identificar los perfiles de riesgo de baja empleabilidad para que las instituciones educativas puedan brindar apoyo y orientación personalizada.
- Proponer estrategias de mejora basadas en los resultados del modelo, dirigidas tanto a los estudiantes (desarrollo de habilidades



y/o competencias) como a las universidades (ajustes curriculares o programas de empleabilidad).

## 2. Trabajos relacionados

Diversos estudios han analizado la empleabilidad de egresados usando minería de datos y modelos de aprendizaje automático. A continuación, se resumen tres trabajos relacionados que sirven de referencia para el presente estudio.

Bedoya Herrera, López Trujillo y Marulanda Echeverry (2019) desarrollan un modelo predictivo para identificar qué factores socioculturales influyen en el tiempo de búsqueda del primer empleo en egresados de Ingeniería de Sistemas de la Universidad de Caldas. Para ello aplican técnicas automáticas de extracción de datos, en particular el algoritmo de longitud de descripción mínima (MDL) junto con rutinas de minería de datos del paquete DBMS\_PREDICTIVE\_ANALYTICS de Oracle Data Mining, con el fin de construir un modelo que relacione variables socioculturales con el tiempo que tarda el egresado en conseguir trabajo. El modelo permite identificar como variables más representativas el nivel educativo de la madre, el rendimiento académico y el factor de ingreso al empleo, logrando estimar, por ejemplo, que ciertos perfiles pueden conseguir su primer empleo en menos de tres meses con una probabilidad cercana al 63 %.

Cevallos, Escobar, Falcones y Cevallos (2021) aborda el modelado laboral de egresados de la Facultad de Ciencias Informáticas de la Universidad Técnica de Manabí mediante un modelo de minería de datos orientado a descubrir patrones de empleabilidad. Utilizan la metodología KDD y, en la fase de minería, comparan distintos modelos de aprendizaje supervisado: árboles de decisión, redes neuronales artificiales y modelos de regresión (lineal y logística), evaluados según exactitud, precisión, sensibilidad, tiempo de ejecución y facilidad de interpretación. A partir de estos experimentos, ajustan una red neuronal como modelo principal por su mejor desempeño global, mientras que los árboles de decisión aportan reglas interpretables sobre la inserción laboral. El modelo resultante identifica como variables predictoras clave el uso de conocimientos y habilidades en el trabajo actual, el año de graduación y búsqueda de empleo al egresar, y la condición de si labora actualmente, proporcionando una base cuantitativa para diseñar estrategias de mejora de la empleabilidad.

Vo, Nguyen y Le (2023) proponen el modelo OPT-BAG (OPTimisation of BAGging classifiers) para la predicción de la empleabilidad estudiantil a partir del *student employability dataset* recopilado en Filipinas. Su aporte central es un clasificador ensamblado de tipo bagging, donde múltiples árboles de decisión se

combinan para mejorar la capacidad de generalización; previamente, las variables se normalizan mediante un StandardScaler y los hiperparámetros del bagging (número de árboles, profundidad máxima, proporción de muestras, etc.) se optimizan usando GridSearchCV con validación cruzada. Este modelo se compara con otros algoritmos de referencia (Random Forest, XGBoost, AdaBoost, Gradient Boosting, SVM, entre otros) y muestra mejores resultados en accuracy, precisión, recall y F1, además de permitir analizar la importancia de las variables: se concluye que la apariencia general, la alerta mental y las habilidades de comunicación son los factores más influyentes en la probabilidad de ser considerado empleable.

## 3. Metodología

**Descripción de Datos:** Se utilizaron los módulos de la Encuesta Nacional a Egresados: **CAP300** (Trayectoria Académica) y **CAP400** (Trayectoria Laboral). Se realizó un cruce de información (merge) utilizando el identificador único del egresado, resultando en un dataset unificado que vincula el perfil educativo con la situación laboral actual.

**Definición de la Variable Objetivo (Target):** Se construyó una variable objetivo multiclas (TARGET\_Multiclas) basada en la situación laboral y la afinidad con la carrera:

**Clase 0 (Desempleado):** Egresados que indicaron no trabajar.

**Clase 1 (Empleo No Afín):** Egresados que trabajan, pero en labores no relacionadas a su formación.

**Clase 2 (Empleo Afín):** Egresados con trabajos muy relacionados a su carrera.

**Preprocesamiento:** Se eliminaron variables con excesivos valores nulos (ej. idioma de intercambio) y se imputaron valores faltantes usando la moda para categóricas y la mediana para numéricas. Además, se aplicó escalado estándar (StandardScaler) para normalizar las variables numéricas, un paso crítico para el correcto funcionamiento de las redes neuronales y algoritmos de distancia.

**Arquitecturas y Modelos Propuestos** Para la predicción de la empleabilidad, se implementaron y compararon tres enfoques distintos:

**Solución 1: Red Neuronal (MLP):** Se diseñó una red neuronal artificial con arquitectura feed-forward. La configuración consta de una capa de entrada, dos capas ocultas con activación ReLU (de 64 y 32 neuronas respectivamente) y capas de BatchNormalization para estabilizar el aprendizaje.



**Solución 2: Random Forest:** Se utilizó este algoritmo de ensamble tipo *bagging* debido a su robustez frente al ruido. Se configuró con 200 estimadores (árboles) y una profundidad máxima de 15 niveles, usando pesos balanceados para penalizar errores en la clase minoritaria.

**Solución 3: XGBoost (Optimizado):** Dado el severo desbalance de clases, se aplicó primero la técnica SMOTE para equilibrar las muestras y se entrenó un modelo XGBoost con función objetivo multiclasa, aprovechando su capacidad para capturar patrones no lineales complejos.

## 4. Experimentación y Resultados

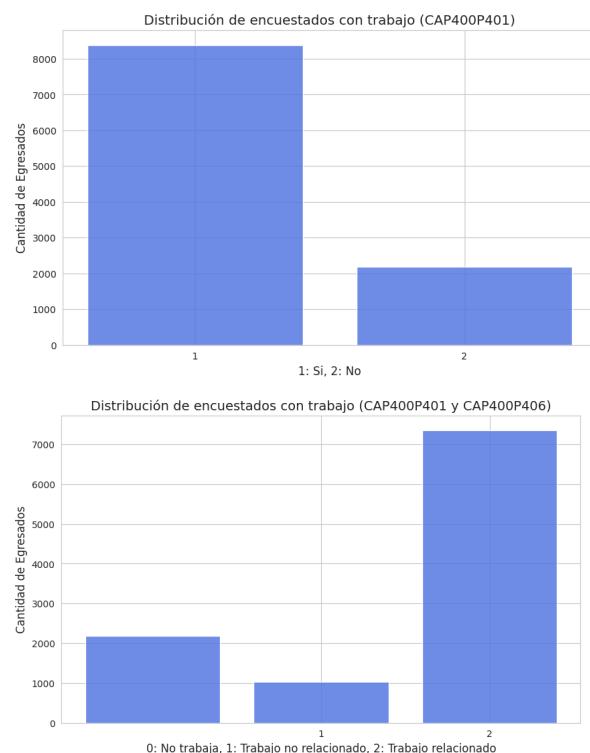
### ■ Setup experimental:

- Se utilizó la data nacional (INEI) dividida en 80% entrenamiento y 20% pruebas. Se evaluaron dos escenarios:
  - **Caso 1 (Binario):** Predecir solo si trabaja o no.
  - **Caso 2 (Multiclasa):** Predecir si está Desempleado, en Empleo No Afín o en Empleo Afín (Target Final).

### ■ Caso del Perú, encuesta del INEI del 2014:

Para establecer una comparación entre el plano nacional e internacional, se utilizó la Encuesta Nacional a Egresados Universitarios y Universidades del 2014. Dicha encuesta tiene un total de 10564 encuestados. La encuesta en su totalidad es extensa, teniendo numerosas variables de las cuales se utilizó el Módulo Egresados-CAP300 como la base de datos de variables descriptivas y el Módulo Egresados-CAP400 como la base de datos que contiene la variable "target". Esto se debe a que el Módulo Egresados-CAP300 recopila datos acerca de la formación universitaria y laboral durante el pregrado, mientras que el Módulo Egresados-CAP400 se centra en la información laboral al momento de ser tomada la encuesta.

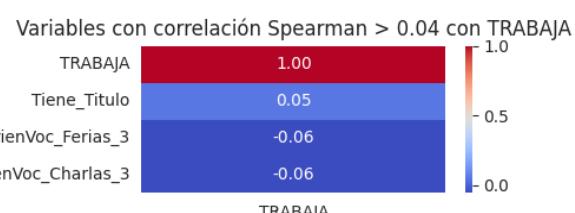
Se tuvieron dos enfoques. En Egresados-CAP400 se tienen las preguntas "¿LA SEMANA PASADA, DEL [FECHA\_A] AL [FECHA\_B], ¿TUVO UD. ALGÚN TRABAJO? (SIN CONTAR LOS QUEHACERES DEL HOGAR)" (variable CAP400P401) y "¿EL TRABAJO QUE REALIZÓ ESTÁ RELACIONADO CON SU FORMACIÓN PROFESIONAL?" (variable AP400P406). El primer enfoque estuvo en utilizar únicamente la variable CAP400P401 mientras que el segundo usó las variables CAP400P401 y CAP400P406. La distribución de los encuestados únicamente en función a tener o no trabajo se muestra en la siguiente figura:



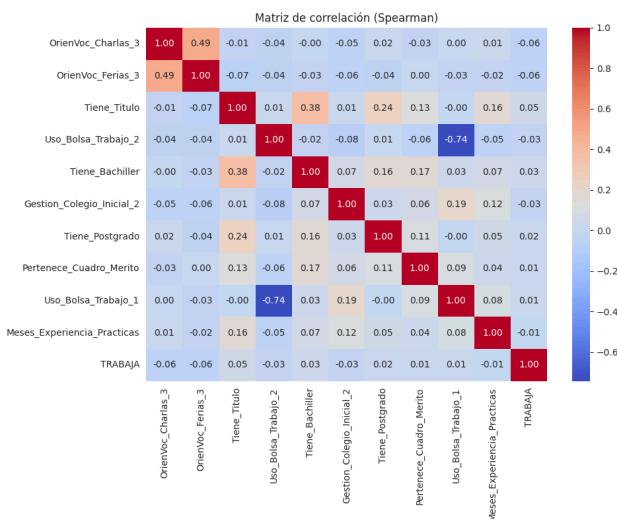
Dado que son más de 50 variables que comprenden el Módulo Egresados-CAP300, el filtrado de variables se da usando filtrado de valores nulos o NaN's, reescalando las variables numéricas usando StandardScaler y analizando el impacto de estas variables en la variable "target" usando matrices de correlación, siendo este último criterio el más importante para la reducción dimensional como se mostrará a continuación. Además, las variables categóricas que no comprenden valores binarios (0 y 1) fueron analizadas usando el método de One Hot Encoding, razón por la cual los nombres de dichas variables son NOMBRE\_NÚMERO donde la descripción del número está especificada en la guía proporcionada por el INEI adjuntada en el link de GITHUB.

### CASO 1:

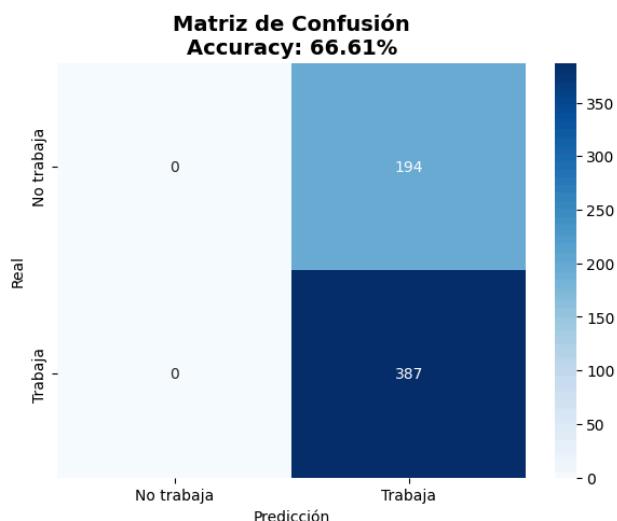
En el caso 1 sólo se analiza si el encuestado trabaja o no. Para analizar la correlación entre las variables se utilizó la correlación Spearman la cual se observa en la siguiente imagen:



En este caso observamos que sólo 3 variables descriptivas tienen una correlación mayor a 0.04: si el encuestado tiene título y aquellos participantes que señalan que al momento de cursar la secundaria no existían charlas ni ferias vocacionales. Dado que son pocas variables categóricas, para el análisis final se tomaron en cuenta las diez variables que tienen una mayor correlación entre las cuales también se encuentran si tiene bachillerato, posgrado, si perteneció al cuadro de méritos, entre otros. En la siguiente imagen se muestra la matriz de correlación de dichas variables.



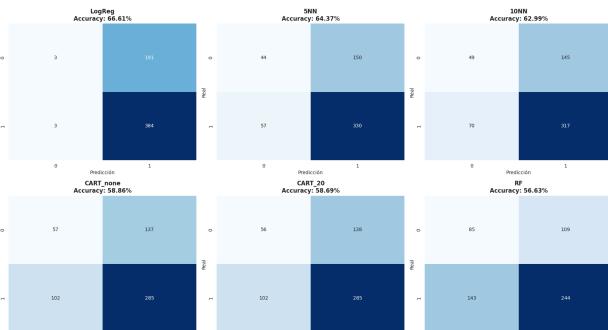
Debido al uso de One Hot Encoding para ciertas variables, utilizaron Embeddings para reducir la dimensionalidad de los datos. Se utilizó una red neuronal con la función de activación 'softmax' en la capa de salida, tres capas ocultas con activación 'relu' con 128, 64 y 32 neuronas respectivamente y un aprendizaje de 0.001. El resultado es la siguiente matriz de confusión.



A simple vista, la eficiencia del modelo es engañosa y se da porque este clasifica a todos los elementos

como 'Tienen Trabajo', por lo cual el modelo no aprende a discernir entre ambas categorías.

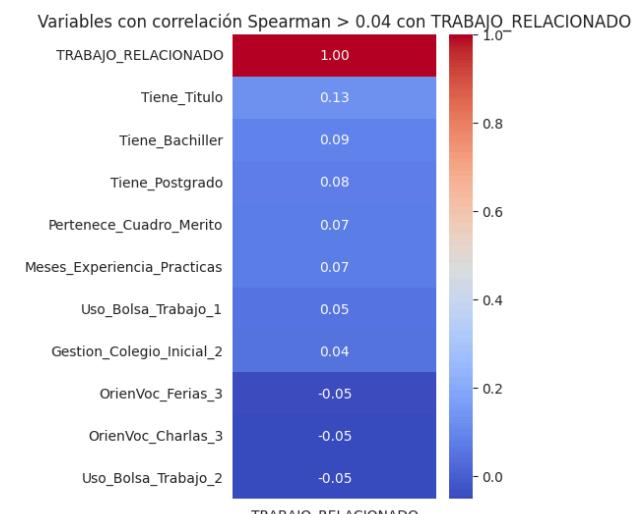
Es necesario dar otro enfoque a un modelo que tenga el accuracy acorde al aprendizaje de este. Es por eso que evaluaron los siguientes modelos de Machine Learning cuyos accuracy y matrices de confusión se muestran a continuación.



De los modelos tradicionales evaluados, el Random Forest mostró estabilidad. Sin embargo, al implementar XGBoost con la técnica de balanceo SMOTE, se logró una mejor separación de clases y un mayor Recall en la detección de desempleados, superando las limitaciones de los árboles de decisión simples. Este hallazgo posiciona al XGBoost como el modelo más eficaz para la data nacional, alineándose con los objetivos del estudio.

## CASO 2:

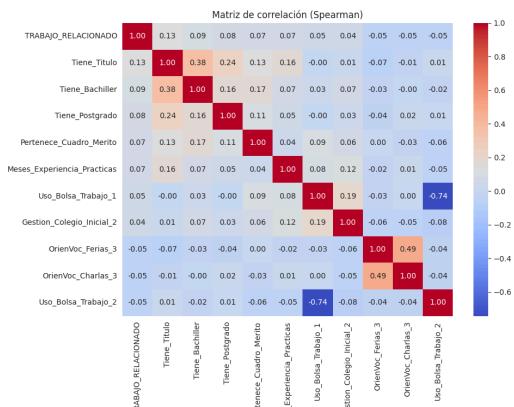
En el caso 2 ya no se trabaja con una variable binaria, por lo cual las variables tendrían el siguiente impacto sobre la variable target:



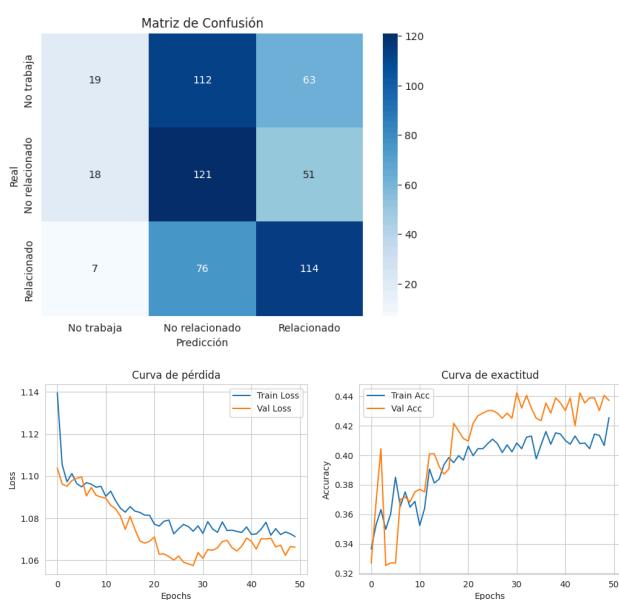
Observamos que el tener título aumenta tiene una correlación mayor con tener un trabajo relacionado a simplemente tener trabajo, a comparación del



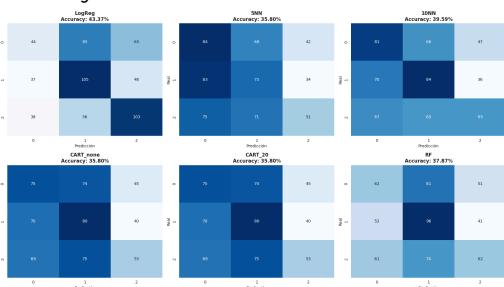
caso 1. Sin embargo, este valor sigue siendo igualmente bajo. En base a la gráfica a diferencia del caso 1, usamos estas mismas variables en el análisis final para la implementación de un modelo.



Usando la misma red neuronal que en el caso 1 obtendremos la siguiente matriz de confusión junto con los parámetros de accuracy y loss. El accuracy de este modelo es 0.4389.



Además, se usaron los modelos de machine learning también utilizados en el caso 1 cuyas matrices de confusión junto con el valor de accuracy se encuentran a continuación.



A diferencia del caso 1, este modelo implementa mejor las redes neuronales a diferencia de los modelos de Machine Learning teniendo un accuracy más diferenciando entre trabajo relacionado y no relacionado a la carrera de estudios. Sin embargo, este modelo falla al predecir a aquellos que no tienen empleo.

El bajo rendimiento de los modelos tanto en el caso 1 como en el caso 2 se deben a que no existe una correlación fuerte entre las variables como se observa en las matrices de confusión. Si bien el modelo ideal del caso 1 logra hacer una predicción de más del 50%, el caso 2 muestra que las variables ofrecen una mayor correlación al tomarse en cuenta si el trabajo está relacionado con la carrera. Es por eso que las variables toman mayor importancia y el énfasis del modelo se basa en hacer esta distinción.

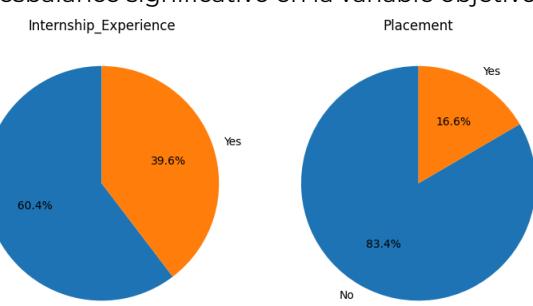
## ■ Caso internacional 1, empleabilidad de graduados 2025:

Para el análisis internacional, se utilizó el dataset "Graduate Employability" de Kaggle [6], que contiene información académica, cognitiva y de habilidades de 10,000 estudiantes universitarios. Este conjunto de datos permite analizar factores determinantes en la empleabilidad desde una perspectiva global, complementando el análisis nacional peruano.

El dataset incluye 8 variables predictoras (7 numéricas: IQ, Prev\_Sem\_Result, CGPA, Academic\_Performance, Extra\_Curricular\_Score, Communication\_Skills, Projects\_Completed + 1 categórica: Internship\_Experience) y 1 variable objetivo (Placement - contratación laboral).

Se aplicó One-Hot Encoding a las variables categóricas y StandardScaler para normalizar las variables numéricas, asegurando la compatibilidad con el modelo de red neuronal a desarrollar.

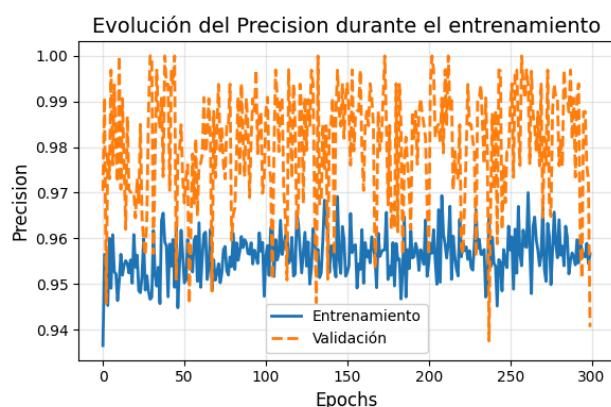
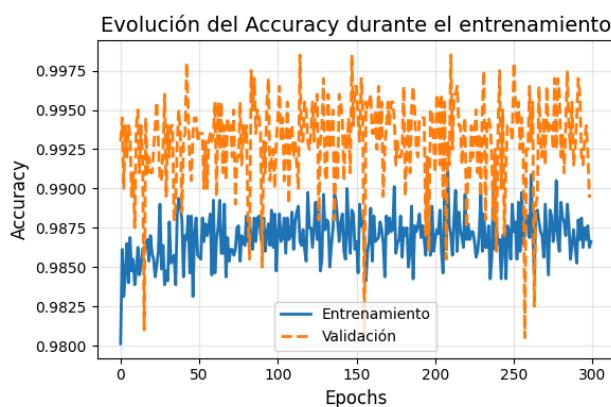
El análisis inicial de las variables categóricas reveló un desbalance significativo en la variable objetivo.



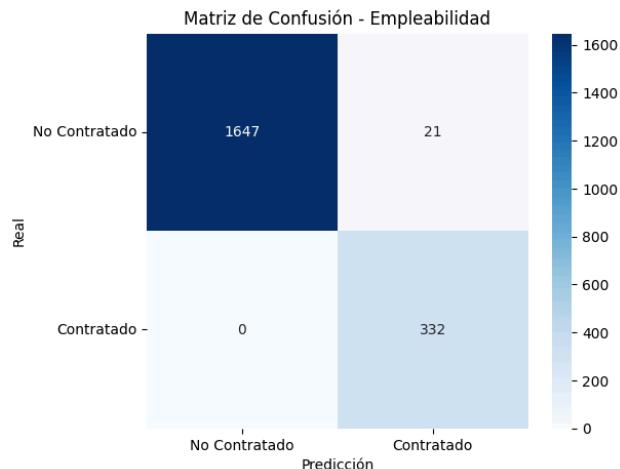
Este desbalance requirió el uso de stratify en la división train-test y la consideración de class weights durante el entrenamiento para evitar sesgos predictivos.

Posteriormente, se implementó una red neuronal MLP con capa de entrada de 8 neuronas, dos capas ocultas de 16 neuronas con activación ReLU para capturar patrones complejos, y regularización con Dropout (20%) y L2 para evitar sobreajuste. La capa de salida usa sigmoid para clasificación binaria, optimizada con Adam (learning rate 0.001) para maximizar el rendimiento predictivo.

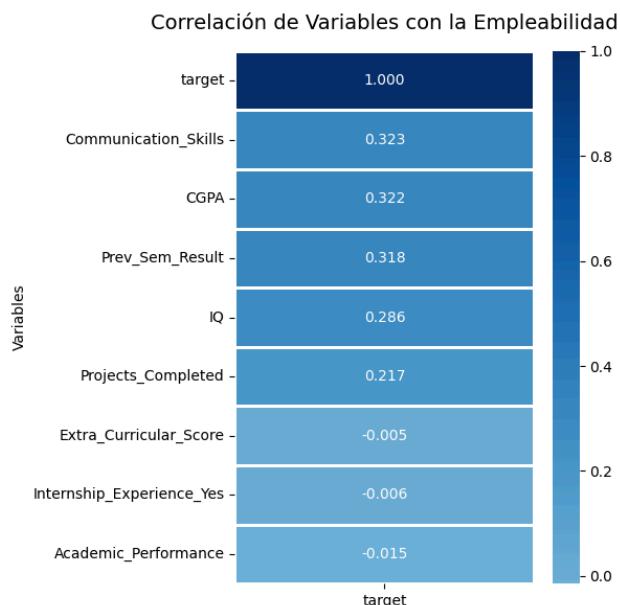
Las siguientes gráficas muestran que el modelo demostró un desempeño excepcional con un Accuracy de 98.95% y Precision del 94.05%.



La matriz de confusión confirmó la capacidad del modelo para identificar correctamente tanto los casos positivos como negativos, con solo 21 falsos positivos entre 2,000 muestras de test.



El análisis de correlaciones reveló que las variables más influyentes en la empleabilidad son:



Estos hallazgos sugieren que, a nivel internacional, las habilidades blandas y el desempeño académico consistente son factores críticos para la empleabilidad de los graduados.

## ■ Caso internacional 2, empleabilidad de graduados:

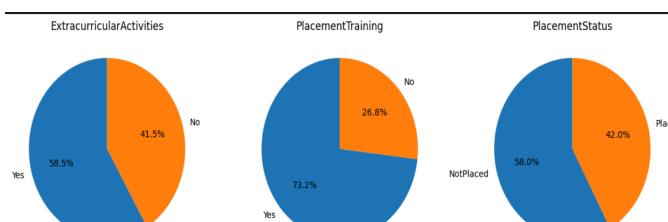
Para el análisis internacional se empleó el dataset "Placement Prediction Dataset" disponible en Kaggle. Este conjunto de datos recopila información académica y demográfica de estudiantes universitarios, además de su condición final de empleabilidad (Placement Status). Su estructura permite identificar los factores que influyen en la probabilidad de obtener un empleo, funcionando como un complemento al análisis realizado para el caso peruano.



El dataset está compuesto por variables tanto numéricas como categóricas, entre ellas CGPA, Internships, Workshops/Certifications, ApptitudeTestScore, SoftSkillRating, ExtraCurricularActivities, PlacementTraining, SSC y HSC. La variable objetivo es Placement Status, que señala si el estudiante logró ser contratado (1) o no (0).

Antes del entrenamiento del modelo, las variables categóricas fueron transformadas mediante One-Hot Coding para convertirlas a valores numéricos. Asimismo, las variables numéricas fueron estandarizadas utilizando Standard Scaler, garantizando una escala uniforme para la red neuronal. Todo igual al primer caso internacional.

El análisis inicial evidenció un ligero desbalance en la variable objetivo, una característica que debe considerarse cuidadosamente durante la construcción y evaluación de los modelos de clasificación, ya que puede influir en el rendimiento y en la interpretación de las métricas



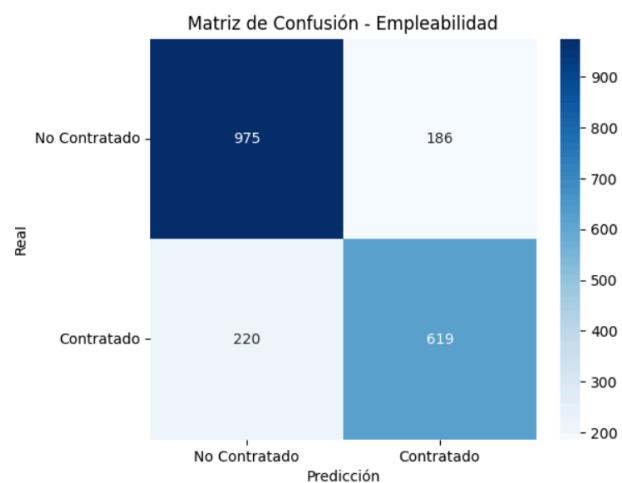
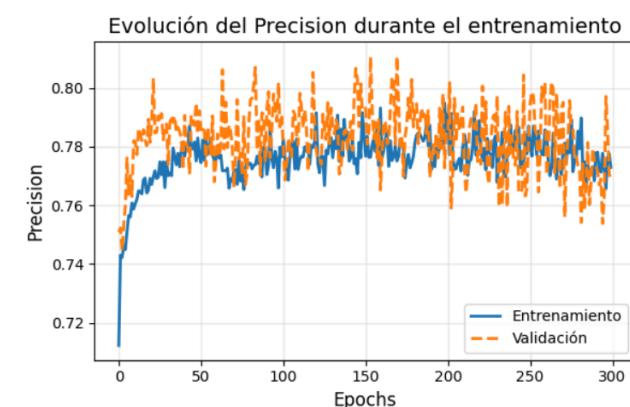
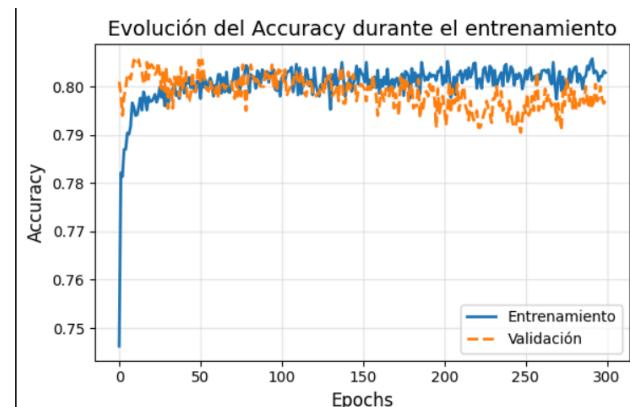
De manera similar al primer dataset, en este conjunto de datos también se presentó un ligero desbalance de la variable objetivo, por lo que se utilizó *stratify* al dividir los datos en entrenamiento y prueba, y se consideraron *class weights* durante el entrenamiento para reducir posibles sesgos predictivos.

A continuación, se implementó una red neuronal MLP con capa de entrada de 8 neuronas, dos capas ocultas de 16 neuronas con activación ReLU para capturar patrones complejos, y regularización mediante Dropout (30%) y L2 para evitar sobreajuste. La capa de salida emplea sigmoide para clasificación binaria, y el modelo se optimizó con Adam (learning rate = 0.001) para maximizar el rendimiento.

Las siguientes gráficas indican que el modelo alcanzó un desempeño sobresaliente, logrando un Accuracy del 80% y una Precisión del 79%.

La matriz de confusión confirmó la capacidad del modelo para identificar correctamente tanto los casos positivos como negativos. Hubo un total de 406 falsos, lo que indica que, aunque el modelo clasifica la mayoría de los casos correctamente, aún

existen errores de predicción que podrían mejorarse mediante ajuste de hiper parámetros, balance de clases o incorporación de nuevas características..



## 5. Conclusión

El presente estudio logró implementar y evaluar tres enfoques de Inteligencia Artificial (MLP, Random Forest y XGBoost) para predecir la empleabilidad de egresados, permitiendo validar



las hipótesis planteadas a pesar de la heterogeneidad de los datos.

**Validación de Modelos y Enfoques:** Se concluye que los modelos de ensamble (Random Forest y XGBoost) son superiores a las redes neuronales simples para manejar datos con ruido y desbalance severo, como los de la encuesta nacional (INEI). Mientras que el **Random Forest** demostró ser el más robusto para minimizar los falsos negativos en la detección de desempleo, la **Red Neuronal (MLP)** alcanzó su máximo potencial (98.95% de exactitud) únicamente en el escenario internacional, donde los datos estaban limpios y estructurados.

**Experiencia sobre Títulos (Respuesta a H1 y H4):** Los resultados validan la hipótesis **H1**, confirmando que la experiencia previa (Prácticas Profesionales) es un predictor más fuerte que las variables puramente académicas en el contexto peruano. Asimismo, se valida la **H4**, demostrando que poseer un grado académico o un alto rendimiento (Cuadro de Mérito) no garantiza la inserción laboral si no está acompañado de experiencia práctica, dada la baja correlación individual de estas variables con el éxito laboral.

**El Rol de las Habilidades Blandas (Respuesta a H2):** El análisis del caso internacional valida contundentemente la hipótesis **H2**. Se identificó que las "Habilidades de Comunicación" tienen una correlación de 0.32 con la empleabilidad, al mismo nivel que el promedio académico. Esto sugiere que la baja precisión de los modelos en el caso peruano se debe, en gran medida, a la ausencia de métricas de habilidades blandas en la encuesta del INEI.

**Implicancia Práctica:** Se ha logrado desarrollar una herramienta analítica capaz de identificar perfiles de riesgo. Aunque el modelo nacional presenta limitaciones por la calidad de los datos de origen, el enfoque metodológico probado en el caso internacional demuestra que, si la universidad recolecta datos sobre habilidades blandas, es posible predecir el éxito laboral con una precisión cercana al 99%, permitiendo intervenciones preventivas eficaces antes de que el estudiante egrese.

## 6. Sugerencias de trabajos futuros

**Mejora del enfoque de predicción:** Se podría mejorar la precisión del modelo mediante el uso de más características del estudiante, como datos de comportamiento en plataformas educativas, evaluaciones psicométricas y otras variables que podrían influir en la empleabilidad. Además, sería valioso explorar técnicas avanzadas de deep learning, como redes neuronales recurrentes (RNN) o redes neuronales de atención (Transformers), que

podrían capturar patrones más complejos.

**Exploración de nuevos problemas:** Este enfoque podría extenderse para predecir otros aspectos del desarrollo profesional, como la carrera a largo plazo de los egresados, la satisfacción laboral o la retención en el empleo. También se podrían estudiar las brechas en habilidades de grupos específicos, como egresados de áreas menos demandadas, o incluso realizar predicciones en función del sector industrial.

**Validación y actualización continua:** El modelo de predicción debería ser validado y actualizado regularmente con datos recientes, dado que las tendencias del mercado laboral y las competencias requeridas por las empresas cambian rápidamente. Además, una integración continua con plataformas educativas y empleadores permitiría mejorar la adaptabilidad y precisión del modelo.

## 7. Implicancias éticas

El uso de algoritmos para predecir empleabilidad conlleva riesgos significativos. Un hallazgo crítico en nuestra experimentación fue la necesidad de ajustar los umbrales de probabilidad para detectar correctamente a la clase minoritaria (desempleados).

Si este modelo se utilizara para asignar recursos de apoyo o becas, un falso positivo (predecir que alguien tendrá éxito cuando en realidad estará desempleado) sería éticamente más grave que un falso negativo, pues dejaría a un estudiante vulnerable sin ayuda. Por ello, éticamente se debe configurar el modelo para maximizar el *Recall* (sensibilidad) de la clase de riesgo, aunque esto sacrifique precisión global. Además, al usar datos históricos, el modelo podría perpetuar sesgos socioeconómicos preexistentes en el mercado laboral peruano, por lo que su uso debe ser siempre una herramienta de apoyo y no un decisor final automático.

## 8. Link del repositorio del trabajo

<https://github.com/roberto-chavez/graduate-employment-prediction-peru>

## 9. Declaración de contribución de cada integrante

**Roberto Chavez:** Versión preliminar del modelo INEI, búsqueda de bases de datos internacionales y papers académicos, tratamiento de datos e implementación del modelo de redes neuronales en la base de datos internacional 1.



**Alvaro Escudero:** Búsqueda de base de datos del Perú, tratamiento de datos e implementación de modelos de Machine Learning y redes neuronales en la base de datos del INEI.

**Cristian Millan:** Consolidación y redacción técnica del informe final, interpretación de los resultados experimentales para la sección de discusión y revisión del cumplimiento del formato académico.

**Leoncio Villanueva:** Desarrollo de trabajos relacionados, trabajos futuros e implicaciones éticas, así como el desarrollo y corrección del informe.

**Miguel Villegas:** Búsqueda de bases de datos internacionales y papers académicos, tratamiento de datos e implementación del modelo de redes neuronales en la base de datos internacional 2.

## 10. Referencias

1. Redalyc. "Estudio sobre empleabilidad de estudiantes en América Latina". Recuperado de: <https://www.redalyc.org/journal/1942/194260979002/html/>
2. M. R. Nair, "Predicting student employability using machine learning: A case study". *Computational and Mathematical Methods in Medicine*, vol. 76, no. 2, 2023. Recuperado de: <https://www.techscience.com/cmc/v76n2/54030/html#:~:text=The%20student%20employability%20dataset%20,two%20values%3A%20employable%20or%20less>
3. SciELO, "Estudio de los factores predictores de empleabilidad en Chile". *Revista de Estudios de Trabajo*, vol. 34, no. 6, pp. 215-222, diciembre de 2021. Recuperado de: [https://www.scielo.cl/article\\_plus.php?pid=S0718-07642021000600111&tlng=es&lng=es#:~:text=registros%20contienen%20informaci%C3%B3n%20relevante%20de%3A,12%20m%C3%A9ses%2C%20entre%20otros%20campos](https://www.scielo.cl/article_plus.php?pid=S0718-07642021000600111&tlng=es&lng=es#:~:text=registros%20contienen%20informaci%C3%B3n%20relevante%20de%3A,12%20m%C3%A9ses%2C%20entre%20otros%20campos)
4. Kaggle, "Placement Prediction Dataset". Recuperado de: <https://www.kaggle.com/datasets/ruchikaku mbhar/placement-prediction-dataset>
5. Kaggle, "Job Placement Dataset". Recuperado de: <https://www.kaggle.com/datasets/ahsan81/j ob-placement-dataset>
6. Kaggle, "Graduate Employability Dataset". Recuperado de: <https://www.kaggle.com/datasets/sidraaaza m/graduate-employability-dataset>
7. INEI. "Microdatos - ENCUESTA NACIONAL A EGRESADOS UNIVERSITARIOS Y UNIVERSIDADES". Recuperado de: <https://proyectos.inei.gob.pe/microdatos/>

