Revista Científica Multidisciplinar G-ner@ndo ISSN: 2806-5905

Análisis Comparativo de algoritmos de Inteligencia Artificial para procesos de selección de personal

Comparative Analysis of Artificial Intelligence algorithms for personnel selection processes

Jonathan Joel Mendoza Vélez & Roberth Abel Alcívar Cevallos

PUNTO CIENCIA.

Julio - diciembre, V°6 - N°2; 2025

Recibido: 15-11-2025
Aceptado: 18-11-2025
Publicado: 20-11-2025

PAIS

- Ecuador, Manabí
- Ecuador, Manabí

INSTITUCION

- Universidad Técnica de Manabí
- Universidad Técnica de Manabí

CORREO:

✉ jmendoza9269@utm.edu.ec
✉ roberth.alcivar@utm.edu.ec

ORCID:

https://orcid.org/0009-0009-5453-6007
https://orcid.org/0000-0001-6282-8493

FORMATO DE CITA APA.

Mendoza, J. & Alcívar, R. (2025). Análisis Comparativo de algoritmos de Inteligencia Artificial para procesos de selección de personal. *Revista G-ner@ndo, V°6 (N°2). Pág. 2925 – 2945.*

Resumen

La Inteligencia Artificial (IA) se ha convertido en una herramienta estratégica para optimizar la selección de personal, al mejorar la precisión en la identificación de candidatos y reducir los tiempos de contratación. Este estudio tuvo como objetivo comparar el desempeño de tres algoritmos de clasificación: Random Forest (RF), Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Profundas (DNN), evaluados mediante métricas como accuracy, recall, F1-score, tiempos de ejecución, gap de generalización y Área Bajo la Curva ROC (AUC). El conjunto de datos utilizado proviene de Kaggle (Predicting Hiring Decisions in Recruitment Data), con 1.500 registros, 10 variables predictoras y la variable objetivo HiringDecision, incluye datos demográficos, historial laboral y evaluaciones del proceso de selección. Se diseñó una metodología propia ajustada a las necesidades de selección de personal, con fases para la preparación de datos, balanceo de clases con SMOTE, entrenamiento de modelos, ajuste de hiperparámetros y análisis de métricas en dos escenarios: sin balanceo y con SMOTE, empleando Python y bibliotecas como Scikit-learn y TensorFlow. Los hallazgos muestran que RF sin balanceo alcanzó el mayor AUC (0.94) y la mejor combinación de accuracy (91%), recall (82%) y F1-score (84%). SVM destacó por su rapidez y bajo gap de generalización (3.2%), y DNN obtuvo un AUC adecuado (0.92), aunque con mayor costo computacional. RF se posiciona como el modelo más apropiado para la selección de personal por su equilibrio entre rendimiento y capacidad predictiva, mientras que SVM se presenta como una alternativa eficaz en escenarios donde la velocidad es prioritaria.

Palabras clave: Inteligencia Artificial, Selección de Personal, Random Forest, SVM, Redes Neuronales.

Abstract

Artificial Intelligence (AI) has become a strategic tool to optimize personnel selection by improving accuracy in candidate identification and reducing hiring times. This study aimed to compare the performance of three classification algorithms: Random Forest (RF), Support Vector Machines (SVM), and Deep Neural Networks (DNN), evaluated through metrics such as accuracy, recall, F1-score, execution times, generalization gap, and Area Under the ROC Curve (AUC). The dataset used comes from Kaggle (Predicting Hiring Decisions in Recruitment Data), with 1,500 records, 10 predictor variables, and the target variable HiringDecision, including demographic data, work history, and selection process evaluations. A custom methodology was designed and adjusted to personnel selection needs, with phases for data preparation, class balancing with SMOTE, model training, hyperparameter tuning, and metrics analysis in two scenarios: without balancing and with SMOTE, using Python and libraries such as Scikit-learn and TensorFlow. The findings show that RF without balancing achieved the highest AUC (0.94) and the best combination of accuracy (91%), recall (82%) and F1-score (84%). SVM stood out for its speed and low generalization gap (3.2%), and DNN obtained an adequate AUC (0.92), although with higher computational cost. RF is positioned as the most appropriate model for personnel selection due to its balance between performance and predictive capacity, while SVM presents itself as an effective alternative in scenarios where speed is a priority.

Keywords: Artificial Intelligence, Personnel Selection, Random Forest, SVM, Neural Networks.

Introducción

En la última década, la Inteligencia Artificial (IA) ha pasado de ser un campo especializado a convertirse en una herramienta ampliamente utilizada gracias a avances en capacidad de cómputo, disponibilidad de datos y algoritmos más eficientes (Pérez López, 2023). En este ámbito, la gestión del talento humano ha encontrado en la IA un aliado para mejorar la eficiencia y la objetividad en la selección de personal (Alameda Castillo, 2021), ya que este proceso ha sido tradicionalmente demandante en tiempo y recursos, y expuesto a sesgos, errores o subjetividad (Quintanilla-Medina & Coral-Ignacio, 2024). Ante esto, los modelos de IA permiten automatizar el análisis, reducir la carga operativa, incrementar la precisión y favorecer decisiones más imparciales (León-Varela et al., 2024).

Los algoritmos de machine learning procesan grandes volúmenes de datos e identifican patrones que predicen la idoneidad de un candidato (Mehdiyev et al., 2025; Raza et al., 2022); técnicas como Random Forest (RF), Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Profundas (DNN) destacan por manejar datos estructurados, adaptarse a distintos contextos y ofrecer métricas sólidas (Medrano et al., 2022). En este trabajo se presenta un análisis comparativo de estos modelos aplicados a un conjunto de datos real, incluyendo preparación de datos, balanceo con SMOTE, entrenamiento y evaluación mediante accuracy, recall, F1-score, tiempo de ejecución, gap de generalización y AUC. Bajo este enfoque surge la pregunta central: ¿Cómo conocer el desempeño de los algoritmos de IA en procesos de selección de personal al comparar precisión, imparcialidad y eficiencia en diferentes etapas del reclutamiento?

Para responder esta cuestión, el estudio se propuso identificar los algoritmos más utilizados, evaluar el rendimiento de cada modelo, analizar fortalezas y limitaciones y determinar cuál ofrece mejores prestaciones, además de formular recomendaciones.

La relevancia de una metodología propia se fundamenta en que cada contexto organizacional presenta características particulares (Paredes Morales, 2023), y como indican Singh y Chakraborty (2024), la personalización de los procesos de IA mejora la interpretabilidad y utilidad de los resultados.

En el Ecuador, la adopción de IA en recursos humanos aún es incipiente (Barragán-Martínez, 2023; Medrano et al., 2022), aunque diversos estudios han mostrado que contribuye a mejorar la eficiencia en la toma de decisiones (Garzón Ayala et al., 2024; Quintanilla-Medina & Coral-Ignacio, 2024). Por ello, este estudio aplicó un diseño experimental que compara el rendimiento con datos originales y balanceados mediante SMOTE, debido a que el desbalance de clases es frecuente (León-Varela et al., 2024; González-Sendino et al., 2024).

La elección de RF, SVM y DNN representa enfoques distintos: RF como método de ensamblado, SVM como algoritmo de optimización y DNN como modelo inspirado en redes neuronales; compararlos bajo una misma metodología permite visualizar sus fortalezas y limitaciones. En conjunto, este trabajo busca determinar cuál modelo ofrece el mejor desempeño predictivo, con el fin de reducir la subjetividad en la contratación y optimizar recursos.

Métodos y Materiales

La presente investigación evaluó el rendimiento de algoritmos de IA aplicados a la selección de personal, respondiendo a la necesidad de sistemas de reclutamiento más eficientes, justos y precisos. Fue una investigación de tipo aplicada, orientada a proponer una solución práctica basada en algoritmos de IA; utilizó un enfoque cuantitativo para asegurar objetividad y reproducibilidad mediante la medición de precision, recall, F1-score, tiempo de ejecución y capacidad de generalización, métricas ampliamente reconocidas para evaluar el desempeño de algoritmos en problemas de clasificación binaria. El diseño experimental permitió comparar sistemáticamente

distintos modelos usando los mismos conjuntos de datos y analizar su desempeño de forma objetiva.

La investigación empleó tres técnicas: revisión documental, revisión bibliográfica y encuesta estructurada. La revisión documental analizó el dataset Predicting Hiring Decisions in Recruitment Data para comprender sus variables y sus limitaciones. La revisión bibliográfica en IEEE Xplore, MDPI Applied Sciences y el International Research Journal of Multidisciplinary Studies permitió seleccionar los modelos (RF, SVM y DNN), definir las métricas y justificar el uso de SMOTE, apoyado por estudios que evidencian su eficacia para mejorar la representación de clases minoritarias (Alameda Castillo, 2021; Pérez López, 2023). Finalmente, se aplicó una encuesta a seis expertos en Recursos Humanos para identificar las variables más relevantes y contrastar estos criterios con los resultados de importancia de variables obtenidos mediante Random Forest.

Para esta investigación se utilizaron herramientas tecnológicas y bibliográficas que facilitaron el procesamiento de datos, la implementación de los modelos y la validación de resultados. En el ámbito de programación se trabajó con Python 3.11.13, empleando Pandas 2.3.0 para la manipulación de datos, NumPy 1.24.4 para cálculos numéricos, Scikit-learn 1.2.2 para el preprocesamiento, los algoritmos y las métricas, e Imbalanced-learn 0.11.0 para aplicar SMOTE. Para la visualización se utilizaron Matplotlib 3.10.3 y Seaborn 0.13.2.

Las redes neuronales profundas se desarrollaron con TensorFlow 2.19.0 y Keras 3.10.0, permitiendo definir la arquitectura DNN, aplicar Dropout y utilizar Early Stopping para evitar sobreajuste. Los datos y resultados se gestionaron mediante archivos CSV, organizando los conjuntos de entrenamiento, prueba y métricas. Además, se aplicó una encuesta a seis expertos en Recursos Humanos mediante Google Forms, con el fin de complementar el análisis sobre la relevancia de las variables del dataset.

Para esta investigación se utilizó el conjunto de datos “Predicting Hiring Decisions in Recruitment Data”, disponible en Kaggle (El Kharoua, 2024). El dataset incluye información sobre edad, experiencia laboral, nivel educativo, habilidades técnicas, desempeño en entrevistas y resultados de pruebas, con un total de 1500 registros, 10 atributos (variables) y una variable objetivo relacionada con la decisión de contratación. Su estructura considera 6 atributos enteros, 2 binarios, 2 categóricos y 1 continuo, como se detalla en la Tabla 1. Este recurso permite analizar de forma completa el perfil de los candidatos y fue elegido por su nivel de detalle y utilidad para estudiar la predicción de decisiones de contratación. Al ser público y actualizado, garantiza transparencia, facilita la replicabilidad de los experimentos y fortalece la validez externa de los resultados.

Tabla 1.

Variables del dataset “Predicting Hiring Decisions in Recruitment Data”

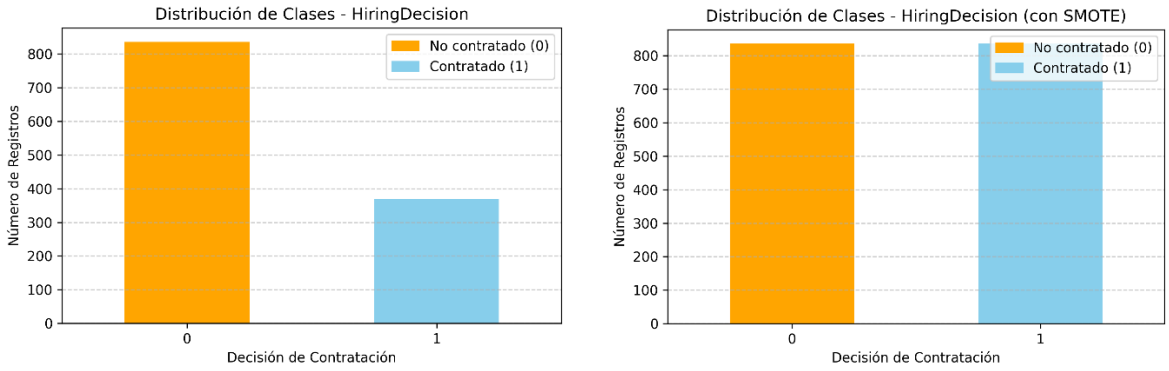
Variable	Descripción	Tipo
Age	Edad del candidato: 20 a 50 años.	Entero
Gender	Género del candidato: Masculino (0) y Femenino (1).	Binario
Education Level	Nivel más alto de educación alcanzado por el candidato: 1: Licenciatura (tipo 1), 2: Licenciatura (tipo 2), 3: Maestría, 4: Doctorado.	Categórico
Experience Years	Número de años de experiencia profesional: 0 a 15 años.	Entero
Previous Companies Worked	Número de empresas anteriores donde el candidato ha trabajado: 1 a 5 empresas.	Entero
Distance From Company	Distancia en km desde la residencia del candidato hasta la empresa contratante: 1 a 50km.	Flotante (continuo)
Interview Score	Puntuación obtenida por el candidato en el proceso de entrevista: 0 a 100.	Entero

Variable	Descripción	Tipo
Skill Score	Puntuación de evaluación de las habilidades técnicas del candidato: 0 a 100.	Entero
Personality Score	Puntuación de evaluación de los rasgos de personalidad del candidato: 0 a 100.	Entero
Recruitment Strategy	Estrategia adoptada por el equipo de contratación para el reclutamiento: 1: Agresivo, 2: Moderado, 3: Conservador.	Categórico
Hiring Decision (variable objetivo)	Resultado de la decisión de contratación: 0: No contratado, 1: Contratado.	Binario (entero)

Previo al uso del dataset se realizó un control de calidad que permitió detectar inconsistencias, principalmente registros donde los años de experiencia eran imposibles respecto a la edad. Por ejemplo, en el registro 6 un postulante de 27 años reportaba 14 años de experiencia, lo que implicaría haber trabajado antes de la edad laboral permitida (18 años). En total, 294 de los 1500 registros (equivalente al 19.6 % del total) presentaron esta anomalía, por lo que se filtraron automáticamente todos los casos donde la experiencia superaba la diferencia entre la edad y la edad mínima legal. También, se eliminó la variable Gender, debido al riesgo ético de introducir sesgos discriminatorios en el modelo. Las demás tareas de limpieza, codificación y normalización se describen en la Fase 3. Además, se verificó el desbalance en la variable objetivo HiringDecision. La Figura 1 muestra una distribución desigual entre candidatos contratados y no contratados, lo que justificó aplicar técnicas de balanceo como SMOTE en etapas posteriores.

Figura 1.

Distribución de la variable objetivo HiringDecision antes y después de equilibrar las clases. (a) Distribución original del conjunto de datos, evidenciando un desequilibrio notable entre las clases "No contratado (0)" y "Contratado. (b) Distribución tras aplicar la técnica de balanceo SMOTE.



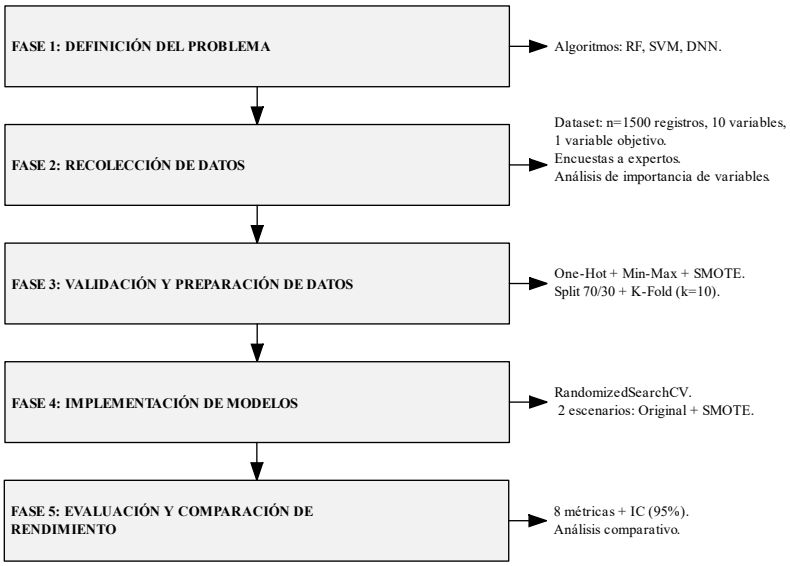
(a) **(b)**

Por esta razón, se decidió aplicar la técnica SMOTE (Synthetic Minority Oversampling Technique), ampliamente utilizada en estudios de atrición laboral y selección de personal. Esta técnica permite aumentar la representatividad de la clase minoritaria mediante instancias sintéticas, reduciendo el sesgo del modelo hacia la clase mayoritaria y mejorando su capacidad de generalización. La literatura respalda su eficacia: Haque et al. (2025) evidenciaron que incorporar SMOTE en modelos como RF optimiza el rendimiento y mejora la detección de patrones de salida laboral. De manera similar, Gómez et al. (2022) demostraron que integrar SMOTE en el preprocesamiento incrementa métricas como recall y F1-score, consolidando su valor en escenarios con alta desproporción entre clases.

Se propuso un modelo metodológico propio compuesto por cinco fases, cada una con un rol clave desde la definición del problema hasta el análisis de resultados. Su estructura garantiza coherencia, justificación en cada paso y facilita la replicabilidad del estudio.

Figura 2.

Modelo metodológico propio de cinco fases para evaluación comparativa de algoritmos de IA.



Fase 1: Definición del problema

Durante la Fase 1 se identificó que muchas organizaciones tienen dificultades en sus procesos de selección debido al alto número de postulantes, el tiempo limitado de los equipos de RR. HH. y la subjetividad en las decisiones, lo que puede llevar a contratar perfiles inadecuados o descartar candidatos aptos. Esto evidencia la necesidad de incorporar herramientas que hagan el proceso más rápido, objetivo y basado en datos. La IA ofrece ese potencial, pero la variedad de modelos disponibles obliga a determinar cuáles funcionan mejor. Dado que la selección de personal es un proceso crítico, su automatización parcial puede mejorar el filtrado de información y la comparación de perfiles, reduciendo errores. Por ello, esta investigación busca identificar qué modelos de IA son más efectivos según su desempeño, accesibilidad y adaptabilidad, ofreciendo orientación para procesos más transparentes y consistentes. Para el estudio se eligieron tres algoritmos ampliamente validados: RF, por su capacidad de evitar sobreajuste; SVM, eficaz para separar clases en conjuntos moderados; y DNN, capaces de aprender patrones complejos. Su selección responde a su uso frecuente en

la literatura, su diversidad metodológica y su versatilidad. La optimización de hiperparámetros se realizó posteriormente en la Fase 4.

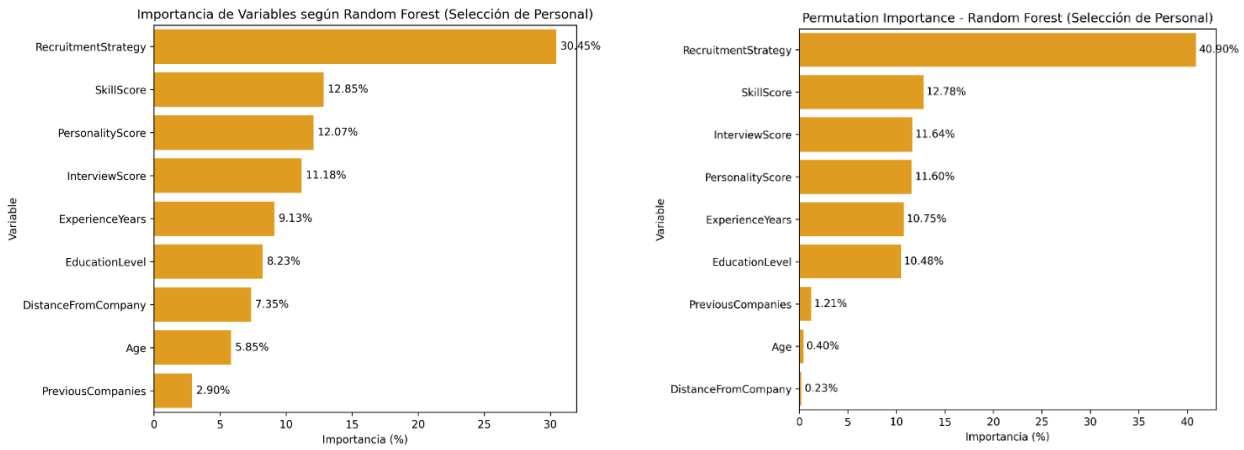
Fase 2: Recolección de datos

En la Fase 2 se procedió a la preparación inicial del conjunto de datos, organizando las variables para su uso en los modelos y recopilando información adicional mediante una encuesta aplicada a seis expertos en recursos humanos de distintos países. Esta encuesta permitió comparar la relevancia que los expertos asignan a cada variable con los resultados del modelo Random Forest, además de establecer bases para estudios futuros sobre la relación entre perfiles y áreas funcionales.

El formulario constó de dos preguntas enfocadas en identificar las variables más importantes para los departamentos de Sales, Human Resources y Research and Development, y en proponer nuevas variables no incluidas en el dataset. Finalmente, para determinar qué variables influían más en la predicción de candidatos aptos, se aplicó un análisis de importancia de características usando Random Forest y la función feature_importances_ de Scikit-learn.

Figura 3.

Importancia de variables en Random Forest: método tradicional (a) e importancia por permutación (b).



(a) **(b)**

La Figura 3(a) muestra que las variables más influyentes fueron RecruitmentStrategy (30.45 %), SkillScore (12.85 %) y PersonalityScore (12.07 %), indicando que la estrategia de reclutamiento, las habilidades técnicas y el desempeño en la entrevista son los factores que más aportan a la predicción. En contraste, Age (5.85 %) y PreviousCompanies (2.90 %) tuvieron un aporte mínimo. Estos resultados coinciden con la encuesta aplicada a los seis expertos en recursos humanos, quienes también destacaron como prioritarias las habilidades técnicas, el desempeño en entrevistas y la estrategia de reclutamiento, reforzando la validez del análisis.

Además, se aplicó Permutation Importance para confirmar la importancia obtenida con Random Forest. Esta técnica mide cómo cambia el rendimiento del modelo al permutar cada variable. Su uso está respaldado por la literatura en contextos de alta dimensionalidad (Mehdiyev et al., 2025) y en estudios de selección de personal donde suele coincidir con valoraciones humanas (Grunenberg et al., 2024). Los resultados ilustrados en la Figura 3(b) confirman que RecruitmentStrategy es la variable más influyente (40.90 %), seguida por SkillScore, InterviewScore y PersonalityScore, con valores similares. En cambio, Age y DistanceFromCompany aportaron muy poco. La coincidencia entre ambas técnicas confirma la solidez del análisis y señala qué atributos deben priorizarse en sistemas de selección de personal.

Fase 3: Validación y preparación de datos

Luego de verificar la calidad del dataset y el desbalance de clases, se aplicaron las transformaciones necesarias para preparar los datos. Las variables categóricas se codificaron con One-Hot Encoding (por ejemplo, Education Level y Recruitment Strategy) y las variables numéricas se normalizaron con Min-Max Scaling para llevarlas al rango 0–1, evitando que atributos como Interview Score o Distance From Company dominaran el modelo. El dataset se dividió en 70 % entrenamiento y 30 % prueba, usando muestreo estratificado para mantener la proporción de clases. Debido al

desbalance, se aplicó SMOTE solo al conjunto de entrenamiento para generar instancias sintéticas y mejorar métricas como Recall y F1-score.

Fase 4: Implementación de modelos

En esta fase se implementaron los tres algoritmos seleccionados: RF, SVM y DNN. Inicialmente, cada modelo se configuró con hiperparámetros basados en la revisión bibliográfica y en pruebas preliminares; sin embargo, estos parámetros se optimizaron posteriormente mediante RandomizedSearchCV y validación cruzada estratificada de 10 folds (K-Fold), valor utilizado por ser el estándar en la literatura y por ofrecer un buen equilibrio entre sesgo y varianza. Esta estrategia permitió obtener modelos más estables y con mejor rendimiento.

Para el modelo RF, el mejor resultado obtenido con RandomizedSearchCV incorporó 300 árboles (n_estimators=300), sin límite de profundidad (max_depth=None), min_samples_split=2, criterion='entropy' y random_state=42, configuraciones que permitieron mitigar el sobreajuste y mantener un desempeño consistente.

Para el modelo SVM, se utilizó un kernel de base radial RBF con C=100, gamma=0.01 y random_state=42, adecuado para modelar relaciones no lineales en tareas de clasificación binaria con variables mixtas (numéricas y categóricas).

En la DNN se empleó una arquitectura de tres capas ocultas con 64, 64 y 16 neuronas (activación ReLU), una capa salida con una única neurona y activación sigmoid, una tasa de aprendizaje de 0.001, Dropout del 40 %, 50 épocas, batch size de 32 y seed=42 en TensorFlow, NumPy y Python. Se aplicó EarlyStopping con paciencia de 10 épocas y el optimizador Adam con binary_crossentropy.

Todos los modelos se entrenaron y evaluaron dos veces: con datos originales y aplicando SMOTE en el conjunto de entrenamiento para analizar su impacto en métricas

como Accuracy, Recall y F1-score. Su usaron scripts independientes que siguieron la misma estructura de preprocesamiento y evaluación.

Se incorporaron mejoras metodológicas como validación cruzada estratificada y el cálculo de intervalos de confianza (IC) para Accuracy, Recall, F1-score y AUC, lo que permitió reforzar la fiabilidad y robustez estadística de los resultados.

Fase 5: Evaluación y comparación del rendimiento de los modelos

En la Fase 5 se evaluó el rendimiento de los modelos RF, SVM y DNN, tanto en su versión original como con SMOTE. Se aplicó un enfoque mixto que combinó validación cruzada estratificada (K-Fold) durante el ajuste y una evaluación final en un conjunto de prueba independiente, garantizando resultados estables y una medición objetiva frente a datos no vistos. El desempeño se cuantificó mediante Accuracy, Recall, F1-score, AUC y el Gap de generalización (Gap F1), cuyas fórmulas empleadas se expresan de la siguiente manera:

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

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 - score = 2 \times \frac{Precisión \times Recall}{Precisión + Recall} \quad (3)$$

$$AUC = \int_0^1 TPR\,(FPR)\,d(FPR) \quad (4)$$

$$Gap_{F1}(\%) = \left|\frac{F1_{train} - F1_{test}}{F1_{train}}\right| \times 100 \quad (5)$$

Donde TP (True Positives), TN (True Negatives), FP (False Positives) y FN (False Negatives) representan los valores obtenidos en la matriz de confusión, y $F1_{train}$ y $F1_{test}$ corresponden a los valores del F1-score calculados sobre los conjuntos de

entrenamiento y prueba, respectivamente. El comportamiento de cada algoritmo fue evaluado sobre el 30 % del conjunto de datos reservado para prueba, utilizando ocho métricas clave:

- Precisión (Accuracy): proporción de candidatos correctamente clasificados.
- Recall: porcentaje de candidatos idóneos que fueron identificados correctamente.
- F1-score: resume el equilibrio entre precisión y recall en una sola métrica.
- AUC-ROC y AUC-PR: áreas bajo las curvas ROC y Precision-Recall, empleadas para evaluar la capacidad discriminativa y el equilibrio entre sensibilidad y precisión, respectivamente.
- Tiempo de entrenamiento: duración total del proceso de ajuste del modelo sobre el conjunto de entrenamiento.
- Tiempo de predicción: tiempo requerido para generar las predicciones sobre el conjunto de prueba.
- Capacidad de generalización: diferencia entre el rendimiento en entrenamiento y prueba (Gap F1), expresada en porcentaje.
- Intervalos de confianza (IC): margen estadístico calculado para cada métrica mediante el método K-Fold, a fin de estimar la variabilidad del rendimiento y la robustez del modelo.

El proceso de evaluación se desarrolló en cinco pasos para cada modelo:

1. entrenamiento con el 70 % del dataset usando los hiperparámetros optimizados mediante RandomizedSearchCV.

2. predicción sobre el 30 % de datos no vistos.

3. cálculo de métricas con funciones de Scikit-learn y medición de tiempos.

4. estimación del Gap F1 y de los intervalos de confianza mediante validación cruzada estratificada.

5. registro y visualización de los resultados mediante tablas comparativas y gráficos ROC y Precision-Recall, para escenarios con y sin SMOTE.

Los resultados se organizaron en tablas y gráficas que se discuten en la sección de Resultados para identificar el modelo más adecuado.

Análisis de resultados

En esta sección se presentan los resultados del análisis comparativo de los algoritmos aplicados al proceso de selección de personal. Los experimentos se realizaron en un equipo con procesador Ryzen 5 3500U, 12 GB de RAM y Windows 11 Pro. Se evaluaron tres modelos de clasificación en dos escenarios: sin balanceo y aplicando SMOTE.

Análisis comparativo del rendimiento de los modelos

Evaluación de métricas tradicionales y computacionales

Los resultados (Tabla 2) muestran que RF sin SMOTE tuvo el mejor desempeño general (Accuracy = 0.91, F1 = 0.84, AUC = 0.94) y gestionó bien el desbalance sin sobremuestreo. Con SMOTE, RF mejoró la detección de la clase minoritaria, pero redujo su precisión y estabilidad. SVM fue el modelo más rápido y con bajo Gap, aunque con rendimiento ligeramente inferior a RF. DNN logró métricas competitivas, pero su alto costo computacional (≈600 s de entrenamiento) limita su utilidad práctica.

Tabla 2. *Resultados comparativos de los modelos sin y con SMOTE.*

Modelo	Accuracy	Recall	F1-score	AUC	Tiempo Entrenamiento(s)	Tiempo Prueba(s)	Gap F1 %
RF	0.91	0.82	0.84	0.94	44.57	0.83	15.74
RF + SMOTE	0.90	0.77	0.82	0.93	65.11	0.06	18.10
SVM	0.90	0.80	0.83	0.93	4.12	0.05	3.20
SVM + SMOTE	0.88	0.77	0.79	0.92	8.93	0.06	20.14
DNN	0.90	0.78	0.83	0.92	624.16	0.30	0.41
DNN + SMOTE	0.89	0.78	0.81	0.92	760.46	0.19	15.59

Validación estadística mediante K-Fold

Para evaluar la consistencia estadística de las métricas de desempeño, se aplicó validación cruzada estratificada (K-Fold con cinco divisiones). Los resultados, junto con los intervalos de confianza al 95 %, se presentan en la Tabla 3, donde se evidencia la estabilidad de los valores de Accuracy, Recall, F1-score y AUC para los tres algoritmos principales.

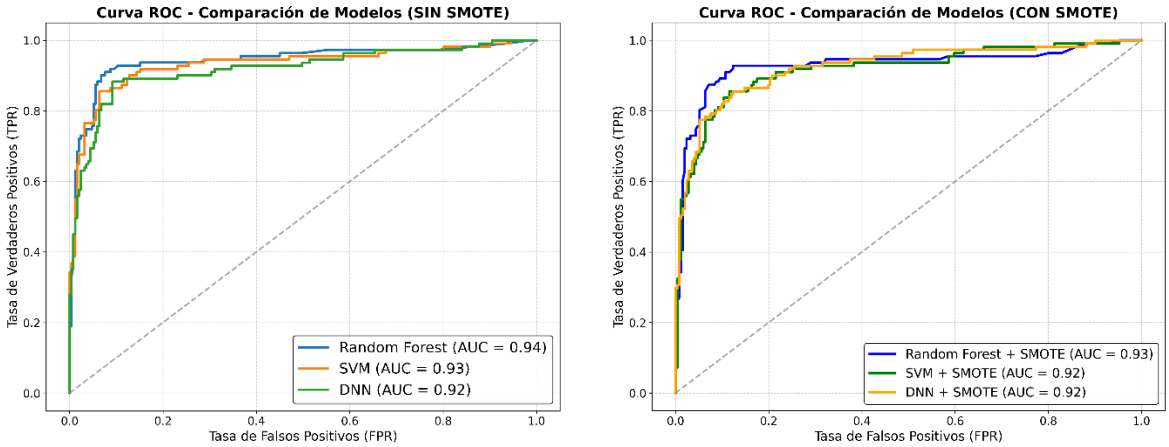
Tabla 3. *Resultados de validación cruzada (IC95 %) para los modelos*

Métrica	CV RF (IC95%)	CV SVM (IC95%)	CV DNN (IC95%)
Accuracy	0.86 (0.86 – 0.91)	0.86 (0.83 – 0.90)	0.87 (0.84 – 0.94)
Recall	0.67 (0.60 – 0.74)	0.73 (0.66 – 0.80)	0.75 (0.68 – 0.82)
F1-score	0.78 (0.73 – 0.83)	0.77 (0.71 – 0.82)	0.78 (0.73 – 0.84)
AUC	0.92 (0.89 – 0.96)	0.92 (0.87 – 0.95)	0.91 (0.85 – 0.94)

Análisis gráfico: ROC, PR y Matriz de Confusión

Las curvas ROC (Figura 4) confirman la superioridad del modelo RF sin SMOTE, que alcanzó un AUC de 0.94, seguido por SVM (0.93) y DNN (0.92). Tras aplicar SMOTE, los tres modelos conservaron un rendimiento destacado (AUC > 0.91), reflejando su capacidad de discriminación con datos balanceados.

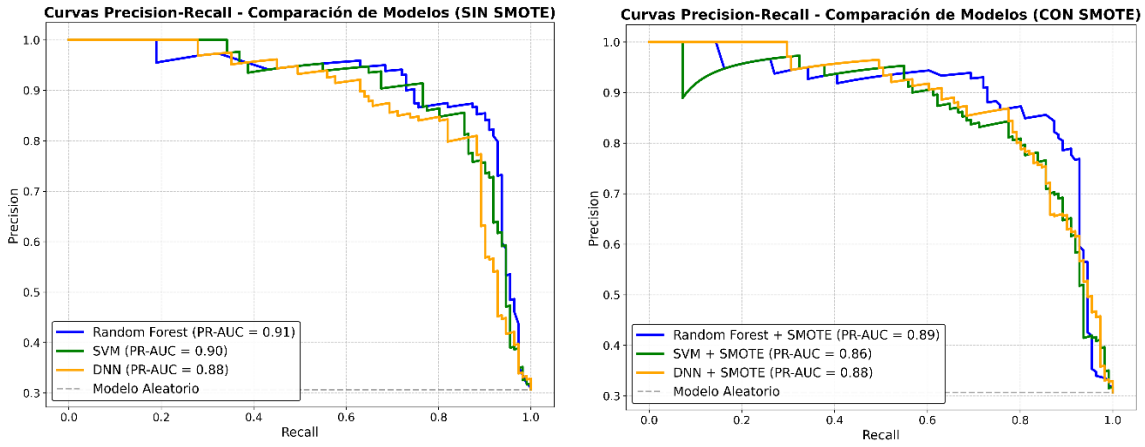
Figura 4. *Curvas ROC comparativas: (a) datos originales; (b) con SMOTE. RF muestra el AUC más alto en ambos casos demostrando el mejor desempeño discriminativo.*



(a) **(b)**

Las curvas Precision-Recall (Figura 5) muestran un comportamiento coherente con las métricas previas. RF sin SMOTE alcanzó el PR-AUC más alto (0.91), seguido por SVM (0.90) y DNN (0.88). Tras aplicar SMOTE, el modelo RF + SMOTE mantuvo el mejor desempeño (PR-AUC = 0.89), lo que confirma que el balanceo de clases mejoró la detección de candidatos aptos sin afectar de forma notable la precisión.

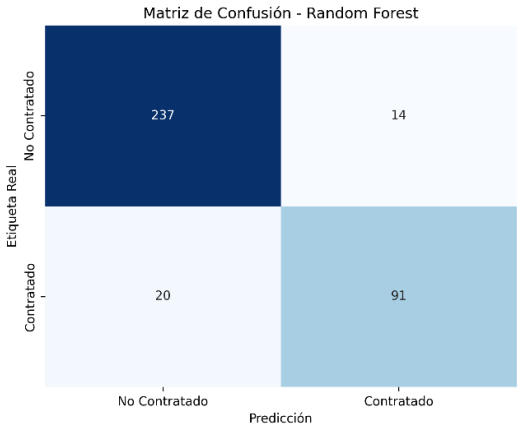
Figura 5. *Curvas Precision-Recall: (a) sin SMOTE; (b) con SMOTE. RF presenta el mayor AUC-PR en ambos escenarios.*



(a) **(b)**

Además de las métricas generales, se analizó la matriz de confusión del modelo RF sin SMOTE, ya que obtuvo el mejor desempeño global. La Figura 6 muestra una alta tasa de aciertos en ambas clases: de los 362 registros, el modelo clasificó correctamente a 237 “No Contratados” y 91 “Contratados”, cometiendo solo 34 errores entre falsos positivos y falsos negativos. Estos resultados coinciden con su precisión del 91 % y su recall del 82 % (Tabla 2), confirmando que el modelo mantiene un buen equilibrio entre exactitud y capacidad de detección.

Figura 6. *Matriz de confusión – RF (mejor modelo).*



En cuanto a eficiencia computacional, SVM fue el modelo más rápido, mientras que la DNN requirió mucho más tiempo debido a su complejidad, tal como indica la literatura sobre redes profundas. Además, los resultados coinciden con Haque et al. (2025) y Gómez et al. (2022), quienes señalan que SMOTE mejora métricas como Recall y F1-score, especialmente en modelos basados en árboles, aunque su efecto depende del dataset y del nivel de desbalance.

Conclusión de la fase

En términos generales, el RF sin SMOTE fue el modelo con mejor desempeño global, por su alta precisión, buena discriminación y estabilidad. El RF con SMOTE mejoró la detección de la clase minoritaria, siendo más adecuado en escenarios con fuerte desbalance. El SVM destacó por su rapidez y consistencia, mientras que la DNN logró un rendimiento competitivo, aunque con alto costo computacional.

En síntesis, el RF sin SMOTE se posiciona como la opción más sólida y confiable para apoyar procesos de selección de personal.

Conclusiones

La integración de IA en la selección de personal ofrece ventajas claras frente a los métodos tradicionales: permite procesar grandes volúmenes de datos, reduce sesgos humanos y estandariza los criterios de evaluación. Los modelos de aprendizaje automático identifican patrones que apoyan el juicio profesional de los reclutadores, mejorando la eficiencia y la calidad de las decisiones.

Este estudio desarrolló un modelo predictivo comparando tres técnicas de clasificación (RF, SVM y DNN) en escenarios con y sin SMOTE. El RF sin SMOTE obtuvo el mejor rendimiento global, con un Accuracy de 91 %, Recall de 82 %, F1-score de 84 % y AUC de 0.94. Aunque SMOTE mejoró la detección de la clase minoritaria, redujo ligeramente la precisión total.

El SVM destacó por su rapidez y estabilidad, con tiempos de entrenamiento y prueba muy bajos (4.12 s y 0.05 s) y un gap de generalización de 3.20 %. La DNN mostró métricas competitivas, pero con un costo computacional alto (624.16 s), lo que limita su uso en aplicaciones que requieren eficiencia.

La validación cruzada y los intervalos de confianza confirmaron la estabilidad de los resultados. Además, el análisis de Permutation Importance identificó como variables más influyentes a Recruitment Strategy, Skill Score, Interview Score y Personality Score, en concordancia con estudios recientes y con la opinión de los expertos consultados.

En términos aplicados, los hallazgos muestran que los modelos de aprendizaje automático pueden mejorar significativamente los procesos de selección al reducir sesgos y aumentar la precisión en la identificación de candidatos aptos. En este

contexto, RF se posiciona como la opción más robusta y equilibrada, mientras que SVM es ideal cuando se prioriza la velocidad.

Para trabajos futuros se recomienda explorar arquitecturas de aprendizaje profundo más avanzadas e incorporar nuevas variables psicométricas, de desempeño y organizacionales, con el fin de fortalecer la capacidad predictiva de los sistemas de selección inteligente.

Referencias bibliográficas

Alameda Castillo, M. T. (2021). Reclutamiento tecnológico: Sobre algoritmos y acceso al empleo. Temas Laborales, 159, 11–52. https://dialnet.unirioja.es/servlet/articulo?codigo=8250088

Barragán-Martínez, X. (2023). Situación de la inteligencia artificial en el Ecuador en relación con los países líderes de la región del Cono Sur. FIGEMPA: Investigación y Desarrollo, 16(2), 23–38. https://doi.org/10.29166/revfig.v16i2.4498

El Kharoua, R. (2024). Predicting hiring decisions in recruitment data [Dataset]. Kaggle. https://doi.org/10.34740/kaggle/dsv/8715385

Garzón Ayala, A. S., Cedeño Torres, J. M., & León Varela, B. F. (2024). Machine learning approaches to enhance candidate selection: A comparative study in HR recruitment. Dom. Cien., 10(3), 160–176. https://doi.org/10.22214/ijraset.2024.63745

Gómez, J. A., García, J. C., & Herrera, F. (2022). An empirical study on the impact of SMOTE for imbalanced classification. Applied Sciences, 12(6424), 1–23.

González-Sendino, R., Serrano, E., & Bajo, J. (2024). Mitigating bias in artificial intelligence: Fair data generation via causal models for transparent and explainable decision-making. Future Generation Computer Systems, 155, 384–401. https://doi.org/10.1016/j.future.2024.02.023

Grunenberg, E., Peters, H., Francis, M. J., Back, M. D., & Matz, S. C. (2024). Machine learning in recruiting: Predicting personality from CVs and short text responses. Frontiers in Social Psychology, 1, Article 1290295. https://doi.org/10.3389/frsps.2023.1290295

Haque, M., Paralkar, T. A., Rajguru, S., Goyal, A. A., Patil, T., & Upreti, K. (2025). Featuring machine learning models to evaluate employee attrition: A comparative analysis of workforce stability-relating factors. International Research Journal of Multidisciplinary Scope, 6(2), 862–873. https://doi.org/10.47857/irjms.2025.v06i02.03512

León-Varela, B. F., Arroyo-Carrillo, L. A., Vargas-Monteleage, A. R., & Reigosa-Lara, A. (2024). Inteligencia artificial para los procesos de gestión del talento humano. Dom. Cien., 10(4), 182–203. https://doi.org/10.23857/dc.v10i4.4057

Medrano, L. A., Velázquez, M. T., & Serrano, A. M. (2022). Evaluación de algoritmos de IA en América Latina. En Congreso Regional de Ciencias Computacionales (pp. 56–72). Bogotá, Colombia.

Mehdiyev, N., Majlatow, M., & Fettke, P. (2025). Integrating permutation feature importance with conformal prediction for robust explainable artificial intelligence in predictive process monitoring. Engineering Applications of Artificial Intelligence, 149, 110363. https://doi.org/10.1016/j.engappai.2025.110363

Paredes Morales, C. (2023). Optimización de procesos de selección en Ecuador mediante IA. Estudios Empresariales del Ecuador, 5, 34–50.

Pérez López, J. I. (2023). Inteligencia artificial y contratación laboral. Revista de Estudios Jurídicos y Sociales, 7, 186–205. https://doi.org/10.24310/rejlss7202317557

Quintanilla-Medina, M. C., & Coral-Ignacio, M. A. (2024). Personnel selection system based on the selection algorithm. Revista DYNA, 91(231), 105–111. https://doi.org/10.15446/dyna.v91n231.110412

Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting Employee Attrition Using Machine Learning Approaches. Applied Sciences, 12(13), 6424. https://doi.org/10.3390/app12136424

Singh, N., & Chakraborty, S. (2024). Machine learning approaches to enhance candidate selection: A comparative study in HR recruitment. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 12(7), 1286–1296. https://doi.org/10.22214/ijraset.2024.63745.