Estudios y Perspectivas Revista Científica y Académica
ISSN en línea 3005-2599, abril-junio 2026,
Volumen 6, Número 2.

DOI: https://doi.org/10.61384/r.c.a.v6i2

DESARROLLO DE UN MODELO PREDICTIVO BASADO EN INTELIGENCIA ARTIFICIAL PARA LA GESTIÓN DE PROYECTOS EN LAS PYMES

DEVELOPMENT OF AN ARTIFICIAL INTELLIGENCE-BASED PREDICTIVE MODEL FOR PROJECT MANAGEMENT IN SMES

Angel Leonel Condoy Ramirez
Universidad Técnica de Machala, Ecuador

E. COUTURE
Educación y Ciencias Sociales

Estudios y Perspectivas
Revista Científica y Académica

Estudios y Perspectivas
Revista Científica Multidisciplinaria

DOI: https://doi.org/10.61384/r.c.a..v6i2.1977

Desarrollo de un Modelo Predictivo Basado en Inteligencia Artificial para la Gestión de Proyectos en las Pymes

Angel Leonel Condoy Ramirez
angelcondoy19@gmail.com
https://orcid.org/0009-0001-6003-3758
Universidad Técnica de Machala, Ecuador

RESUMEN

Las plataformas de financiamiento colectivo, como Kickstarter, se han consolidado como una alternativa de financiación para emprendedores, iniciativas creativas y startups que buscan desarrollar proyectos sin recurrir a mecanismos tradicionales. No obstante, una proporción importante de las campañas no alcanza su meta de recaudación, lo que afecta su viabilidad y sostenibilidad. La presente investigación tuvo como objetivo desarrollar un modelo predictivo basado en técnicas de aprendizaje automático para estimar la probabilidad de éxito o fracaso de campañas de crowdfunding antes de su lanzamiento. El estudio se fundamentó en el análisis de un conjunto de datos históricos públicos de Kickstarter, integrado por más de 300.000 proyectos clasificados como exitosos o fallidos. Se adoptó un enfoque cuantitativo y aplicado, utilizando algoritmos de clasificación supervisada como Árboles de Decisión, Random Forest y Redes Neuronales Artificiales. El desempeño de los modelos se evaluó mediante validación cruzada y métricas de clasificación, entre ellas exactitud, precisión, sensibilidad, puntuación F1 y el área bajo la curva ROC. Los resultados evidenciaron que Random Forest presentó el mejor desempeño predictivo en escenarios ex ante, confirmando la utilidad del aprendizaje automático como herramienta de apoyo para evaluar la viabilidad de proyectos antes del inicio de una campaña de financiamiento colectivo

***Palabras clave*:** crowdfunding, aprendizaje automático, modelos predictivos, Kickstarter, éxito de proyectos

Development of an Artificial Intelligence-Based Predictive Model for Project Management in Smes

ABSTRACT

Crowdfunding platforms, such as Kickstarter, have become an important alternative source of financing for entrepreneurs, creative initiatives, and startups seeking to develop projects without relying on traditional funding mechanisms. However, a significant proportion of campaigns fail to reach their fundraising goals, compromising their viability and long-term sustainability. The objective of this study was to develop a predictive model based on machine learning techniques to estimate the probability of success or failure of crowdfunding campaigns before their launch. The research was based on the analysis of a public historical Kickstarter dataset comprising more than 300,000 projects classified as either successful or failed. A quantitative and applied research approach was adopted, employing supervised classification algorithms, including Decision Trees, Random Forest, and Artificial Neural Networks. Model performance was evaluated using cross-validation and standard classification metrics, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The results showed that the Random Forest algorithm achieved the best predictive performance in ex ante scenarios, confirming the usefulness of machine learning as a decision-support tool for assessing project feasibility before launching a crowdfunding campaign

Keywords**:** crowdfunding*,* machine learning*,* predictive models, Kickstarter, project success

Recibido: 11- abril-2026 / Aprobado: 30-junio-2026

INTRODUCCIÓN

En un entorno cada vez más competitivo e innovador, los emprendedores, las startups y las pequeñas iniciativas creativas enfrentan importantes dificultades para acceder al financiamiento. Esta situación ha impulsado el crecimiento de plataformas de financiamiento colectivo (crowdfunding), como Kickstarter, que permiten obtener recursos directamente de los patrocinadores sin recurrir a intermediarios financieros. Sin embargo, una proporción significativa de las campañas no alcanza su meta de recaudación, limitando la ejecución de los proyectos.

Ante este escenario, surge la necesidad de desarrollar herramientas que permitan estimar la probabilidad de éxito de una campaña antes de su lanzamiento. En este contexto, la inteligencia artificial (IA), particularmente el aprendizaje automático (machine learning), ha demostrado un importante potencial para optimizar la toma de decisiones en escenarios de incertidumbre. Carrillo Zenteno et al. (2024) destacan que la IA mejora los procesos de planificación y fortalece la gestión de proyectos, capacidad especialmente relevante en el ámbito del crowdfunding.

Diversos estudios respaldan la aplicación del aprendizaje automático en la predicción de resultados. Pedroso (2017) desarrolló una plataforma basada en técnicas de machine learning para identificar tempranamente riesgos en proyectos, mientras que investigaciones recientes evidencian que algoritmos como Random Forest y las Redes Neuronales Artificiales superan, en numerosos casos, a los métodos estadísticos tradicionales para predecir el éxito de campañas de crowdfunding (Elitzur et al., 2024; Putri Raflesia et al., 2023).

Aunque Kickstarter no constituye un sistema formal de gestión de proyectos, varios de sus componentes pueden analizarse desde esta perspectiva. Ilieș et al. (2010) sostienen que la aplicación de buenas prácticas de dirección incrementa la probabilidad de alcanzar los objetivos propuestos. En consecuencia, variables como la meta financiera, la duración de la campaña y la identificación del público objetivo pueden interpretarse como factores clave para la planificación y el éxito de los proyectos.

Con base en estos antecedentes, la presente investigación propone desarrollar un modelo predictivo basado en algoritmos de aprendizaje automático, entrenado con un conjunto de datos históricos públicos de más de 300.000 campañas de Kickstarter. A diferencia de los enfoques descriptivos, el estudio se orienta a estimar la probabilidad de éxito de una campaña antes de su lanzamiento, utilizando variables estructurales como la meta financiera (goal), la categoría principal (main category), la duración de la campaña (duration) y el país de origen (country).

La investigación adopta un enfoque cuantitativo, con un diseño no experimental y de corte transversal. Para el análisis se emplean algoritmos de clasificación supervisada, entre ellos Árboles de Decisión, Random Forest y Redes Neuronales Artificiales. El desempeño de los modelos se evalúa mediante validación cruzada y métricas estándar de clasificación, como exactitud, precisión, sensibilidad, puntuación F1 y el área bajo la curva ROC (AUC), con el propósito de desarrollar una herramienta de apoyo para la toma de decisiones que contribuya a reducir el riesgo antes del lanzamiento de campañas de crowdfunding.

Trabajos Relacionados Y Fundamentos Teóricos

Las pequeñas y medianas empresas (PYMES) y los emprendimientos independientes desempeñan un papel fundamental en la generación de empleo y la innovación, aunque muchas enfrentan dificultades para acceder al financiamiento tradicional. En este contexto, el crowdfunding se ha consolidado como una alternativa para impulsar nuevos proyectos mediante plataformas como Kickstarter.

La ausencia de herramientas que permitan estimar previamente la probabilidad de éxito de una campaña incrementa la incertidumbre en la toma de decisiones. Frente a ello, la inteligencia artificial y el aprendizaje automático representan una alternativa para anticipar resultados y optimizar recursos. Burggräf et al. (2024) señalan que la incorporación de soluciones basadas en IA mejora la eficiencia y la capacidad de respuesta de las PYMES, aun cuando disponen de recursos limitados.

El aprendizaje automático permite identificar patrones en grandes volúmenes de datos y generar predicciones útiles para la gestión de proyectos y el emprendimiento. En campañas de crowdfunding, estas técnicas facilitan la estimación de la probabilidad de éxito mediante variables como la categoría

del proyecto, la meta financiera, la duración y el país de origen. Ugbebor et al. (2024) destacan que los modelos predictivos pueden implementarse eficazmente en organizaciones pequeñas siempre que exista una base histórica de datos adecuada.

Desde la perspectiva de la gestión de proyectos, la aplicación de buenas prácticas incrementa las probabilidades de éxito. Hussain et al. (2022) sostienen que estas prácticas mejoran significativamente el desempeño de los proyectos, mientras que Provost y Fawcett (2013) señalan que el análisis sistemático de datos históricos constituye una base sólida para generar modelos predictivos que apoyen la toma de decisiones.

Para la construcción del modelo se emplearon algoritmos supervisados de clasificación binaria, entre ellos Árboles de Decisión, Random Forest y Redes Neuronales Artificiales. Los Árboles de Decisión destacan por su interpretabilidad; Random Forest mejora la precisión mediante métodos de ensamblado, y las Redes Neuronales permiten modelar relaciones complejas entre variables. Yoo et al. (2023) demostraron la eficacia de estos algoritmos para predecir el desempeño de proyectos de I+D en PYMES.

Finalmente, la evaluación de los modelos predictivos se realizó mediante métricas estándar como exactitud, precisión, recall, puntuación F1 y el área bajo la curva ROC (AUC), complementadas con validación cruzada estratificada. Peral (2015) destaca que la combinación de estas métricas con técnicas de validación constituye una práctica esencial para seleccionar modelos predictivos confiables.

METODOLOGÍA

Enfoque metodológico

La presente investigación adopta un enfoque cuantitativo basado en el análisis de datos históricos provenientes de campañas publicadas en la plataforma Kickstarter. Su propósito es identificar patrones relevantes y desarrollar modelos predictivos capaces de estimar la probabilidad de éxito o fracaso de una campaña utilizando únicamente variables disponibles antes de su lanzamiento (James et al., 2013).

El estudio presenta un diseño no experimental y de corte transversal, ya que analiza datos observacionales sin manipular las variables de estudio. Para ello se emplearon técnicas de aprendizaje automático (machine learning) mediante algoritmos supervisados de clasificación binaria, específicamente Árboles de Decisión, Random Forest y Redes Neuronales Artificiales (ANN), seleccionados por su capacidad para modelar relaciones complejas entre variables numéricas y categóricas (Jordan & Mitchell, 2015; Bang et al., 2022).

Diseño y estructura del estudio

La investigación se desarrolló en dos fases complementarias. La primera correspondió al análisis exploratorio-descriptivo, orientado a caracterizar el comportamiento histórico de las campañas mediante variables como el país de origen, la categoría del proyecto, la meta financiera, el monto recaudado y la duración de la campaña, siguiendo los principios del análisis exploratorio de datos (Seltman, 2018).

La segunda fase consistió en el desarrollo de modelos predictivos para estimar el éxito de una campaña antes de su lanzamiento. Con el fin de evitar la fuga de información (data leakage), únicamente se utilizaron variables disponibles en la etapa ex ante, siguiendo las recomendaciones de buenas prácticas para el desarrollo de modelos predictivos (Zinkevich, 2015).

Tipo y nivel de investigación

La investigación es de carácter aplicado, ya que propone una solución para apoyar la evaluación anticipada de campañas de financiamiento colectivo mediante técnicas de aprendizaje automático (Joakin, 2021).

Asimismo, presenta un alcance exploratorio-descriptivo, debido a que analiza el comportamiento histórico de campañas de crowdfunding e identifica patrones relevantes asociados a variables como la meta financiera, la categoría del proyecto, la duración, el país de origen y el número de patrocinadores, constituyendo la base para el desarrollo del modelo predictivo.

Fuente de datos

Se utilizó el conjunto de datos público Kickstarter Projects, disponible en Kaggle, integrado por más de 300.000 campañas publicadas hasta 2018 (Kaggle, 2018).

Cada registro incluye variables como la meta financiera (goal), el monto recaudado (usd_pledged_real), el número de patrocinadores (backers), la categoría del proyecto, el país de origen, las fechas de inicio y cierre de la campaña y el estado final (successful o failed). Las variables usd_pledged_real y backers fueron empleadas únicamente en el análisis descriptivo y en el modelo de referencia (baseline), pero no como variables predictoras del modelo ex ante.

Preprocesamiento de los datos

El procesamiento se realizó en Python utilizando las bibliotecas pandas, NumPy y scikit-learn. El preprocesamiento incluyó la depuración de registros inconsistentes, la codificación de variables categóricas mediante Label Encoding y One-Hot Encoding, la normalización de variables numéricas y la generación de variables derivadas, como la duración de la campaña.

Posteriormente, el conjunto de datos se dividió en subconjuntos de entrenamiento (80 %) y validación (20 %). Géron (2019) señala que esta etapa resulta fundamental para garantizar un adecuado desempeño de los modelos predictivos.

Selección y entrenamiento de los modelos

Se entrenaron modelos supervisados utilizando Árboles de Decisión, Random Forest y Redes Neuronales Artificiales. Cada algoritmo fue evaluado mediante partición entrenamiento-prueba y validación cruzada (k-fold cross-validation), empleando métricas como exactitud (accuracy), precisión (precision), sensibilidad (recall), puntuación F1 y el área bajo la curva ROC (AUC). Raschka y Mirjalili (2019) destacan que una adecuada selección del algoritmo y un proceso riguroso de evaluación son determinantes para obtener modelos confiables.

Evaluación de los modelos

El desempeño de los modelos se evaluó mediante métricas estándar para problemas de clasificación binaria, incluyendo exactitud, precisión, sensibilidad, puntuación F1 y AUC.

Además, se aplicó validación cruzada estratificada (Stratified K-Fold Cross-Validation), técnica que contribuye a reducir el sobreajuste y mejorar la capacidad de generalización del modelo (Géron, 2019).

Desarrollo del prototipo

Como aplicación práctica se desarrolló un prototipo funcional en Python, capaz de procesar archivos en formato Excel o CSV y estimar automáticamente la probabilidad de éxito o fracaso de nuevos proyectos utilizando únicamente variables ex ante.

La interfaz gráfica fue implementada mediante Tkinter y la ejecución de los modelos mediante scikit-learn, permitiendo disponer de una herramienta sencilla y accesible para apoyar la toma de decisiones. Moore (2021) destaca que Tkinter constituye una alternativa eficiente para el desarrollo de interfaces gráficas en aplicaciones de análisis de datos.

RESULTADOS Y DISCUSIÓN

A. Preparación de datos

Para esta investigación se utilizó un conjunto de datos histórico de campañas de Kickstarter, disponible en Kaggle (ks-projects-201801.csv). El conjunto de datos fue procesado mediante la biblioteca pandas de Python, realizando una depuración inicial que eliminó registros incompletos y campañas con estados diferentes de successful o failed.

A partir de las variables launched y deadline se calculó la duración de cada campaña en días y la variable objetivo (state) fue codificada de forma binaria, asignando 1 a las campañas exitosas y 0 a las fallidas. La preparación de los datos se estructuró en dos fases: análisis exploratorio-descriptivo y desarrollo del modelo predictivo ex ante.

B. Preparación de datos para la fase exploratoria-descriptiva

La fase exploratoria-descriptiva utilizó variables predictoras y de resultado para analizar el comportamiento histórico de las campañas e identificar patrones relevantes. Se consideraron la meta financiera (goal), el monto recaudado (usd_pledged_real), el número de patrocinadores (backers), la categoría principal (main_category), el país de origen (country) y la duración de la campaña (duration).

Estas variables permitieron analizar la tasa de éxito por país y categoría, la relación entre la meta financiera y el monto recaudado, así como el comportamiento general de las campañas publicadas en Kickstarter.

C. Preparación de datos para la fase predictiva (modelo ex ante)

Para el desarrollo del modelo predictivo se emplearon únicamente variables disponibles antes del lanzamiento de la campaña (variables ex ante), evitando la fuga de información (data leakage) y garantizando la aplicabilidad del modelo en escenarios reales.

Como variables predictoras se utilizaron la meta financiera (goal), la categoría principal (main_category), el país de origen (country) y la duración planificada de la campaña (duration). Las variables usd_pledged_real y backers fueron excluidas por corresponder a información disponible durante o al finalizar la campaña.

Las variables categóricas fueron codificadas mediante Label Encoding, mientras que las numéricas se estandarizaron con StandardScaler. Finalmente, el conjunto de datos se dividió en subconjuntos de entrenamiento (80 %) y prueba (20 %) mediante train_test_split de la biblioteca scikit-learn.

D. Análisis exploratorio-descriptivo

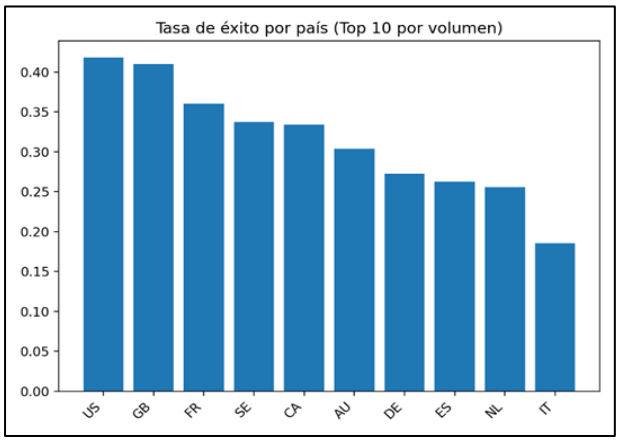
1) Tasa de éxito por país

La Figura 1 presenta la tasa de éxito de las campañas de crowdfunding según el país de origen, considerando los diez países con mayor número de proyectos publicados en Kickstarter. Los resultados evidencian diferencias significativas en el desempeño de las campañas según el contexto geográfico.

Estados Unidos (US) y el Reino Unido (GB) registran las mayores tasas de éxito, cercanas al 42 %, seguidos por Francia (FR), Suecia (SE) y Canadá (CA), con porcentajes entre el 33 % y el 36 %. En contraste, España (ES), Países Bajos (NL) e Italia (IT) presentan los menores niveles de éxito, siendo este último el país con la proporción más baja de campañas exitosas.

Estos resultados sugieren que el contexto geográfico constituye un factor asociado a la probabilidad de éxito de las campañas de crowdfunding.

Figura 1. Tasa de éxito de campañas de *crowdfunding* por país.



Fuente: Elaboración propia

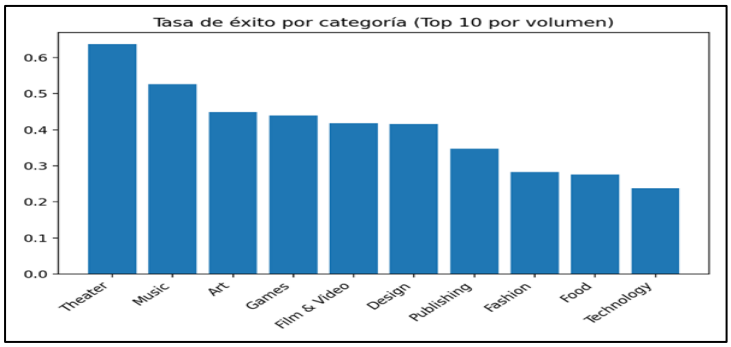
2) Tasa de éxito por categoría del proyecto

La Figura 2 presenta la tasa de éxito de las campañas según la categoría principal del proyecto, considerando las diez categorías con mayor número de campañas registradas en Kickstarter. Los resultados muestran diferencias significativas en el desempeño según el tipo de proyecto.

Las categorías Theater y Music registran las mayores tasas de éxito, superiores al 50 %, seguidas por Art, Games, Film & Video y Design, que también presentan porcentajes favorables. En contraste, Fashion, Food y Technology muestran las tasas de éxito más bajas.

Estos resultados indican que la categoría del proyecto constituye un factor asociado a la probabilidad de éxito de las campañas de crowdfunding.

Figura 2. Tasa de éxito de campañas por categoría del proyecto.



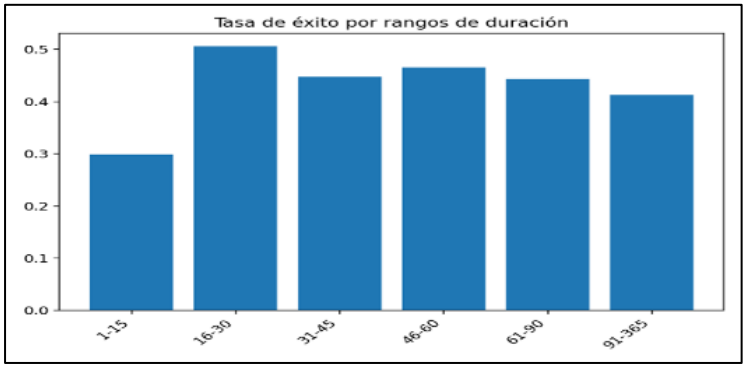
Fuente: Elaboración propia

3) Tasa de éxito según la duración de la campaña

La Figura 3 presenta la relación entre la duración de la campaña y su probabilidad de éxito. Los resultados muestran que las campañas con una duración de 16 a 30 días registran las mayores tasas

de éxito, mientras que las de muy corta duración presentan porcentajes considerablemente inferiores. Estos hallazgos evidencian que la duración de la campaña constituye un factor asociado a su probabilidad de éxito.

Figura 3. Tasa de éxito de campañas por rangos de duración.

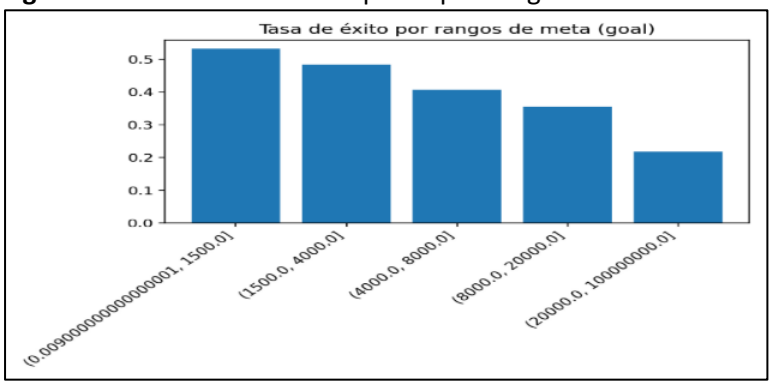


Fuente: Elaboración propia

4) Tasa de éxito según la meta financiera

La **Figura 4** presenta la tasa de éxito de las campañas según diferentes rangos de meta financiera (*goal*). Los resultados muestran una relación inversa entre el monto solicitado y la probabilidad de éxito: las campañas con metas financieras más bajas registran mayores tasas de éxito, mientras que aquellas con objetivos más elevados presentan porcentajes significativamente inferiores.

Figura 4. Tasa de éxito de campañas por rangos de meta financiera.



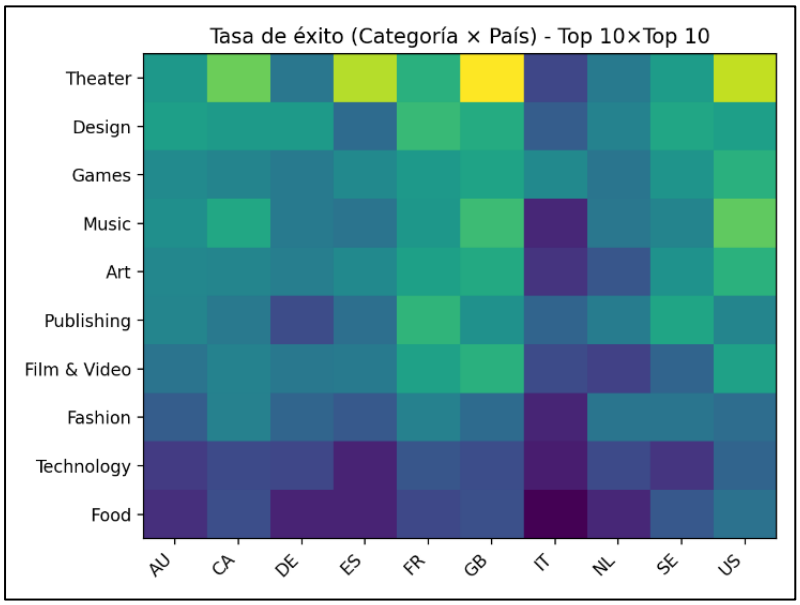
Fuente: Elaboración propia

5) Relación entre la categoría del proyecto y el país de origen

La **Figura 5** presenta un mapa de calor que muestra la tasa de éxito según la combinación de la categoría del proyecto y el país de origen. Los resultados evidencian que el desempeño de una misma categoría varía entre países, indicando que la probabilidad de éxito de una campaña de *crowdfunding*

depende de la interacción entre la categoría del proyecto, el contexto geográfico, la duración de la campaña y la meta financiera.

Figura 5. Relación entre la categoría del proyecto y el país de origen.



Fuente: Elaboración propia

E. Síntesis del análisis exploratorio

El análisis exploratorio-descriptivo permitió identificar patrones asociados al éxito de las campañas de crowdfunding, particularmente en variables como la categoría del proyecto, el país de origen, la duración de la campaña y la meta financiera. Estos resultados sustentan la selección de las variables empleadas en el modelo predictivo ex ante y facilitan la interpretación de los resultados posteriores.

F. Modelo de referencia (*baseline*)

Como punto de comparación, se entrenaron modelos de clasificación utilizando todas las variables disponibles en el conjunto de datos, incluidas aquellas observadas durante o al finalizar la campaña, como *usd_pledged_real* y *backers*. Este modelo de referencia (*baseline*) se utilizó únicamente con fines exploratorios y comparativos, ya que dichas variables no están disponibles antes del lanzamiento de una campaña y, por tanto, no representan un escenario predictivo real.

Modelo Árbol de Decisión

El **Árbol de Decisión** es un algoritmo de clasificación supervisada basado en reglas jerárquicas que facilita la interpretación del proceso de decisión. Como parte del modelo *baseline*, fue entrenado utilizando la totalidad de las variables disponibles en el conjunto de datos.

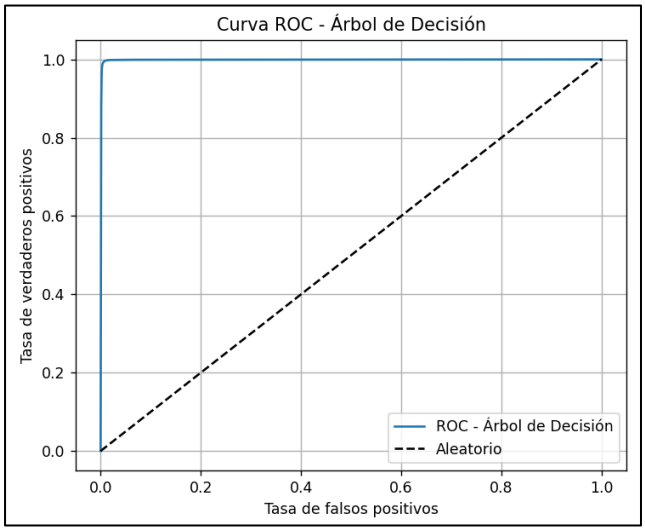
Tabla 1. Métricas de desempeño del modelo Árbol de Decisión (baseline)

Métrica	Valor
Accuracy	0.9938
Precision	0.9910
Recall	0.9936
F1 Score	0.9923
AUC	0.9987

Fuente: Elaboración propia

La **Figura 6** presenta la curva ROC correspondiente al modelo. El área bajo la curva (AUC) alcanza un valor cercano a **1**, lo que evidencia una excelente capacidad discriminativa. No obstante, este desempeño debe interpretarse con cautela, ya que se obtuvo utilizando variables observadas durante o después de la campaña, las cuales no se encuentran disponibles en un escenario predictivo real.

Figura 6. Curva ROC del modelo Árbol de Decisión (*baseline*).



Fuente: Elaboración propia

Modelo Random Forest

Random Forest es un método de aprendizaje por ensamblado (ensemble learning) que combina múltiples árboles de decisión para mejorar la capacidad predictiva y reducir el riesgo de sobreajuste mediante un mecanismo de votación mayoritaria. El modelo fue entrenado utilizando el mismo conjunto de variables empleado en el Árbol de Decisión.

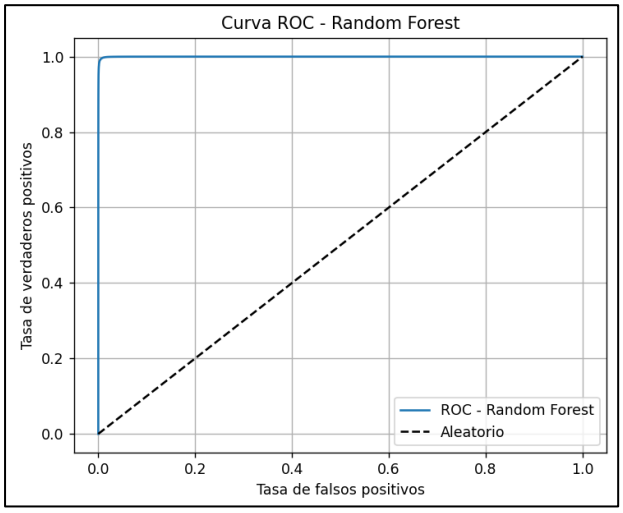
Tabla 2. Métricas de desempeño del modelo Random Forest (baseline)

Métrica	Valor
Accuracy	0.9943
Precision	0.9917
Recall	0.9941
F1 Score	0.9929
AUC	0.9998

Fuente: Elaboración propia

La Figura 7 muestra la curva ROC del modelo Random Forest, la cual evidencia un desempeño ligeramente superior al obtenido por el Árbol de Decisión individual. Esta mejora se consigue a costa de un mayor tiempo de entrenamiento, aunque proporciona un incremento en la estabilidad y capacidad de generalización del modelo.

Figura 7. Curva ROC del modelo Random Forest (baseline).



Fuente: Elaboración propia

Modelo Red Neuronal Artificial

Se implementó una Red Neuronal Artificial del tipo Perceptrón Multicapa (MLP) mediante la biblioteca scikit-learn, configurada con dos capas ocultas. Al igual que los modelos anteriores, fue entrenada utilizando la totalidad de las variables disponibles como parte del modelo de referencia (baseline).

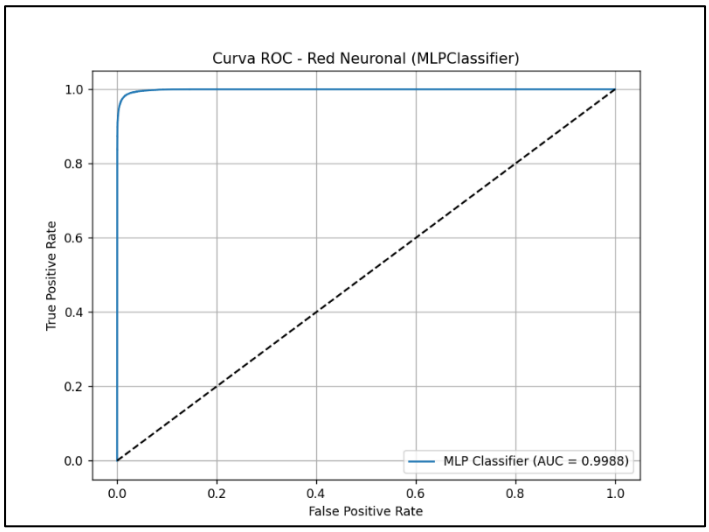
Tabla 3. Métricas de desempeño del modelo Red Neuronal Artificial (baseline)

Métrica	Valor
Accuracy	0.9835
Precision	0.9772
Recall	0.9821
F1 Score	0.9796
AUC	0.9988

Fuente: Elaboración propia

Aunque el modelo obtuvo un desempeño elevado, sus resultados fueron ligeramente inferiores a los alcanzados por los modelos basados en árboles. Este comportamiento puede atribuirse tanto a la arquitectura relativamente simple empleada como a la naturaleza tabular del conjunto de datos, escenario en el que los métodos de ensamblado suelen ofrecer un mejor rendimiento.

Figura 8. Curva ROC del modelo Red Neuronal Artificial (baseline).



Fuente: Elaboración propia

Comparación de los modelos (baseline)

Tabla 4. Comparación de modelos del enfoque baseline.

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Tiempo Entrenamiento
Árbol de Decisión	0.9938	0.9910	0.9936	0.9923	0.9987	Bajo
Random Forest	0.9943	0.9917	0.9941	0.9929	0.9998	Medio
Red Neuronal (MLP)	0.9835	0.9772	0.9821	0.9796	0.9988	Medio-Alto

Fuente: Elaboración propia

Los resultados corresponden al modelo de referencia (*baseline*) y se presentan únicamente con fines comparativos, ya que incorporan variables no disponibles antes del lanzamiento de la campaña.

Análisis

Los resultados muestran que los modelos basados en árboles, especialmente **Random Forest**, alcanzaron el mejor desempeño en términos de exactitud, precisión y capacidad discriminativa. La Red Neuronal Artificial obtuvo resultados ligeramente inferiores, mientras que el Árbol de Decisión destacó por su mayor interpretabilidad. En conjunto, estos resultados respaldan la selección de **Random Forest** como algoritmo para el desarrollo del modelo predictivo *ex ante*.

F. Resultados del modelo predictivo ex ante

Se evaluó el desempeño de modelos de clasificación entrenados exclusivamente con variables disponibles antes del lanzamiento de una campaña de crowdfunding (variables ex ante), con el fin de valorar su utilidad en escenarios reales de apoyo a la toma de decisiones.

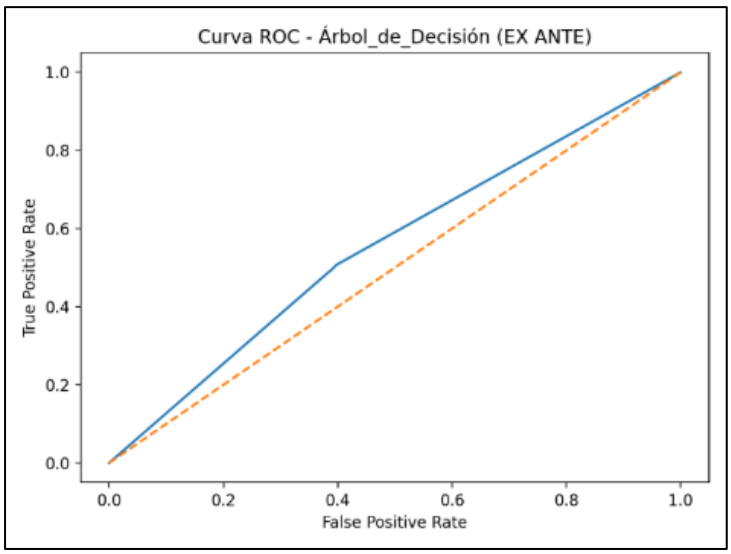
Los algoritmos analizados fueron Árbol de Decisión, Random Forest y una Red Neuronal Artificial (MLP). Su desempeño se evaluó mediante métricas estándar de clasificación, con énfasis en el área bajo la curva ROC (AUC), utilizada como indicador de la capacidad discriminativa de los modelos.

Modelo Árbol de Decisión

El Árbol de Decisión fue entrenado utilizando exclusivamente variables disponibles antes del lanzamiento de la campaña (variables ex ante).

La Figura 9 presenta la curva ROC del modelo, cuyo valor AUC fue aproximadamente 0,55, evidenciando una capacidad limitada para diferenciar entre campañas exitosas y fallidas. Aunque el algoritmo destaca por su simplicidad e interpretabilidad, su capacidad predictiva resulta reducida cuando solo se dispone de información ex ante.

Figura 9. Curva ROC del modelo Árbol de Decisión (ex ante).



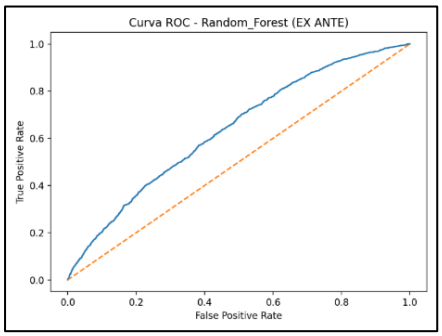
Fuente: Elaboración propia

Modelo Random Forest

El modelo Random Forest fue entrenado utilizando las mismas variables ex ante empleadas en el Árbol de Decisión.

La Figura 10 presenta la curva ROC del modelo, cuyo valor AUC fue aproximadamente 0,64, el más alto entre los algoritmos evaluados en la fase ex ante. Este resultado evidencia una mayor capacidad para diferenciar entre campañas exitosas y fallidas y confirma el mejor desempeño de Random Forest para apoyar la toma de decisiones en escenarios reales de crowdfunding.

Figura 10. Curva ROC del modelo Random Forest (ex ante).



Fuente: Elaboración propia

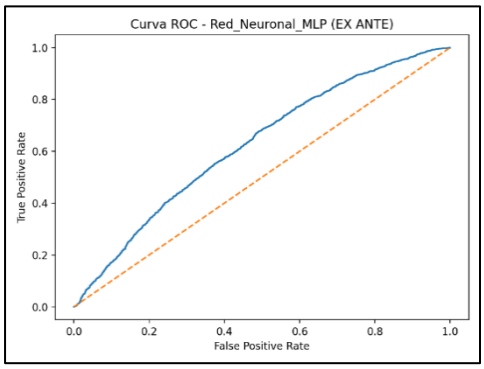
Modelo Red Neuronal Artificial (MLP)

Se implementó una Red Neuronal Artificial del tipo Perceptrón Multicapa (MLP), entrenada con las variables ex ante definidas para el estudio.

La Figura 11 presenta la curva ROC del modelo, cuyo valor AUC fue aproximadamente 0,62, superior al obtenido por el Árbol de Decisión, aunque ligeramente inferior al de Random Forest. Estos resultados evidencian una capacidad moderada para discriminar entre campañas exitosas y fallidas utilizando únicamente información disponible antes del lanzamiento de la campaña.

El desempeño observado sugiere que, para este conjunto de datos, los modelos basados en árboles y métodos de ensamblado ofrecen mejores resultados que una red neuronal con una arquitectura simple.

Figura 11. Curva ROC del modelo Red Neuronal Artificial (ex ante).



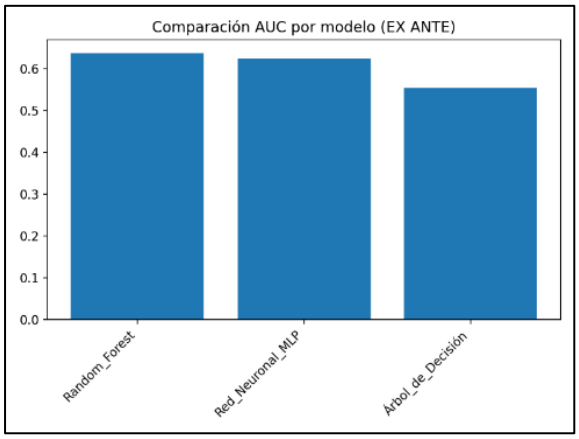
Fuente: Elaboración propia

Comparación de los modelos predictivos ex ante

La Figura 12 compara el área bajo la curva ROC (AUC) de los modelos evaluados. Los resultados muestran que Random Forest obtuvo el mejor desempeño predictivo, seguido por la Red Neuronal Artificial (MLP) y el Árbol de Decisión.

Aunque el desempeño fue inferior al del modelo de referencia (baseline), debido al uso exclusivo de variables ex ante, Random Forest mantuvo una capacidad discriminativa superior al azar, consolidándose como la alternativa más adecuada para apoyar la toma de decisiones antes del lanzamiento de campañas de crowdfunding.

Figura 12. Comparación del área bajo la curva ROC (AUC) de los modelos predictivos ex ante.



Fuente: Elaboración propia

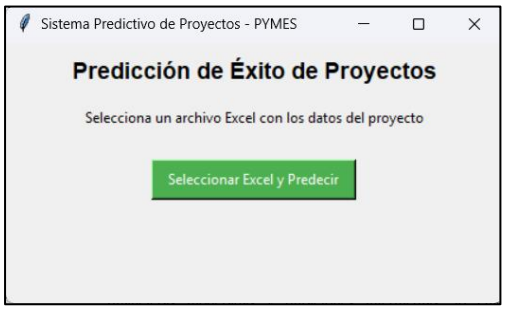
G. Uso del prototipo: caso de aplicación

Con el propósito de demostrar la aplicabilidad del modelo predictivo, se desarrolló un prototipo funcional en Python que estima la probabilidad de éxito de proyectos antes de su lanzamiento utilizando variables ex ante.

El prototipo incorpora una interfaz gráfica que permite cargar un archivo en formato Excel con información como la meta financiera (goal), la duración de la campaña (duration), el mes de lanzamiento (launch_month), el día de la semana (launch_dayofweek), la categoría del proyecto (main_category) y el país de origen (country).

Una vez cargados los datos, el sistema ejecuta automáticamente el modelo Random Forest y genera la clasificación del proyecto (exitoso o fallido), la probabilidad estimada de éxito y una recomendación para apoyar la toma de decisiones.

Figura 13. Interfaz gráfica del prototipo.



Fuente: Elaboración propia

La Figura 13 presenta la interfaz gráfica desarrollada, diseñada para facilitar la selección del archivo de entrada y simplificar el proceso de evaluación por parte del usuario.

Con el fin de ilustrar el funcionamiento del prototipo, se evaluó un proyecto con una meta de **2.000 dólares**, duración de **25 días**, categoría **Music** y país **Estados Unidos**. El modelo estimó una probabilidad de éxito del **38,21 %**, clasificándolo como **"Fallido"**, lo que sugiere revisar la estrategia antes del lanzamiento.

Figura 14. Archivo de entrada utilizado por el prototipo.

goal	duration	launch_month	launch_dayofwee	main_categ	country
1000	20	3	1	Theater	US

Fuente: Elaboración propia

La Figura 14 muestra un ejemplo del archivo Excel utilizado como entrada, el cual contiene las variables requeridas por el modelo para realizar la predicción.

En un segundo escenario se evaluó un proyecto con una meta financiera de 800 dólares, una duración de 20 días, categoría Theater y país Estados Unidos. El sistema estimó una mayor probabilidad de éxito y lo clasificó como "Exitoso".

Finalmente, un tercer proyecto, con una meta de 500 dólares, una duración de 18 días, categoría Art y país Estados Unidos, también fue clasificado como "Exitoso", confirmando la influencia de variables como la meta financiera y la categoría del proyecto.

En conjunto, estos resultados evidencian que el prototipo permite evaluar distintos escenarios antes del lanzamiento de una campaña, proporcionando una herramienta de apoyo para la planificación estratégica y la toma de decisiones.

Figura 15. Resultados generados por el prototipo.

goal	duration	launch_ month	launch_dayo fweek	main_cat egory	country	resultado	probabilidad_exit o (%)	recomendacion
1000	20	3	1	Theater	US	Exitoso	82.75	Alta probabilidad de éxito

Fuente: Elaboración propia

La Figura 15 presenta la salida generada por el sistema, la cual incluye la clasificación del proyecto, la probabilidad estimada de éxito y la recomendación correspondiente para apoyar la toma de decisiones.

DISCUSIÓN DE RESULTADOS

Los resultados obtenidos permiten diferenciar dos enfoques de análisis: el comportamiento histórico de las campañas de crowdfunding y la predicción ex ante de su probabilidad de éxito. Esta distinción es fundamental, ya que separa el análisis retrospectivo de la construcción de modelos orientados a apoyar la toma de decisiones antes del lanzamiento de una campaña.

El análisis exploratorio-descriptivo evidenció que variables como el país de origen, la categoría del proyecto, la duración de la campaña y la meta financiera influyen en el desempeño de las campañas. En particular, los proyectos de categorías artísticas y culturales, junto con campañas de duración intermedia y metas financieras moderadas, registraron mayores tasas de éxito. Estos resultados coinciden con estudios previos sobre crowdfunding (Elitzur et al., 2024; Putri Raflesia et al., 2023).

Respecto a los modelos predictivos, la diferencia entre el modelo de referencia (baseline) y los modelos ex ante se explica por el uso, en el primero, de variables observadas durante o después de la campaña, como usd_pledged_real y backers, las cuales generan fuga de información (data leakage). Aunque este modelo constituye un referente metodológico, no representa un escenario predictivo real.

Por el contrario, los modelos ex ante fueron entrenados exclusivamente con información disponible antes del lanzamiento de la campaña, proporcionando resultados metodológicamente más consistentes para escenarios reales de aplicación. Entre los algoritmos evaluados, Random Forest presentó el mejor desempeño, en concordancia con la literatura que destaca la eficacia de los métodos de ensamblado en problemas de clasificación con datos tabulares y relaciones no lineales (Burggräf et al., 2024; Yoo et al., 2023).

Si bien los valores de AUC obtenidos fueron inferiores a los del modelo baseline, este comportamiento era esperable por la ausencia de información posterior al lanzamiento. En consecuencia, los resultados

demuestran que Random Forest constituye una herramienta útil para apoyar la evaluación temprana de campañas de crowdfunding, favoreciendo una toma de decisiones más informada antes del inicio de los proyectos.

CONCLUSIONES

La presente investigación demostró la viabilidad de aplicar técnicas de aprendizaje automático para la predicción ex ante del éxito de campañas de crowdfunding utilizando exclusivamente variables disponibles antes del lanzamiento de los proyectos. Los resultados confirmaron que el análisis de datos históricos permite identificar patrones asociados a la categoría del proyecto, el país de origen, la duración de la campaña y la meta financiera, factores que influyen en la probabilidad de éxito.

Asimismo, se comprobó que, aunque los modelos entrenados únicamente con variables ex ante presentan un desempeño inferior al del modelo de referencia (baseline), ofrecen resultados metodológicamente consistentes para escenarios reales, al evitar la fuga de información (data leakage) y permitir estimaciones basadas exclusivamente en información disponible antes del lanzamiento de la campaña.

Entre los algoritmos evaluados, Random Forest presentó el mejor equilibrio entre capacidad predictiva, estabilidad y robustez, consolidándose como la alternativa más adecuada para el prototipo desarrollado. Aunque su desempeño fue moderado, demostró una capacidad discriminativa superior al azar, validando su utilidad como herramienta de apoyo para la evaluación preliminar de campañas de crowdfunding y la toma de decisiones estratégicas.

Finalmente, este estudio aporta evidencia sobre la utilidad de los modelos predictivos ex ante en plataformas de financiamiento colectivo y presenta un prototipo funcional para apoyar la evaluación anticipada de proyectos. Como líneas futuras de investigación, se recomienda incorporar variables exógenas, optimizar los hiperparámetros y evaluar el modelo en otras plataformas de crowdfunding, con el propósito de mejorar su capacidad predictiva y generalización.

REFERENCIAS BIBLIOGRÁFICAS

Bang, S., Aarvold, M. O., Hartvig, W. J., Olsson, N. O. E., & Rauzy, A. (2022). Application of machine learning to limited datasets: Prediction of project success. *ITcon, 27*, 732–752.

Burggräf, P., Steinberg, F., Sauer, C. R., & Nettesheim, P. (2024). Machine learning implementation in small and medium sized enterprises: Insights and recommendations from a quantitative study. *Computer Assisted Approaches in Production Engineering*, 1–10.

Carrillo Zenteno, J. A., Ormaza Vintimilla, A. D., & Santacruz Espinoza, J. J. (2024). El impacto de la inteligencia artificial en la gestión de proyectos. *Revista Tecnológica ESPOL, 36*(2), 52–66.

Elitzur, R., Katz, N., Muttath, P., & Soberman, D. (2024). The power of machine learning methods to predict crowdfunding success: Accounting for complex relationships efficiently. *Elsevier*, 3–14.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.

Hussain, S. B., Mouzam Uddin, M. K., Sayeed Uddin, S. M., & Ali, M. F. (2022). The impact of project management practices on project success. *International Research Journal of Engineering and Technology, 9*(5), 995–1001.

Ilieş, L., Crişan, E., & Mureşan, I. N. (2010). Best practices in project management. *Review of International Comparative Management, 11*(1), 43–49.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Joakin, I. (2021). *Aplicación de tecnologías de aprendizaje automático para predecir negocios y tomar decisiones* [Trabajo de grado, Universidad Nacional de La Plata].

Kaggle. (2018). *Kickstarter projects dataset*.

Moore, A. D. (2021). *Python GUI programming: Design and build functional and user-friendly GUI applications* (2nd ed.). Packt Publishing.

Pedroso, M. (2017). *Application of machine learning techniques in project management* [Master's thesis, Instituto Superior Técnico, Universidade de Lisboa].

Peral, G. (2015). *Elaboración y evaluación de modelos predictivos de negocio* [Trabajo Fin de Grado, Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid].

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

Putri Raflesia, S., Lestarini, D., Dhini Kurnia, R., & Yunika Hardiyanti, D. (2023). Using machine learning approach towards successful crowdfunding prediction. *Bulletin of Electrical Engineering and Informatics, 12*(4), 2439–2443.

Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2* (3rd ed.). Packt Publishing.

Seltman, H. J. (2018). Exploratory data analysis. En *Experimental design and analysis* (pp. 61–99). Carnegie Mellon University.

Ugbebor, F. O., Adeteye, D. A., & Ugbebor, J. O. (2024). Predictive analytics models for SMEs to forecast market trends, customer behavior, and potential business risks. *Journal of Knowledge Learning and Science Technology, 3*(3), 356–377.

Yoo, H. S., Jung, Y. L., & Jun, S. (2023). Prediction of SMEs' R&D performances by machine learning for project selection. *Scientific Reports, 13*, Article 7598.

Zinkevich, M. (2015). *Rules of machine learning: Best practices for ML engineering*. Google Research. https://developers.google.com/machine-learning/guides/rules-of-ml