



Universidad
Internacional
de Valencia

ELABORACIÓN DE PROPUESTA DE PROYECTO TFM

SERGIO MOYA MARÍN

ÍNDICE

1. Título	3
2. Resumen	3
3. Introducción y Justificación	4
Antecedentes.....	4
Planteamiento del Problema	4
Motivación, Relevancia e Impacto Potencial	5
4. Objetivos del Proyecto	6
Objetivo General	6
Objetivos Específicos	6
5. Metodología Propuesta	7
Conjunto de Datos y Preprocesamiento	7
Modelos de Aprendizaje Automático a Evaluar	7
Técnicas para el Manejo del Desbalance de Clases	8
Métricas de Evaluación	8
Herramientas y Recursos	8
6. Resultados Esperados	9
7. Contribución del Proyecto	9
8. Planificación Temporal	10
9. Referencias Bibliográficas Preliminares.....	10

1. Título

Aplicación de Técnicas de Aprendizaje Automático para la Detección de Transacciones Anómalas en Operaciones Financieras

2. Resumen

El crecimiento exponencial de las transacciones digitales ha posicionado la detección de fraude como un desafío de alta prioridad para la industria financiera. La constante evolución de las tácticas fraudulentas, junto con las complejidades inherentes a los conjuntos de datos del mundo real exige el desarrollo y la validación de sistemas de detección que no solo sean precisos, sino también robustos, interpretables y metodológicamente sólidos. Esta propuesta describe un proyecto de investigación diseñado para abordar estas necesidades mediante un análisis sistemático y comparativo de diversas técnicas de aprendizaje automático.

Este proyecto se centra en la detección de fraude con tarjetas de crédito, un problema definido por desafíos técnicos que a menudo son subestimados en la literatura. El principal de ellos es el desbalance extremo de clases, donde las transacciones fraudulentas constituyen una fracción minúscula del total ($\approx 0.17\%$), invalidando métricas tradicionales como la exactitud. Adicionalmente, una porción significativa de la literatura existente adolece de fallos metodológicos, como la fuga de datos (*data leakage*), que inflan artificialmente los resultados y conducen a conclusiones poco fiables. Finalmente, existe una tensión fundamental entre la precisión de los modelos complejos (considerados "cajas negras") y la necesidad de interpretabilidad, un requisito indispensable para la auditoría y el cumplimiento normativo en el sector financiero.

El objetivo central de esta investigación es establecer un benchmark reproducible y metodológicamente riguroso para corregir estas deficiencias. En lugar de proponer una única solución, se realizará un estudio sistemático y comparativo sobre la efectividad de algoritmos de clasificación supervisada (Regresión Logística, Random Forest y XGBoost) y la estrategia de ponderación de clases para manejar el desbalance. Para ello, se utilizará el conjunto de datos público y de referencia *creditcard.csv* de Kaggle. El rendimiento se medirá con métricas robustas como el Recall, F1-Score y el Área Bajo la Curva de Precisión-Recall (AUPRC).

La contribución esperada de este proyecto es una guía metodológica rigurosa que valide las técnicas genuinamente efectivas y establezca una línea base fiable para futuras investigaciones. Al ofrecer recomendaciones para evitar errores comunes, el estudio beneficiará tanto al ámbito académico, promoviendo una mayor

reproducibilidad, como a las aplicaciones prácticas en el sector financiero, facilitando la construcción de sistemas de detección de fraude más fiables y transparentes.

3. Introducción y Justificación

Para establecer la necesidad de un nuevo estudio comparativo, esta sección analiza críticamente el estado del arte en la detección de fraude, exponiendo las deficiencias metodológicas sistémicas que invalidan una porción significativa de los resultados publicados y que, por tanto, motivan esta investigación. Al delinear los antecedentes, los desafíos técnicos y la motivación subyacente, se articula el valor y la relevancia del proyecto, sentando las bases para el desarrollo de objetivos claros y una metodología rigurosa.

Antecedentes

La transición global hacia una economía sin efectivo ha consolidado a las tarjetas de crédito como el pilar del comercio digital. Esta creciente dependencia, si bien ofrece conveniencia y eficiencia, ha amplificado simultáneamente los riesgos asociados al fraude. El fraude con tarjetas de crédito se ha convertido en una de las amenazas más persistentes y de mayor impacto financiero para el sector, exigiendo la implementación de sistemas de detección cada vez más sofisticados y adaptativos para proteger tanto a las instituciones como a los consumidores.

Planteamiento del Problema

La detección de fraude con tarjetas de crédito es un problema de clasificación binaria cuya aparente simplicidad oculta desafíos técnicos y metodológicos que, si no se abordan explícitamente, conducen a soluciones ineficaces y conclusiones inválidas.

- **Desbalance Extremo de Clases:** Los conjuntos de datos de transacciones se caracterizan por un severo desbalance. En el dataset de referencia de este estudio, los casos de fraude representan solo el 0.17% del total. Este desequilibrio provoca que los algoritmos de clasificación estándar tiendan a ignorar la clase minoritaria (fraude), optimizando su rendimiento para la clase mayoritaria (no fraude) y fallando así en su objetivo principal.
- **Insuficiencia de Métricas Tradicionales:** Como consecuencia directa del desbalance, métricas como la exactitud (*accuracy*) son engañosas. Un clasificador trivial que prediga "no fraude" para todas las transacciones alcanzaría una exactitud del 99.83%, aparentando un rendimiento excelente mientras es completamente inútil en la práctica, al no detectar un solo caso de fraude.

- **Deficiencias en la Literatura Existente:** Una revisión crítica de la literatura revela fallos metodológicos recurrentes que comprometen la validez de muchos estudios. El error más crítico es la **fuga de datos (*data leakage*)**, un fallo cuya prevalencia ha sido documentada por autores como Pozzolo et al. (2024). Ocurre cuando técnicas de remuestreo (como SMOTE) se aplican al conjunto de datos completo *antes* de la división en entrenamiento y prueba. Esta práctica contamina el conjunto de prueba con información sintética, lo que conduce a una sobreestimación irreal del rendimiento del modelo y a conclusiones inválidas, perpetuando un ciclo de resultados inflados y no reproducibles.
- **El Dilema de la Interpretabilidad:** Existe un compromiso inherente entre la precisión predictiva y la interpretabilidad. Los modelos más complejos y de mayor rendimiento, como las redes neuronales profundas o los ensambles, a menudo funcionan como "cajas negras". Su opacidad dificulta la auditoría, merma la confianza de los analistas y complica el cumplimiento de normativas financieras que exigen transparencia. Este desafío, como señalan Hasan & Gazi (2025), es un problema de investigación reconocido que busca cerrar la brecha entre la precisión y la explicabilidad.

Motivación, Relevancia e Impacto Potencial

La motivación central de este Trabajo de Fin de Máster (TFM) no es simplemente construir otro modelo de detección, sino abordar la necesidad de un análisis comparativo que sea metodológicamente riguroso y transparente. El propósito es determinar qué técnicas de aprendizaje automático y estrategias de manejo de desbalance son genuinamente efectivas, bajo qué condiciones y evitando los fallos comunes que invalidan gran parte de la investigación existente.

El impacto potencial de este proyecto es doble:

1. **A nivel académico:** Contribuir con un estudio reproducible que sirva como una referencia sólida y un benchmark fiable para futuras investigaciones. Al documentar y evitar explícitamente los errores metodológicos que han plagado el campo, el proyecto busca elevar el estándar de rigor en el área.
2. **A nivel industrial:** Ofrecer a las instituciones financieras una visión clara y basada en evidencia para la selección, implementación y validación de sistemas de detección de fraude. Los resultados proporcionarán una guía práctica para construir soluciones que sean a la vez robustas, eficientes e interpretables.

Al justificar la necesidad de esta investigación, se establece la base para definir un conjunto de objetivos claros y alcanzables que guiarán su ejecución.

4. Objetivos del Proyecto

La definición de objetivos claros, específicos y medibles es fundamental para guiar el desarrollo del proyecto, delimitar su alcance y permitir una evaluación objetiva de sus resultados. Esta sección articula la finalidad de la investigación y desglosa las metas concretas que se persiguen para abordar el problema planteado.

Objetivo General

Evaluar y comparar sistemáticamente la efectividad de diversas técnicas de aprendizaje automático y estrategias de manejo de desbalance de clases para la detección de transacciones financieras anómalas, con un enfoque en la rigurosidad metodológica y la interpretabilidad de los modelos.

Objetivos Específicos

Para alcanzar el objetivo general, se plantean los siguientes objetivos específicos:

1. **Analizar el estado del arte** en la detección de fraude con tarjetas de crédito para identificar los algoritmos más prometedores y documentar los fallos metodológicos más comunes reportados en la literatura científica.
2. **Implementar un pipeline de preprocesamiento de datos** que garantice la correcta separación de los conjuntos de entrenamiento y prueba para prevenir la fuga de datos, utilizando el dataset público de Kaggle sobre fraude en tarjetas de crédito.
3. **Desarrollar y entrenar un conjunto de modelos de referencia** supervisados, específicamente Regresión Logística, Random Forest y XGBoost, para establecer una línea base de rendimiento.
4. **Evaluar el impacto de la ponderación de clases (`class_weight`)** para el manejo del desbalance, comparando el rendimiento de los modelos con y sin la aplicación de esta estrategia.
5. **Validar y comparar el rendimiento de los modelos** utilizando un conjunto de métricas de evaluación apropiadas para datos desbalanceados, priorizando el Recall, F1-Score y el Área Bajo la Curva Precisión-Recall (AUPRC).
6. **Analizar la importancia de las características (`feature importances`)** nativas del modelo en los modelos con mejor rendimiento para ofrecer una explicación de sus predicciones.

La consecución de estos objetivos se llevará a cabo siguiendo una metodología estructurada, que se detalla en la siguiente sección.

5. Metodología Propuesta

Este apartado constituye el núcleo técnico de la propuesta, donde se detallan de forma estructurada las fases, herramientas y procedimientos que se seguirán para alcanzar los objetivos planteados. La metodología está diseñada para garantizar la reproducibilidad de los experimentos y el rigor científico del estudio, abordando explícitamente los desafíos metodológicos identificados en la literatura.

Conjunto de Datos y Preprocesamiento

- **Fuente de Datos:** Se utilizará el conjunto de datos Credit Card Fraud Detection disponible en la plataforma Kaggle. Este dataset es un estándar de facto en la investigación de la detección de fraude con tarjetas de crédito y contiene 284,807 transacciones. Incluye 492 casos de fraude, lo que representa un 0.17% del total, y consta de 30 características: 28 componentes principales anonimizados (V1-V28) obtenidos mediante PCA, y las variables Time y Amount.
- **Pipeline de Preprocesamiento:** El tratamiento de los datos seguirá un pipeline estricto para asegurar la integridad de los resultados:
 1. **Escalado y Selección de Características:** Se aplicará StandardScaler a la columna Amount para normalizar su distribución. La característica Time, que representa segundos transcurridos, es de naturaleza cíclica y requeriría una ingeniería de características compleja (ej., transformaciones de seno/coseno) para ser útil. Para mantener el enfoque en la comparación de los algoritmos y las técnicas de remuestreo, se excluirá esta variable del alcance del estudio, garantizando una comparación más directa de los métodos centrales.
 2. **División de Datos:** Se realizará una división estratificada (stratified split) del conjunto de datos en un 80% para entrenamiento y un 20% para prueba. Se enfatiza que esta división se realizará **antes** de aplicar cualquier técnica de remuestreo, un paso crucial para evitar la fuga de datos y garantizar una evaluación imparcial del rendimiento de los modelos.

Modelos de Aprendizaje Automático a Evaluar

Se implementará y evaluará un conjunto de modelos para realizar una comparación exhaustiva:

- **Regresión Logística (Logistic Regression):** Seleccionado como modelo lineal de línea base por su simplicidad y alta interpretabilidad.

- **Random Forest:** Seleccionado por su robustez y buen rendimiento general en problemas de clasificación.
- **XGBoost:** Elegido por su eficiencia y su capacidad nativa para manejar datos desbalanceados mediante el parámetro `scale_pos_weight`.

Técnicas para el Manejo del Desbalance de Clases

Se comparará el rendimiento de los modelos bajo diferentes estrategias de manejo de desbalance:

- **Línea Base:** Entrenar los modelos sobre los datos de entrenamiento originales sin ninguna modificación para establecer un punto de referencia de rendimiento.
- **Ponderación de Clases (Cost-Sensitive Learning):** Ajustar el parámetro `class_weight` o `scale_pos_weight` en los modelos que lo soporten para penalizar más fuertemente los errores en la clase minoritaria.

Métricas de Evaluación

La selección de métricas es crítica en problemas con datos desbalanceados. Aunque la exactitud se reportará por completitud, el análisis principal se basará en las siguientes métricas:

- **Recall (Sensibilidad):** Mide la proporción de fraudes reales que fueron identificados correctamente ($TP / (TP + FN)$). Es una métrica crucial, ya que el objetivo principal es minimizar los fraudes no detectados.
- **Precision (Precisión):** Mide la proporción de transacciones marcadas como fraude que realmente lo eran ($TP / (TP + FP)$). Es importante para minimizar el número de falsas alarmas.
- **F1-Score:** La media armónica de Precision y Recall, ofreciendo una métrica que balancea ambos objetivos.
- **Área Bajo la Curva Precisión-Recall (AUPRC):** Considerada una de las métricas más informativas para evaluar clasificadores en datos con alto desbalance, ya que se centra en el rendimiento de la clase minoritaria.

Herramientas y Recursos

El proyecto se desarrollará utilizando el siguiente stack tecnológico de código abierto:

- **Lenguaje de Programación:** Python 3.x.
- **Librerías Principales:**

- Pandas y numpy para manipulación de datos
- Scikit-learn para preprocesamiento y modelos base
- **XGBoost** para el modelo de boosting
- Funciones nativas de los modelos para la interpretabilidad (**feature importances**).

A partir de esta metodología se generará un conjunto de resultados concretos y analizables.

6. Resultados Esperados

Esta sección define los entregables tangibles y verificables que se producirán al finalizar el proyecto, los cuales permitirán medir de forma objetiva el éxito en la consecución de los objetivos planteados y servirán como la principal contribución del trabajo.

Se espera generar los siguientes resultados:

1. Un **informe comparativo** detallado en formato de tabla Markdown que resuma el rendimiento (Recall, Precision, F1-Score, AUPRC, MCC) de cada combinación de modelo y técnica de manejo de desbalance.
2. Un conjunto de visualizaciones de apoyo y gráficos de **importancia de características nativas** para los modelos de mejor rendimiento.
3. Un **documento de análisis y conclusiones** que interprete los resultados, discuta las fortalezas y debilidades de cada enfoque, y ofrezca recomendaciones fundamentadas sobre la selección de técnicas para problemas de CCFD.
4. Un **repositorio de código fuente** completo y comentado que permita la total reproducibilidad de los experimentos realizados.

7. Contribución del Proyecto

Esta sección articula el valor añadido del proyecto, destacando cómo sus resultados contribuirán al cuerpo de conocimiento existente en el campo de la detección de fraude, con un enfoque en el rigor metodológico y el valor académico.

- **Contribución Metodológica:** El principal aporte del proyecto será la realización de un estudio sistemático, riguroso y reproducible que clarifique la efectividad real de diferentes técnicas de aprendizaje

automático y estrategias de manejo de desbalance. Al evitar explícitamente los errores metodológicos comunes identificados en la literatura, como la fuga de datos, este trabajo establecerá un benchmark fiable que servirá como guía para que tanto investigadores como profesionales puedan diseñar y evaluar sistemas de detección de manera correcta.

- **Contribución al Conocimiento Académico:** El proyecto generará una guía empírica que contrastará el rendimiento de modelos clásicos supervisados, enfoques no supervisados de detección de anomalías y arquitecturas híbridas bajo un marco de evaluación unificado y apropiado para el problema. Al publicar los resultados y el código de manera abierta, se proporcionará una base sólida y transparente para futuras investigaciones, fomentando una mayor reproducibilidad en el campo.

8. Planificación Temporal

Para garantizar una ejecución ordenada y la consecución de los objetivos dentro del plazo establecido, se propone el siguiente cronograma de trabajo, distribuido en fases clave a lo largo de cinco meses.

Fases del proyecto	Noviembre	Diciembre	Enero	Febrero	Marzo
1. Revisión Bibliográfica y Estado del Arte					
2. Preparación del Entorno y Preprocesamiento de Datos					
3. Implementación y Entrenamiento de Modelos					
4. Experimentación y Evaluación de Resultados					
5. Análisis de Resultados y Redacción del TFM					
6. Preparación de la Defensa Final					

9. Referencias Bibliográficas Preliminares

La siguiente es una lista preliminar de referencias científicas que proporcionan el marco teórico y metodológico para esta propuesta, fundamentales para comprender el estado del arte, los desafíos del problema y las mejores prácticas en el campo de la detección de fraude.

- Abdelhamid, M., & Desai, A. (2024). Balancing the Scales: A Comprehensive Study on Tackling Class Imbalance in Binary Classification. *arXiv preprint*.

- Dal Pozzolo, A., et al. (2017). *Reproducible Machine Learning for Credit Card Fraud detection - Practical handbook*.
- Gaav, T. A., et al. (2025). *Recent Advances in Credit Card Fraud Detection: An Analytical Review of Frameworks, Methodologies, Datasets, and Challenges*. Journal of Future Artificial Intelligence and Technologies.
- Hasan, N. G. M. R., & Gazi, M. S. (2025). *Explainable AI for credit card fraud detection: Bridging the gap between accuracy and interpretability*. World Journal of Advanced Research and Reviews.
- Hayat & Magnier (2025). *Data Leakage and Deceptive Performance*
- Shanaa, M. (2025). *A Hybrid Anomaly Detection Framework Combining Supervised and Unsupervised Learning for Credit Card Fraud Detection*. F1000Research.