

Anexo I. Registro del Título del Trabajo Fin de Grado (TFG)

NOMBRE DEL ALUMNO: Luis Mielgo Larriba

PROGRAMA: E2+BA

GRUPO: B

FECHA: 27/10/2023

Director Asignado: Knop Muszynski _____, Roberto _____
Apellidos Nombre

Título provisional del TFG:

Análisis comparativo de técnicas de Machine Learning en modelos de Riesgo de Crédito

ADJUNTAR PROPUESTA (máximo 4 páginas: Índice provisional, objetivos, metodología y bibliografía)

Introducción

En el contexto dinámico de la gestión financiera, la evaluación precisa del riesgo crediticio se ha vuelto imperativa para la estabilidad de las instituciones. Este Trabajo de Fin de Grado se enfocará en realizar un análisis comparativo de técnicas de Machine Learning aplicadas a modelos de riesgo de crédito, abarcando desde regresión logística hasta redes neuronales. La complejidad inherente a los datos financieros demanda enfoques más sofisticados, y este estudio se propone evaluar la eficacia, interpretabilidad y robustez de diversas técnicas. El objetivo es proporcionar una guía informada para la elección de modelos en la práctica financiera, contribuyendo así al avance continuo en la gestión eficiente del riesgo crediticio.

- **Riesgo Crediticio:** El riesgo crediticio, como componente fundamental en la gestión financiera, se refiere a la posibilidad de que un prestatario incumpla con sus obligaciones de pago. Este riesgo no solo impacta a las instituciones financieras, sino que también influye en la estabilidad y salud del sistema económico en su conjunto. La evaluación precisa del riesgo crediticio es, por lo tanto, crucial para garantizar una asignación de recursos responsable y la toma de decisiones fundamentada.

- **Evolución de Modelos Tradicionales a Técnicas de Machine Learning:** Los modelos de riesgo crediticio tradicionales, basados en estadísticas y reglas predefinidas, han demostrado cierta efectividad, pero enfrentan limitaciones en su capacidad para capturar relaciones no lineales y adaptarse a patrones emergentes. La evolución tecnológica ha propiciado la adopción de técnicas de Machine Learning, las cuales poseen la capacidad intrínseca de aprender de los datos y adaptarse a la complejidad inherente en los comportamientos crediticios.
- **Tipos de Técnicas de Machine Learning Aplicadas a Riesgo de Crédito:** Este marco teórico se enfocará en explorar diversas técnicas de Machine Learning que han demostrado ser pertinentes en la predicción y evaluación del riesgo crediticio. Esto incluye modelos clásicos como la regresión logística, máquinas de soporte vectorial (SVM), y árboles de decisión, así como enfoques más avanzados como las redes neuronales y los ensambles de modelos. Se analizará la idoneidad de cada técnica en función de la naturaleza de los datos financieros y la complejidad de las relaciones subyacentes.

Índice

1. Introducción
2. Marco Teórico
 - a. Fundamentos de Machine Learning en Riesgo de Crédito
 - i. Tipos de Algoritmos de Machine Learning
 - ii. Aplicaciones Específicas en Riesgo de Crédito
 - b. Selección de Variables y Preprocesamiento de Datos
 - c. Evaluación de Modelos en Riesgo de Crédito
 - d. Consideraciones Éticas y Legales en el Uso de Datos Financieros
3. Metodología
 - a. Descripción de la Base de Datos
 - b. Selección de Algoritmos de Machine Learning
 - c. Diseño Experimental y Métricas de Evaluación
 - d. Herramientas y Tecnologías Utilizadas
4. Análisis y Resultados
 - a. Descripción de Resultados Obtenidos
 - b. Comparación de Rendimiento entre Algoritmos
 - c. Interpretación de Modelos y Variables Relevantes
 - d. Discusión de Hallazgos
5. Conclusiones
 - a. Resumen de Resultados
 - b. Contribuciones del Estudio
 - c. Implicaciones Prácticas
6. Referencias Bibliográficas
7. Anexos

Objetivos

- Evaluar el Rendimiento Predictivo de los distintos modelos. Analizar y comparar el rendimiento predictivo de diversas técnicas de machine learning en la estimación del riesgo de crédito, para luego identificar cuáles de estas técnicas ofrecen resultados más precisos y confiables en la predicción de la probabilidad de incumplimiento crediticio.

- Identificar Variables Relevantes. Utilizar técnicas de selección y análisis de variables para mejorar la comprensión de los factores que impactan significativamente en la evaluación del riesgo, lo que puede ser crucial para la toma de decisiones informada en instituciones financieras.
- Analizar la Interpretabilidad de los Modelos. Realizar un análisis detallado de la interpretabilidad de los modelos. Este objetivo se centra en comprender cómo explican los modelos sus decisiones, especialmente en entornos financieros donde la transparencia y la comprensión son esenciales para la confianza de los stakeholders y la aceptación regulatoria.

Metodología

- Selección de la Base de Datos: Seleccionaremos una base de datos relevante y representativa para el estudio de riesgo de crédito. La base de datos debe contener información histórica de clientes, transacciones financieras y variables relevantes.
- Preprocesamiento de Datos: Realizaremos un análisis exploratorio de los datos para identificar posibles problemas, como valores atípicos, datos faltantes, redundancias o outliers. Implementaremos técnicas de preprocesamiento, como imputación de datos faltantes o normalización, para garantizar la calidad de los datos.
- Selección de Variables: Utilizaremos técnicas de selección de variables para identificar aquellas características más relevantes en la predicción del riesgo crediticio. Esto puede incluir métodos estadísticos, como análisis de importancia de características, y algoritmos específicos.
- Elección de Algoritmos de Machine Learning: Seleccionaremos una variedad de algoritmos de machine learning, como regresión logística, árboles de decisión, máquinas de vectores de soporte (SVM), y redes neuronales, para comparar su rendimiento en la predicción de riesgo crediticio.
- División de Datos y Cross Validation: Dividiremos la base de datos en conjuntos de train y test para evaluar el rendimiento de los modelos. Aplicaremos técnicas de validación cruzada para garantizar la robustez y la generalización de los resultados, evitando problemas de overfitting.
- Entrenamiento y Evaluación de Modelos: Entrenaremos cada modelo seleccionado utilizando el conjunto de entrenamiento y evaluaremos su rendimiento utilizando métricas adecuadas para la predicción de riesgo crediticio, como precisión, sensibilidad, especificidad y área bajo la curva ROC (AUC-ROC).
- Análisis Comparativo: Realizaremos un análisis detallado de los resultados obtenidos, comparando el rendimiento de los diferentes modelos. Identificar fortalezas y debilidades de cada técnica en términos de precisión e interpretabilidad.
- Interpretación de Modelos: Analizaremos la interpretabilidad de los modelos, especialmente en términos de la importancia asignada a diferentes variables.
- Presentación de Resultados: Presentaremos los resultados de manera clara y visual mediante gráficos, tablas y métricas de evaluación, destacando las principales conclusiones derivadas del análisis comparativo.
- Discusión de Limitaciones y Consideraciones Éticas: Identificar y discutir posibles limitaciones del estudio, como sesgos en los datos, limitaciones algorítmicas y cuestiones éticas asociadas al uso de modelos de machine learning en el sector financiero.

Bibliografía preliminar:

EUROPEAN BANKING AUTHORITY. (2021). *EBA DISCUSSION PAPER ON MACHINE LEARNING FOR IRB MODELS*. Paris: EBA.

EUROPEAN BANKING AUTHORITY (2023). *MACHINE LEARNING TOOLS FOR IRB MODELS*. Paris: EBA.

Firma del estudiante:

Fecha: