

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

NOMBRE DEL ALUMNO: Beatriz Tejedor Canet

PROGRAMA: E6-Analytics

GRUPO: A

FECHA: 21/10/2024

Director Asignado: Vallez Fernández

Apellidos

Carlos Miguel

Nombre

Título provisional del TFG-BA:

La aplicación de técnicas de Machine Learning para predecir empresas con mayor potencial de rentabilidad futura: Un enfoque basado en la clasificación.

ADJUNTAR PROPUESTA (máximo 2 páginas: objetivo, bibliografía, metodología e índice preliminares)

Firma del estudiante:



Firma del tutor:

Fecha:

1.ÍNDICE:

1. RESUMEN
PALABRAS CLAVE
2. ABSTRACT
KEY WORDS
3. ÍNDICE DEL CONTENIDO
4. INDICE DE TABLAS Y GRÁFICOS
5. INTRODUCCIÓN
6. OBJETIVOS
7. JUSTIFICACIÓN
8. METODOLOGIA
9. ESTRUCTURA
10. MARCO TEÓRICO
11. CONCEPTOS TEÓRICOS
12. ESTUDIOS PREVIOS APLICADOS DE ML APLICADOS A CLASIFICACIÓN
13. ANÁLISIS EMPÍRICO
14. DATOS
15. METODOLOGÍA
16. REVISIÓN DE RESULTADOS
17. CONCLUSIONES Y FUTURAS LINEAS DE INVESTIGACIÓN
18. DECLARACIONES DE HERRAMIENTAS DE IA
19. BIBLIOGRAFÍA
20. ANEXO

2.OBJETIVO:

El objetivo principal de este estudio es predecir qué empresas tendrán mayor rentabilidad en el futuro aplicando diversos modelos de machine learning, con el fin de comparar su efectividad y desempeño. Estos modelos serán de gran valor para empresas de private equity, ya que permitirán automatizar los procesos de análisis y, en cuestión de segundos, identificar las mejores oportunidades de inversión y adquisición mediante la aplicación de los modelos más precisos, optimizando la formación de carteras de inversión.

El análisis se centrará específicamente en las empresas que forman parte del índice S&P 500, también conocido como Standard & Poor's 500, uno de los índices bursátiles más relevantes y seguidos a nivel mundial. Este índice representa a las 500 empresas más grandes y líquidas que cotizan en los principales mercados de valores de Estados Unidos, como la Bolsa de Nueva York (NYSE) y el NASDAQ.

3. METODOLOGÍA:

La metodología que se va a seguir en este estudio estará basada en un enfoque cuantitativo, para ello, vamos a utilizar técnicas de ML para poder predecir y con ello poder clasificar las posibles empresas que puedan tener la mejor rentabilidad futura. Este proceso se va a realizar en varias fases secuenciales clave para garantizar la precisión y efectividad de los modelos predictivos. Cada una de ellas estará diseñada para optimizar la extracción, transformación y análisis de los datos, así como para realizar una comparación rigurosa de los diferentes modelos de machine learning utilizados.

Para ello, la primera etapa del proceso será la fase de ETL (extracción, transformación y carga). Esta fase se centrará en la recopilación de datos financieros históricos de las empresas, ya sean sus ingresos, costes, márgenes de beneficio y capitalización bursátil y entre otros, a través del índice S&P, proporcionando una base sólida para el análisis posterior.

Posteriormente, se llevará a cabo la transformación de los datos, este paso es esencial para garantizar la calidad de estos. Durante esta etapa los datos se limpiarán, eliminando outliers, y se normalizarán los valores para así hacerlos comparables. Finalmente, los datos transformados se cargarán en un entorno adecuado para su posterior uso en el entrenamiento de los modelos.

Una vez concluida la primera fase de preparación de datos, se procederá a la creación de los modelos de Machine Learning. En esta fase se entrenarán los diversos modelos los cuales serán los de regresión logística, K-Nearest Neighbors (k-NN), árboles de decisión, Random Forest, Gradient Boosting Machines (GBM), y redes neuronales artificiales (ANN). Cada uno de estos modelos será entrenado y ajustado para predecir la clasificación de las empresas, permitiendo así comparar su rendimiento. Se utilizarán técnicas como la validación cruzada y el ajuste de hiperparámetros al entrenar los modelos para garantizar que los modelos puedan generalizarse adecuadamente cuando se trabaje con nuevos datos. Intentaremos optimizar el rendimiento de cada modelo para obtener las predicciones más precisas posibles.

Una vez obtenidos esos modelos, la siguiente fase será compararlos. En este proceso, se utilizará un enfoque basado en la metodología Agile, esta permitirá trabajar de forma iterativa. A través de ciclos de interacción y refinamiento, los modelos se ajustarán a los requisitos específicos identificados en cada iteración. La retroalimentación recibida durante cada ciclo optimiza el rendimiento de los modelos durante todo el proceso.

Finalmente, llevaremos a cabo la implementación y evaluación de los modelos que nos permitirá verificar la aplicabilidad de las soluciones desarrolladas en escenarios reales. Para ello, utilizaremos técnicas de validación cruzada y evaluación en conjuntos de datos de prueba para garantizar que los modelos no solo tengan un buen rendimiento en el conjunto de entrenamiento, sino que también generalicen de manera efectiva a datos no vistos. Las métricas de rendimiento como la de precisión, recall, F1-score y AUC-ROC serán clave para evaluar y comparar los modelos, asegurando que seleccionemos el más adecuado para predecir la rentabilidad futura de las empresas.

BIBLIOGRAFIA:

1. Erickson, B. J., & Kitamura, F. (2021). Magician's Corner: 9. Performance Metrics for Machine Learning Models. *Radiology Artificial Intelligence*, 3(3), e200126.
<https://doi.org/10.1148/ryai.2021200126>.
2. Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, 63(2), 183-193. <https://doi.org/10.1016/j.bushor.2019.11.003>

3. IEEE. (2021, January 22-24). *Proceedings of the 2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*. Shenyang, China.
<https://doi.org/10.1109/ACCTCS52002.2021.00077>
4. Kotane, I., & Kuzmina-Merlino, I. (2011). *Assessment of financial indicators for evaluation of business performance*. ResearchGate.
https://www.researchgate.net/publication/268352828_Assessment_of_financial_indicators_for_evaluation_of_business_performance