

Título: MACHINE LEARNING MEETS ASSET MANAGEMENT (APLICACIONES DE APRENDIZAJE AUTOMÁTICO EN LA GESTIÓN DE ACTIVOS FINANCIEROS)

Nombre: MIRETE FERRER, PEDRO MANUEL

Universidad: Universidad de Murcia

Departamento: Escuela Internacional de Doctorado de la Universidad de Murcia

Fecha de lectura: 13/12/2023

Programa de doctorado: Programa de Doctorado en Economía por la Universidad de Alicante; la Universidad de Murcia; la Universidad Miguel Hernández de Elche; la Universidad Nacional de Educación a Distancia y la Universidad Politécnica de Cartagena

Dirección:

- > **Director:** MARÍA ASUNCIÓN PRATS ALBENTOSA
- > **Director:** JUAN SAMUEL BAIXAULI SOLER

Tribunal:

- > **presidente:** SUSANA ÁLVAREZ DIEZ
- > **secretario:** Luis Moreno Izquierdo
- > **vocal:** JOSÉ GARCÍA RODRÍGUEZ

Descriptores:

- > GESTION FINANCIERA
- > INVESTIGACION OPERATIVA EN LA EMPRESA
- > DIRECCION DE VENTAS

El fichero de tesis ya ha sido incorporado al sistema

- > <http://hdl.handle.net/10201/136652>

Resumen: RESUMEN

El propósito de esta tesis es proporcionar una revisión exhaustiva, desde una perspectiva tanto teórica como práctica, de la utilidad que el Machine Learning (ML) aporta a la disciplina financiera de la Gestión de Activos. En años recientes, el mundo de la Gestión de Activos ha presenciado una transformación notable con la integración de tecnologías de vanguardia. Entre estos avances, la aplicación del ML destaca como una herramienta que ha transformado la toma de decisiones de inversión, capacitando a los gestores de activos para tomar decisiones de inversión más informadas basadas en datos. La tesis realiza contribuciones en dos áreas principales. En primer lugar, proporciona una revisión global exhaustiva del estado del arte en cuanto a la contribución del ML a la Gestión de Activos. Buscando un equilibrio entre Finanzas, Estadística y Computación, la revisión intenta mejorar el análisis de la literatura reciente, a fin de que investigadores y profesionales puedan explorar sin fisuras sus áreas de interés. En segundo lugar, la tesis se centra en la aplicación práctica de herramientas de ML en la gestión de activos, buscando evidencia empírica de su utilidad para mejorar los

resultados obtenidos por enfoques tradicionales como la econometría. La primera sección empírica aborda el problema más general dentro de la disciplina de Gestión de Activos, la predicción de precios. A pesar de las limitaciones que la hipótesis de Eficiencia impone en cuanto a la predicción de precios de activos financieros, aplicamos los algoritmos de ML más innovadores, específicamente las Gated Recurrent Units (GRUs), a los últimos 33 años de precios diarios del índice SP 500, previamente procesados con un innovador algoritmo de descomposición de señales para series no estacionarias (EMD). Los resultados son prometedores, ya que nuestro modelo EMD-GRU, de acuerdo a los estándares estadísticos, muestra un nivel notable de precisión en la predicción de movimientos direccionales de precios de acciones. La Directional Accuracy (DA) asciende al 75,15% en el caso del mejor modelo cuando se prueba en escenarios reales de trading. En el segundo ejercicio empírico, la tesis aborda otra estrategia de predicción de precios muy conocida entre los gestores de activos, llamada estrategia de pares o arbitraje estadístico. El experimento examina la viabilidad de utilizar técnicas de aprendizaje no supervisado de ML para encontrar posibles pares de acciones que incorporar a una cartera long-short. Además de usar autoencoders para la reducción de dimensionalidad, proponemos aplicar un enfoque específico de clustering para datos de series temporales, llamado k-means clustering con DTW. Utilizando datos mensuales desde 2010 para los componentes individuales del índice SP500, el modelo propuesto logra un rendimiento total del 89% en 8 años, con una DA del 63% y un ratio de Sharpe (SR) de 2,658. En comparación con los dos benchmarks seleccionados, el índice y una estrategia short-term reversal, nuestro modelo es el mejor en términos de rendimiento total y otras medidas de rendimiento. Finalmente, el tercer ejercicio se centra en los modelos de tres y cinco factores de Fama y French, quizás los modelos de valoración de activos más seguidos por la comunidad científica en los últimos años. En este caso, la propuesta es más metodológica que empírica, ya que sugerimos una estimación alternativa de estos modelos a través del uso de algoritmos de ML. Para ser más específicos, se aplica una combinación de Support Vector Regression (SVR) y Grid Search (GS) con Validación Cruzada (CV) de 10 capas para pronosticar rendimientos mensuales en cinco industrias del índice SP500, utilizando los modelos de tres y cinco factores de Fama y French. El modelo propuesto GS-CV-SVR arroja coeficientes de correlación fuera de muestra entre 92-97% para los diferentes sectores. Estos hallazgos muestran que la ganancia de precisión en la predicción es notable en comparación con el método de optimización OLS convencional utilizado en el trabajo original de Fama-French y en trabajos posteriores.

The purpose of this thesis is to provide a comprehensive review, from both theoretical and practical perspectives, of the usefulness that Machine Learning (ML) brings to the financial discipline of Asset Management. In recent years, the world of Asset Management has witnessed a remarkable transformation with the integration of cutting-edge technologies. Among these advancements, the application of ML stands out as a revolutionary tool that has revolutionized how investment decisions are made, empowering asset managers to make more informed and data-driven investment choices. The thesis makes significant contributions in two main areas. Firstly, it provides a comprehensive global review of the state of the art regarding the contribution of ML to Asset Management. By striking a balance between Finance, Statistics, and computational fields, our review enhances the analysis of recent literature, ensuring that researchers and practitioners can explore their areas of interest without encountering gaps. Secondly, the thesis focuses on the practical application of ML tools to typical asset management topics. It seeks to find empirical evidence of the usefulness of ML techniques in enhancing the results of traditional econometrics approaches. The first empirical section addresses the most general problem within the discipline of Asset Management, price forecasting. Despite the limitations that the EMH imposes on financial markets regarding the prediction of financial assets prices, we apply the most innovative ML algorithms, specifically Gated Recurrent Units (GRUs), to the last 33 years of SP 500 daily prices, previously preprocessed with an innovative decomposition signal algorithm for non-stationary series (EMD). The

results are quite promising, as our model EMD-IGRU, according to the statistical standards, exhibits a reasonable level of accuracy in predicting stock prices and directional movements. The directional accuracy (DA) rises to 75.15% in the case of the best model when is tested in real-world scenarios. In the second empirical exercise, the thesis faces another price forecasting topic, very well-known between asset managers, called pairs trading strategy or statistical arbitrage. The second experiment of the thesis looks at the viability of using unsupervised learning techniques of ML to find possible stock pairs for a long-short portfolio. Besides using autoencoders for dimensionality reduction, we propose to apply a specific clustering approach to time series data, called k-means clustering with DTW. As an additional contribution, we suggest using financial data (firm characteristics) as additional features in the pairs selection. Using monthly data since 2010 for the individual constituents of the US index SP500, the proposed model achieves a total return of 89% in 8 years, with a DA of 63%, a Sharpe ratio (SR) of 2.658, and a Calmar ratio of 0.678. In comparison with the two benchmarks selected, the market and the short-term reversal strategy, our model is the best in terms of total return, but also in terms of the different trading performance measures, mainly SR figures. Finally, the third exercise is focused on the three- and five-factor models from Fama and French, maybe the most reproduced asset pricing models in recent years. In this case, the proposal is more methodological than empirical, because we suggest an alternative estimation of these models through the use of ML algorithms. To be more specific, we apply a combination of Support Vector Regression (SVR) and mixed Grid Search (GS) with 10-fold Cross Validation (CV) optimization to forecast monthly returns in five industries of US S&P500 using the three and five factors presented by Fama and French. The proposed hybrid GS-CV-SVR model produce out-of-sample (testing data sets) correlation coefficients between 92-97% for the different sectors. These findings show that the prediction accuracy gain is quite notable when compared to the conventional OLS optimization method used in the original work of Fama-French and the traditional approach of the literature globally inspired in them.