



MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MÁSTER

APLICACIÓN DE LOS AVANCES DE APRENDIZAJE AUTOMÁTICO PARA INVERSIÓN EN MERCADOS FINANCIEROS

Autor: Borja de Olazabal Bernaldo de Quirós

Co-Director: José Portela González

Co-Director: Juan Luis Zamora Macho

Madrid

Julio de 2020

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
*Aplicación de los avances de aprendizaje automático para inversión en
mercados financieros*

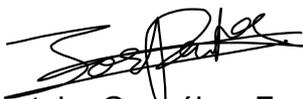
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2019-2020 es de mi autoría, original e inédito y
no ha sido presentado con anterioridad a otros efectos. El Proyecto no es
plagio de otro, ni total ni parcialmente y la información que ha sido tomada
de otros documentos está debidamente referenciada.



Fdo.: Borja de Olazabal Bernaldo de Quirós Fecha: 04/07/2020

Autorizada la entrega del proyecto

LOS DIRECTORES DEL PROYECTO



Fdo.: José Portela González Fecha: 04/07/2020



Fdo.: Juan Luis Zamora Macho Fecha: 04/07/2020



MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MÁSTER APLICACIÓN DE LOS AVANCES DE APRENDIZAJE AUTOMÁTICO PARA INVERSIÓN EN MERCADOS FINANCIEROS

Autor: Borja de Olazabal Bernaldo de Quirós

Co-Director: José Portela González

Co-Director: Juan Luis Zamora Macho

Madrid

Julio de 2020

Resumen Ejecutivo

Introducción y Estado del Arte

En este resumen se va a seguir la estructura utilizada en el trabajo. Primero se presentarán las principales tendencias relacionadas con las técnicas utilizadas. Se seguirá con la metodología propuesta. Después se presentará el análisis exploratorio de los datos, consistente en inspección visual de las principales correlaciones, análisis de componentes principales y clusterización. Estas dos últimas son técnicas de aprendizaje automático no supervisado. Por último se presentarán los resultados de la metodología aplicada a una estrategia de inversión y las conclusiones y futuros desarrollos.

Los mercados bursátiles han sido durante mucho tiempo objeto de interés para el análisis de datos, especialmente para tratar de obtener mayores beneficios. En los últimos años se han desarrollado especialmente el uso de redes neuronales para predecir el comportamiento de los valores bursátiles. En concreto, ha habido grandes desarrollos en el uso de redes neuronales recurrentes y con memoria al corto y largo plazo que consiguen predecir de cierta forma el precio que tendrá un valor en el futuro cercano, así como mejorar la gestión de carteras. Por otro lado, también se han desarrollado en los últimos años el uso de características de las series temporales para clasificar e identificar patrones en series temporales de distinta naturaleza. En este contexto, se propone el cálculo y uso de características de series temporales antes de realizar este tipo de predicciones.

Metodología Propuesta

Los datos, compuestos por más de 1200 series de precios de cierre diarios entre 2009 y 2019, de distintas fuentes (Yahoo! Finance y Bloomberg) se prestan a realizar este tipo de análisis. Dado que existen tantas compañías en las que invertir, la decisión de invertir o no en cada una de éstas puede necesitar de grandes y profundos estudios, por ello, resulta interesante desarrollar este tipo de metodologías para ayudar en la inversión. Con este objetivo, en este trabajo se realiza el cálculo de múltiples características relacionadas con la descomposición STL, funciones de auto-correlación o auto-correlación parcial, raíces unitarias, tests de Portmanteau y otros muchas más. Con este método, se reduce notablemente el número de datos con los que trabajar. Un ejemplo de este tipo de transformación se muestra en el figura 1.

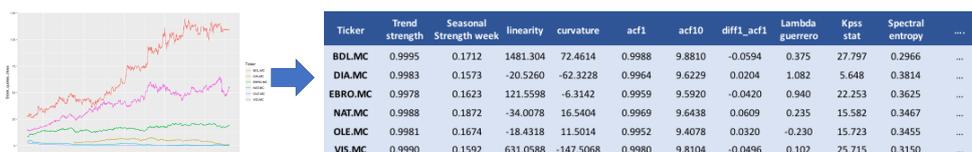


Figura 1: Esquema de la transformación realizada

Análisis Exploratorio

Una vez calculadas todas las características, las series son representadas por un punto. Sobre estos datos se pueden aplicar técnicas de análisis exploratorio de los datos y de aprendizaje no supervisado. En términos de análisis exploratorio se han encontrado una serie de relaciones a través de la visualización de los datos. Estas relaciones concluyen que el mercado inglés es el que obtiene valores más extremos de linealidad y curvatura, lo cual resulta en que estas compañías reciban un análisis distinto al del resto de compañías. También otras características como la función de auto-correlación de la serie diferenciada permite encontrar ciertas anomalías en los datos, que pueden ser síntomas de falta de liquidez o de insuficiencia de datos. La representación de los datos que han llevado a estas conclusiones se muestran en la figura 2.

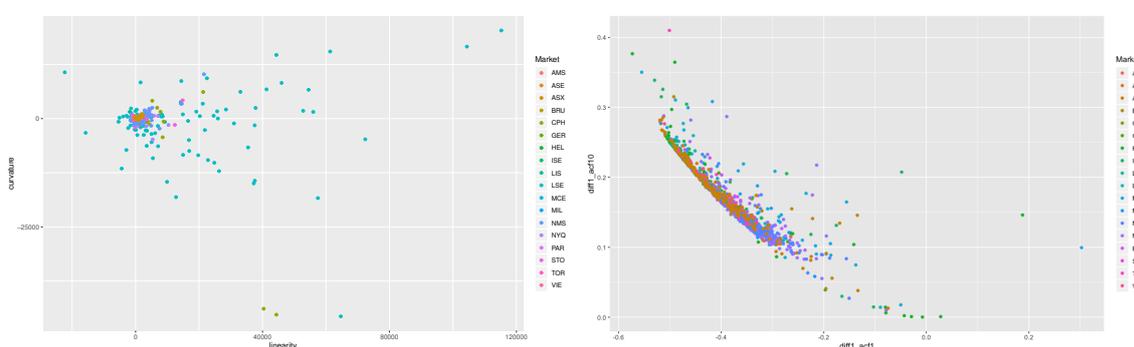


Figura 2: Algunas de las relaciones encontradas a través de la visualización de datos

Siguiendo este análisis y siendo inviable en este trabajo analizar las más de 800 relaciones entre variables, se realiza un análisis de componentes principales. De este análisis se obtiene que algunas características representan relaciones similares dentro de los datos. Este es el caso de la fuerza de la tendencia y el estadístico de Phillips Perron. Esto deriva en que estas características tengan una fuerte representación en la primera componente, y que en las series de los percentiles más bajos de esta componente se aprecien fuertes tendencias. La segunda componente, en cambio, tiene fuertes relaciones con las funciones de auto-correlación de las series diferenciadas una y dos veces. Este tipo de características resultan difíciles de apreciar a simple vista, incluso conociendo su existencia.

Además se han identificado una serie de características que explican poco la varianza de los datos. Algunas de éstas están relacionadas con la estacionalidad, lo cual quiere decir que esta característica debería o bien cambiarse (ampliando el periodo de estacionalidad) o eliminarse. La aportación acumulada de cada característica a las primeras cinco componentes se muestra en la figura 3. Por último se muestra, también, la explicación de la varianza de cada componente y la varianza explicada acumulada de la misma. Entre las 4 primeras componentes se explica más del 50% de la varianza, y entre las 11 primeras explican el 80%.

El siguiente paso que sigue al análisis es el de realizar una clusterización de las características. Esta clusterización está muy marcada por los grandes valores de las características de linealidad y curvatura. Se realiza, por tanto, una clusterización con los valores

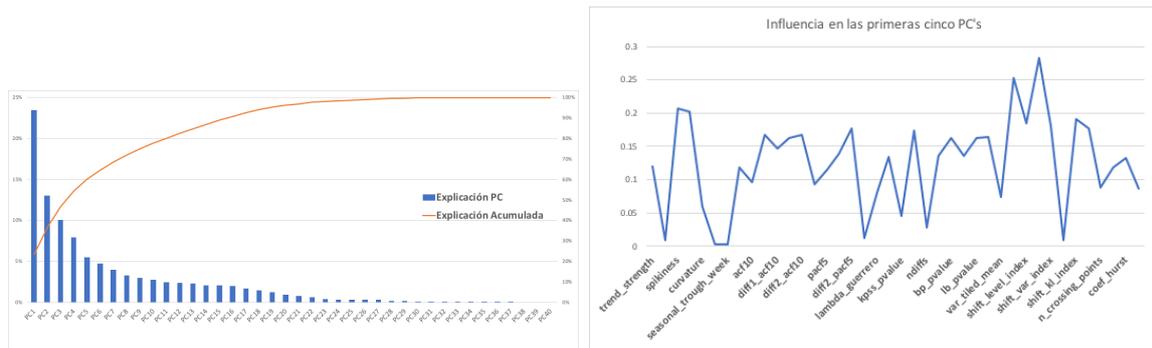


Figura 3: Explicación de la varianza e influencia de las componentes principales

normalizados. Con esta clusterización se identifican 9 clústers, los cuales, al evaluar la rentabilidad que dan a lo largo de los 10 años (siguiendo todos la misma estrategia), presentan ciertas diferencias en su rentabilidad. En la figura 4 se representan los clústers sobre las dos primeras componentes principales, así como la distribución de la rentabilidad que dan las compañías de cada clúster a través de un diagrama de cajas y bigotes.

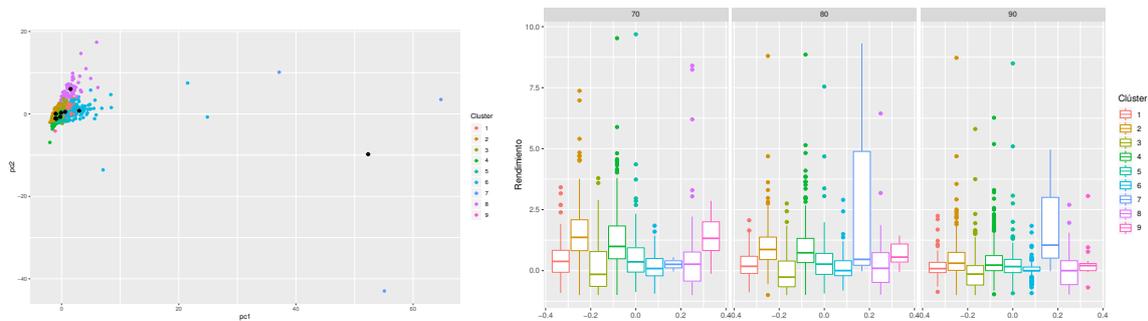


Figura 4: Representación de los clústers sobre sus CP y sus rentabilidades

Modelo de Inversión

En cuanto al aprendizaje supervisado, se ha realizado sobre estrategias a un año basadas en el índice de fuerza relativa (RSI) [1]. Para ello, se transforman las series de valor de cierre de forma que indiquen cuándo una acción está infra-comprada o sobre-vendida. Definiendo unos umbrales se obtiene un modelo que da lugar a señales de compra y venta que, al realizarse, suelen dar mayores o menores beneficios. Ilustrando esta estrategia se muestra en la figura 5 un ejemplo de RSI con sus umbrales, con las señales de venta (verde) y compra (rojo), así como estas señales sobre la evolución del precio de cierre.

Estas series RSI se dividen por años y se optimizan los umbrales de compra y venta de forma que se maximiza la rentabilidad del valor para el año analizado. Sobre estas series (anuales) del índice de fuerza relativa se aplica el cálculo de características. También se calcula la rentabilidad económica que habría obtenido la estrategia óptima del año analizado sobre el año siguiente, por ejemplo, los umbrales óptimos de 2014 aplicados sobre 2015.

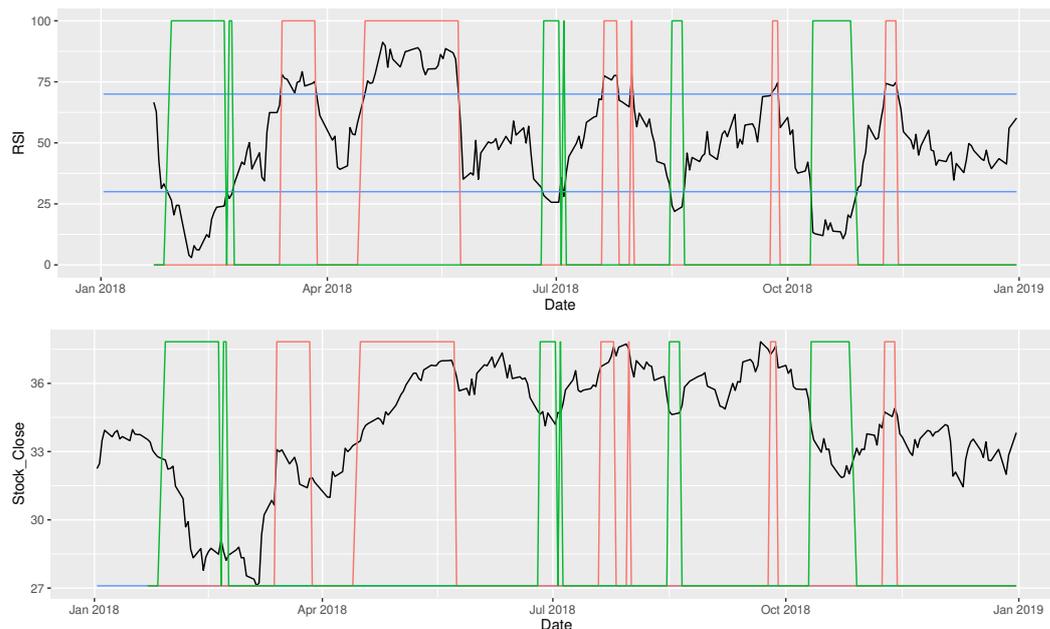


Figura 5: Ejemplo de estrategia RSI

El objetivo que se persigue es el de predecir el rentabilidad que se obtendría en el año siguiente manteniendo la estrategia a partir de la información que se tiene: datos de la estrategia y características de la serie. El conjunto de entradas sería, por tanto, el conjunto de las características y las métricas de la estrategia del año anterior, mientras que la salida sería la rentabilidad de dicha estrategia en el año siguiente. Una vez predicha la rentabilidad, se tomará como criterio de decisión invertir en aquellas empresas que se prevean que den una rentabilidad igual o superior a un cierto índice.

Se desarrollan tres tipos de predicciones. La primera consiste en suponer que la rentabilidad será la misma que el año anterior. Esta estrategia será la base contra la que comparar. El segundo consiste en una regresión lineal. El tercero consiste en una red neuronal. De este último tipo se analizan varios tamaños de redes, de entre 5 y 20 neuronas, siendo el de 15 neuronas en la capa intermedia el óptimo. En la figura 6 se muestran las rentabilidades de las distintas opciones y según la mínima rentabilidad esperada escogida.

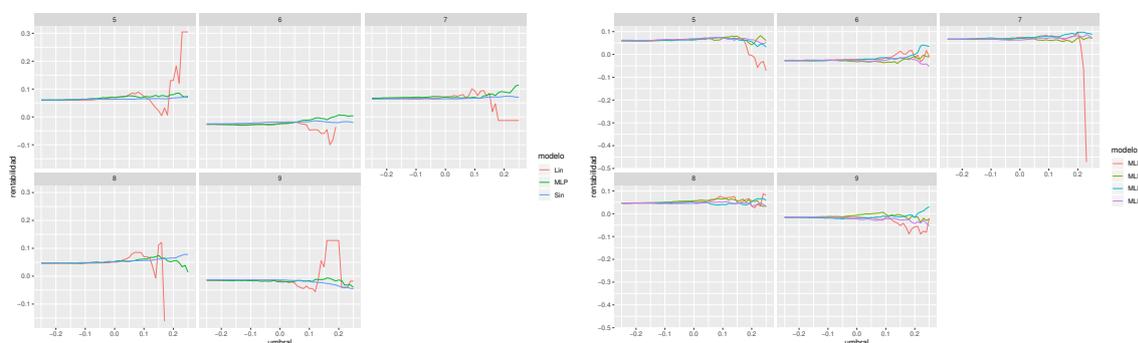


Figura 6: Comparativa de las rentabilidades obtenidas por los distintos modelos

Además, se han desarrollado dos estrategias a la hora de tratar las predicciones. La

primera, y la que se ha aplicado en el gráfico anterior, es la de seleccionar un umbral mínimo e invertir en todas las compañías cuyas rentabilidades esperadas sean superiores a dicho umbral. La segunda consiste en invertir solo en las empresas en un espacio de $\pm 2,5\%$ sobre el umbral escogido. Estas estrategias se muestran en la figura 7, siendo la primera estrategia la de la izquierda.

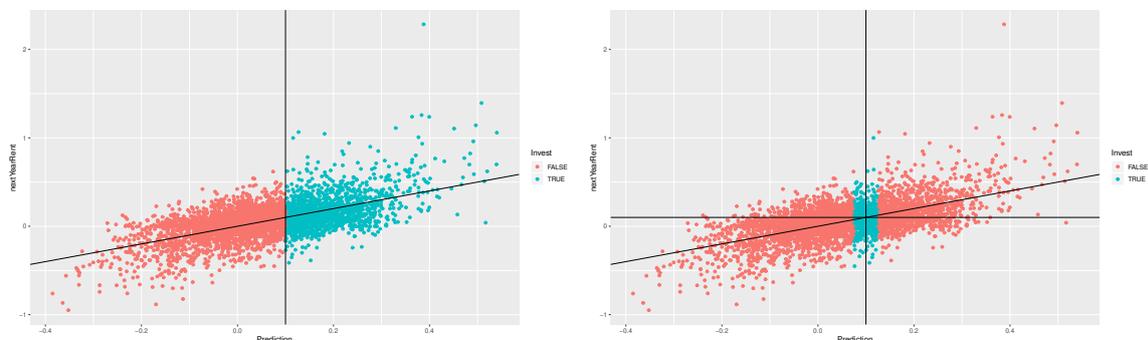


Figura 7: Representación gráfica de las dos estrategias probadas

Con la segunda estrategia se realizan las mismas pruebas, solo siendo más eficaz que la original en los casos donde el umbral mínimo sea superior o igual al 20%. Es por ello que se sugiere una combinación de ambas estrategias, aplicando la primera para umbrales inferiores al 20% y la segunda para los restantes.

Conclusiones y Futuros Desarrollos

Con estos resultados se concluye que la metodología resulta en mayores beneficios a la hora de invertir. Por tanto, se puede afirmar que se ha conseguido el objetivo de usar la selección de características de las series temporales para ayudar en la toma de decisiones a la hora de invertir en el mercado bursátil.

En cuanto a futuros desarrollos se propone el uso de esta metodología pero usando las características de la serie original, sin procesar; utilizar otras estrategias de inversión, tales como la convergencia-divergencia de la media móvil (MACD) [2]; aumentar la frecuencia del análisis; y utilizar las redes neuronales para ajustar los parámetros de la estrategia de inversión para el año siguiente.

Bibliografía

- [1] James Chen. *Relative Strength Index (RSI)*. Investopedia, 2020.
- [2] Adam Hayes. *Moving Average Convergence Divergence–MACD*. Investopedia, 2020.

Executive Summary

Introduction and State of the Art

In this summary the thesis's structure is going to be used. In the first place, the main trends that are related with the used techniques will be presented. The proposed methodology will follow. Afterwards, the exploratory analysis will be presented, composed of a visual inspection of the main correlations, a principal component analysis and clustering. These last two are unsupervised learning techniques for machine learning. Finally, the results of applying the methodology to an investment strategy will be presented, followed by the conclusions and future developments.

Stock markets have been, for many years object of interest for data analysis, specially when trying to acquire greater financial benefits. In the last years the use of neural networks for the forecasting of stock price behavior has been specially developed. More precisely, advances on recurrent and long-short term memory networks have allowed, to some extent, to forecast the price of certain stocks in the near future, and to enhance the portfolio management. Additionally, also in the last years the use of time series features for classification and pattern identification has been developed in many fields. In this context, the calculation and use of these time series features is proposed before performing any forecasting.

Proposed Methodology

The data, composed of over 1200 series of daily stock prices between 2009 and 2019, from different sources (Yahoo! Finance and Bloomberg) are open to perform these kind of analysis. Since there are many companies in which we can invest, the decision of investing or not in each of these companies can require great and profound studies. Therefore, it is interesting to develop these kind of methodologies to help in these decisions. With this objective, in this thesis many features related with STL decomposition, auto-correlation functions, partial auto-correlations, unit roots, Portmanteau tests and many others are performed. With this method, the number of data with which to work is notably reduced. An example of this methodology is shown in figure 1.



Figure 1: Diagram of the performed transformation

Exploratory Analysis

Once all features are calculated, the time series are represented by a single point. It is on these data where the exploratory analysis techniques and unsupervised learning will be performed. Related with the exploratory analysis, some relations have been found though data visualization. These relations conclude on the British market having the more extreme values relating with linearity and curvature, which indicates that these companies should have different analyses from the other companies. Other features like the auto-correlation function of the differentiated time series allows to find data anomalies, which may be symptoms of lack of liquidity of the stock or insufficient data. The data that have led to these insights are shown in figure 2.

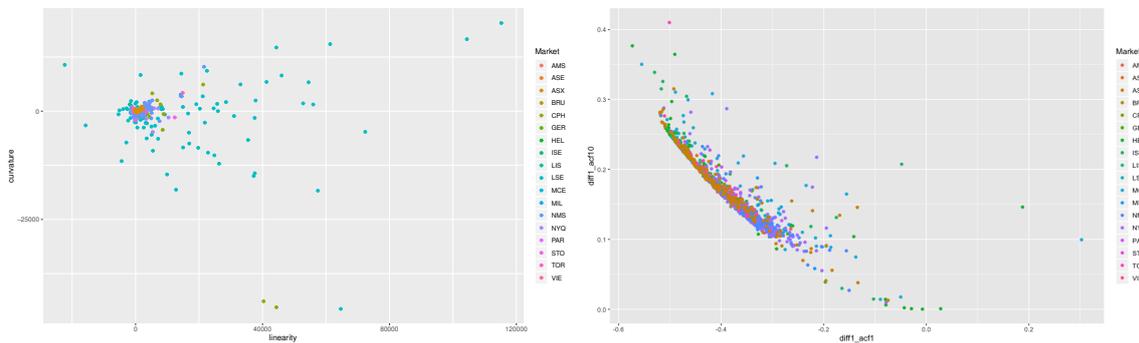


Figure 2: Some of the insights found though data visualization

Following the analysis and being unfeasible in this thesis to analyze the resulting over 800 relationships between the features, a principal component analysis is performed. From this analysis we obtain that some features represent similar relationships among the data. This is the case of the trend strength and the Phillips Perron statistic. This derives that these features have a strong representation in the first component and that the series in the lowest percentile of this component show strong trends. However, the second component has strong relations between the auto-correlation functions of the series differentiated once and twice. This kind of features are very difficult to appreciate at first sight, despite knowing of its existence.

Furthermore, some features that explain little variance have been identified. Some of these features are related with the seasonality, which means that these feature should either be changed (by increasing the seasonality period) or eliminated. The accumulated contribution of each feature of the first five components is shown on figure 3. Lastly, the variance explanation of each component and the accumulated explanation of the previous components is shown in figure 3. Between the first 4 components 50% of the variance is explained, while between the first eleven explain 80% of the variance.

The next step in the analysis is to perform a clustering of the features. This clustering is strongly marked by the extreme values of linearity and curvature. For this reason, the clustering is done with the normalized values of the features. With this clustering, 9 different clusters are found, among which, after evaluating the profitability we would have obtained along the last 10 years, we find noticeable differences in the profitability

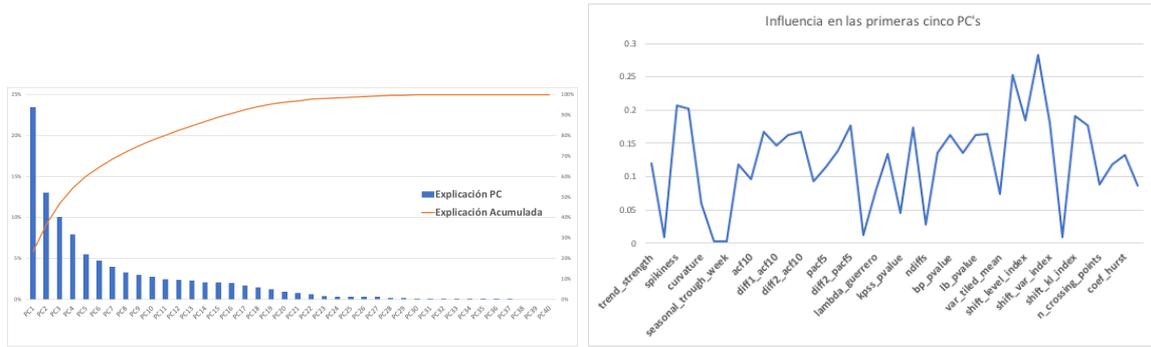


Figure 3: Variance explanation and influence of the principal components

by cluster. O figure 4, the clusters are represented on the principal components, as well as the profitability from the companies in each cluster in the form of a box and whiskers plot.

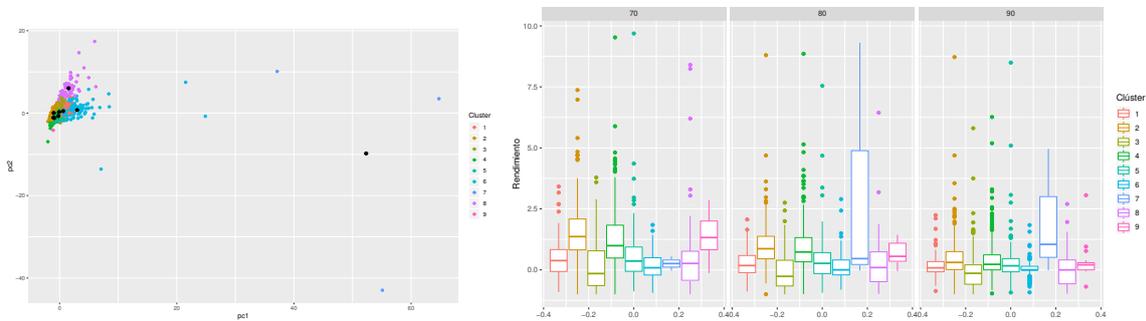


Figure 4: Representation of the clusters on the PC and its profitability

Investment model

Regarding supervised learning, it has been performed over strategies based on the relative strength index (RSI) [1] in each year. For this, the series of the closing stock quotes are transformed so that we obtain when a stock is over or under sold. By defining some thresholds we obtain a model that gives buying and selling signals, that, when performed, tend to give benefits. Illustrating this strategy, figure 5 is shown, where an example of RSI with its thresholds is presented, with the selling (green) and buying (red) signals, as well as these signals over the original evolution of the stock price.

These RSI series are divided in years and the buying and selling thresholds are optimized so that the profitability is maximized for a certain stock for the analyzed year. Over these annual RSI series is where we apply the new feature calculation. The profitability that applying the past year's strategy in the following year is calculated. For example, we apply the optimized thresholds of 2014 to the 2015 series to estimate the profitability in 2015.

The pursued objective is to forecast the profitability that would be obtained next year by maintaining the optimal strategy of the past year on the basis of the information

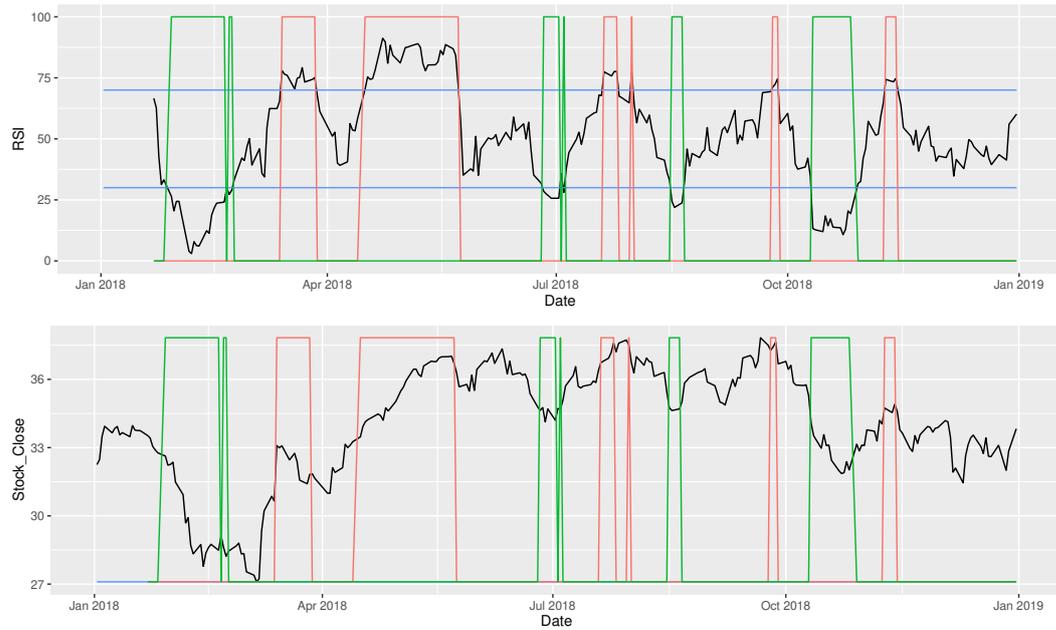


Figure 5: Example of an RSI strategy

we can extract from this last year: optimal strategy data and series features. The input would, therefore, be the group of the features and strategy metrics of last years RSI, while the output would be the profitability of the same strategy next year. Once the profitability is forecasted, a criteria to decide in which stocks to invest is taken. As a first approximation, we will use the criteria of investing in all stocks whose profitability is higher than a certain value.

Three types of forecasting models are developed. The first model consists on supposing that next year's profitability will be the same that the previous year. This will be the base to which we can compare the other models. The second model consists on a linear regression. The third one consists on a neural network. Many sizes have been analyzed, between 5 and 20 neurons, being 15 the optimal size. On figure 6 we can see the profitability of the different options according to the minimum expected profitably.

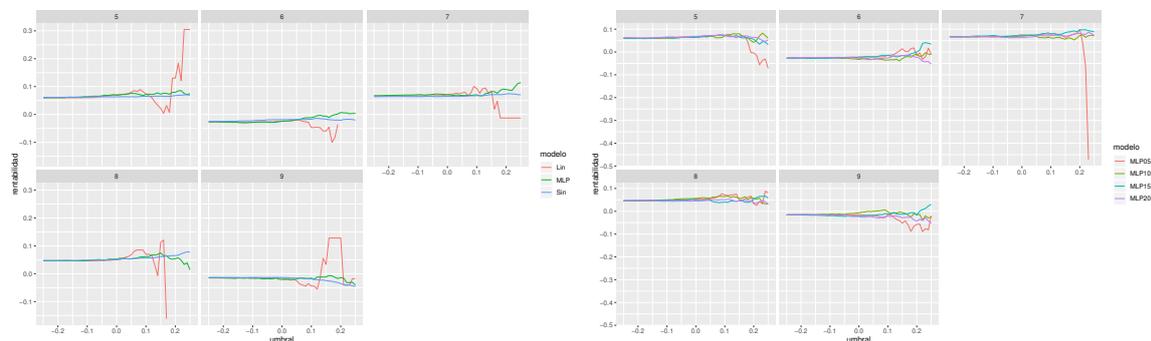


Figure 6: Profitability comparison of the different models

Furthermore, two strategies have been developed when addressing the predictions. The first, and the one that was applied previously, consists on selecting a minimum pro-

profitability threshold and investing in all companies whose expected profitability is higher than that threshold. The second one consists on investing on companies in a range between $\pm 2,5\%$ over the selected threshold. These strategies are graphically presented on figure 7, being the first strategy the one on the left.

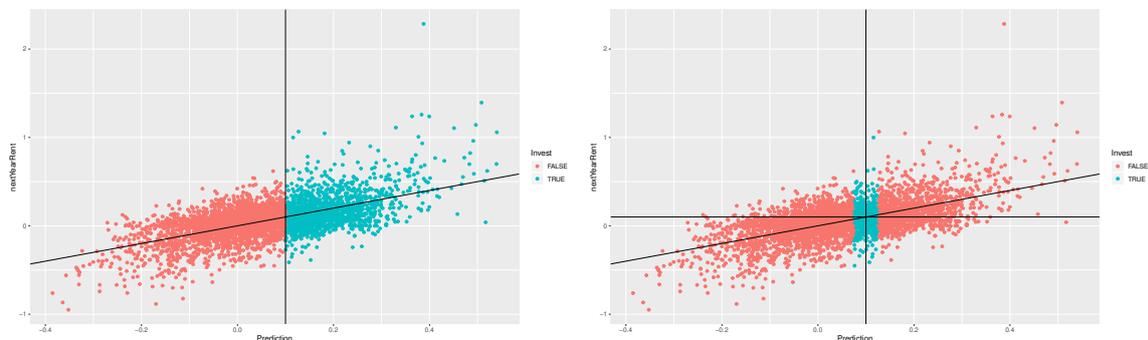


Figure 7: Graphic representation of the two proposed strategies

The same tests are performed with the second strategy. The results are that this strategy is that the original strategy results in higher profits for the cases where the selected threshold is lower than 20%. For this reason, the combination of both models is proposed, applying the original for thresholds lower than 20% and the second one for the remaining thresholds.

Conclusions y Future Developments

With these results we can conclude that the methodology results on higher profits when taking investment decisions. Therefore, we can confirm that we have achieved the pursued objective of using feature selection on stock price time series to help in investment decision.

In terms of future developments, many are proposed: the use of this methodology with the features of the original time series, before calculating the RSI; the use of other investment strategy, such as the moving average convergence-divergence (MACD) [2]; increasing the frequency of the analysis; and the use of the neural networks to adjust the parameters of the investment strategy for the following year.

Bibliography

- [1] James Chen. *Relative Strength Index (RSI)*. Investopedia, 2020.
- [2] Adam Hayes. *Moving Average Convergence Divergence–MACD*. Investopedia, 2020.



TRABAJO DE FIN DE MÁSTER

Aplicación de los avances de aprendizaje automático
para inversión en mercados financieros

MEMORIA FINAL

Autor:

Borja DE OLAZABAL BERNALDO DE QUIRÓS

Directores:

José PORTELA GONZÁLEZ
Juan Luis ZAMORA MACHO

Julio 2020

Dedicado a Boli, a mis padres y a JAO

Agradecimientos

En primer lugar querría agradecer a mis directores, José y Juan Luis, por su dedicación y guía en todo este tiempo, por su paciencia y por el trato recibido. También me gustaría agradecer toda la formación recibida en la Universidad Pontificia de Comillas durante estos seis años, donde he tenido la oportunidad de tener profesores, compañeros y amigos que me han enseñado, ayudado y transmitido el espíritu del ICAI y me han hecho sentir como en casa.

No puedo dejar de agradecer a mis padres y a Boli por su apoyo incondicional, por el ejemplo que siempre me han dado y por los cuales he tenido esta gran oportunidad. También querría agradecer a mis hermanos por su ayuda y apoyo durante estos años.

Por último, querría agradecer a mis amigos que me han apoyado y aguantado durante todos estos años, especialmente a Gonzalo, con el cual he compartido todos estos años y que tanto me ha ayudado y apoyado a lo largo de toda la carrera.

A.M.D.G.

Resumen

El aprendizaje automático ha sido utilizado en numerosas aplicaciones, incluida la predicción del precio de las acciones. El objetivo de este trabajo es el de combinar el aprendizaje automático con las técnicas de selección de características para series temporales en el mercado de valores de forma que ayude a distinguir y clasificar las compañías así como en la decisión sobre en qué valores invertir, consiguiendo mayores rentabilidades.

Con este objetivo, se ha aplicado a la selección de características una serie de técnicas de aprendizaje no supervisado, como el análisis de componentes principales o la clusterización, para clasificar las empresas que puedan ser más rentables a la hora de invertir. Gracias a este aprendizaje se consigue entender mejor las dinámicas internas que gobiernan este tipo de series.

También se han transformado las series originales para aplicar estrategias de inversión y, a través de distintos modelos de predicción, tratar de predecir la rentabilidad que pueden dar ciertas estrategias de inversión al medio plazo sobre distintas compañías. Con esta predicción, se proponen distintos criterios a la hora de elegir las compañías en las que invertir. Por último, se comparan los beneficios obtenidos al aplicar las distintas combinaciones de modelo de predicción y criterios de elección a lo largo de los años.

Abstract

Machine learning has been used in numerous applications, included the prediction of the stock prices. The objective of this thesis is to combine the machine learning with the time series feature selection techniques for the stock market in a way that helps distinguishing and classifying the different companies as well as in the decision on which companies to invest, obtaining higher profitability.

Following this objective, unsupervised learning techniques have been applied, such as principal component analysis or clustering, to these selected features in order to classify those companies. This classification may distinguish the most profitable for the investors. Thanks to this kind of learning, we can achieve a better understanding of the internal dynamics that govern this kind of series.

Furthermore, the time series have been transformed in order to apply investment strategies and, through different forecasting models, try to forecast the profitability that certain investing strategies term give in the medium term over different companies. With this profitability forecast, different criteria are proposed when selecting the companies in which to invest. Finally, the different combinations of prediction model and selecting criteria along the years are bench-marked using the profitability that we would have obtained if the strategy had been followed.

Índice

1. Introducción	1
2. Estado del Arte	5
3. Metodología Propuesta	9
4. Análisis Exploratorio de los Datos	15
4.1. Métricas principales	15
4.2. Análisis de componentes principales	25
4.3. Clusterización	32
4.4. Conclusiones	41
5. Modelo de inversión	43
5.1. Cálculo y uso del RSI	43
5.2. Aplicación de la metodología a la estrategia RSI	46
5.3. Conclusiones	58
6. Conclusiones y Futuros Desarrollos	59
Referencias	61
A. Alineación del Proyecto con los ODS	63

Índice de figuras

1.	Tipos de aprendizaje automático y algoritmos utilizados	2
2.	Principales tipos de redes neuronales y de neuronas	2
3.	Distribución de los tipos de empresas que gestionan acciones en el mercado americano	3
4.	Distribución del comercio de acciones del mercado americano según el tipo de compañía que los gestionan	3
5.	Estructura de una BLSTM	7
6.	Evolución de los valores españoles a lo largo de 10 años	10
7.	Esquema de la transformación realizada	14
8.	Distribución de las empresas según mercado y sector	15
9.	Fuerza de la tendencia y la estacionalidad para el volumen de acciones	16
10.	Fuerza de la tendencia y la estacionalidad de los indicadores	17
11.	Fuerza de la tendencia y la estacionalidad del valor de apertura	17
12.	Valor correspondiente a percentil 10 en trend_strength	18
13.	Valor correspondiente a percentil 50 en trend_strength	18
14.	Valor correspondiente a percentil 90 en trend_strength	19
15.	Valor correspondiente a percentil 10 en seasonal_strength_week	19
16.	Valor correspondiente a percentil 50 en trend_strength	19
17.	Valor correspondiente a percentil 90 en trend_strength	20
18.	Matriz de correlación de las principales características STL	21
19.	Relación entre curvatura y linealidad según los mercados	21
20.	Comparativa de las variables diff1_acf y diff10_acf	22
21.	Compañías con diff_acf1 positivo (outliers)	23
22.	Matriz de correlación completa	24
23.	Valores proyectados sobre las componentes principales	25
24.	Explicación de la varianza de cada PC y acumulada	26

25.	Influencia de cada indicador en las primeras dos PCs	26
26.	Influencia acumulada de cada indicador en las primeras cinco PCs	27
27.	Gráfico biplot	27
28.	Compañía correspondiente al percentil 1 de la primera PC	28
29.	Compañía correspondiente al percentil 50 de la primera PC	28
30.	Compañía correspondiente al percentil 99 de la primera PC	28
31.	Compañía correspondiente al percentil 1 de la segunda PC	29
32.	Compañía correspondiente al percentil 50 de la segunda PC	29
33.	Compañía correspondiente al percentil 99 de la segunda PC	29
34.	Compañía correspondiente al percentil 35 de la segunda PC	30
35.	Compañía correspondiente al percentil 95 de la segunda PC	31
36.	Compañía diferenciada correspondiente al percentil 1 de la segunda PC .	31
37.	Compañía diferenciada correspondiente al percentil 50 de la segunda PC	31
38.	Compañía diferenciada correspondiente al percentil 95 de la segunda PC	31
39.	Compañía doblemente diferenciada correspondiente al percentil 1 de la segunda PC	32
40.	Compañía doblemente diferenciada correspondiente al percentil 50 de la segunda PC	32
41.	Compañía doblemente diferenciada correspondiente al percentil 95 de la segunda PC	32
42.	Clustering de las series mostrado sobre las componentes principales . . .	33
43.	Método del codo para definir el número de clústers	33
44.	Número de compañías dentro de cada clúster	34
45.	Coordenadas de los centroides	34
46.	Evolución de los valores de apertura de las compañías de los clústers 2 y 3	36
47.	Nueva clusterización para el valor de apertura	36
48.	Centroides de la nueva clusterización para el valor de apertura	37
49.	Clústers normalizados	38

50.	Centroides de los clústers normalizados	38
51.	Muestra de curvas del clúster 6	39
52.	Muestra de curvas del clúster 7	39
53.	Muestra de curvas del clúster 8	39
54.	Muestra de curvas del clúster 9	39
55.	Rendimientos de cada clúster	40
56.	Valores de cierre de ACS	44
57.	Cambios al alza y baja de ACS	44
58.	Medias móviles de los cambios al alza y baja de ACS	44
59.	Fuerza relativa de ACS	45
60.	Índice de fuerza relativa de ACS	45
61.	RSI de ACS con sus límites	45
62.	Señales de compra y venta para el RSI	46
63.	Señales de compra y venta con la evolución del precio de cierre	46
64.	Rendimiento en función de los umbrales	47
65.	Relación entre las rentabilidades reales de un año y el siguiente	49
66.	Comparativa dummy vs red neuronal - Conjunto de entrenamiento	49
67.	Predicción vs realidad de las empresas seleccionadas - Red Neuronal	50
68.	Rendimiento de la cartera por países	51
69.	Comparativa de rentabilidad según los umbrales escogidos	51
70.	Análisis de sensibilidad de la red neuronal	52
71.	Predicción para el año 7 tomando los anteriores 4 años - Entrenamiento	53
72.	Predicción para el año 7 tomando los anteriores 4 años - Test	54
73.	Rendimientos de los distintos modelos	54
74.	Comparativa entre diferentes tamaños de redes neuronales	55
75.	Empresas seleccionadas según la segunda estrategia- Red Neuronal	56

76.	Empresas seleccionadas según la segunda estrategia- Red Neuronal	56
77.	Resultados de la nueva estrategia	57
78.	Objetivos de desarrollo sostenible	63

Índice de tablas

1.	Métricas para la función de auto correlación	11
2.	Características de las series temporales utilizadas	13
3.	Valores que representan los percentiles de cada indicador	18
4.	Distribución de los clúster por mercado	35
5.	Distribución de los clúster por sector	35
6.	Número de integrantes de cada cada clúster	37
7.	Compañías y estrategias con un rendimiento superior a 10	41

1. Introducción

El aprendizaje automático (más conocido por su nombre en inglés *machine learning*) es una herramienta que, a través del diseño de algoritmos, establece relaciones, de mayor o menor complejidad, entre una serie de variables de entrada y de salida a través de un grupo inicial de datos. Es similar al aprendizaje que seguimos los humanos por la experiencia hasta cierto punto. Esta técnica tiene, a día de hoy, aplicaciones muy diversas, siendo las más comunes el reconocimiento de imágenes y sonido y la predicción de todo tipo de indicadores.

Dentro del aprendizaje automático existen dos ramas: el aprendizaje no supervisado y el supervisado. Ambos aprendizajes parten de un conjunto inicial de datos. En aprendizaje no supervisado los algoritmos buscan relaciones ocultas entre los datos y, por lo tanto, no es necesario indicar valores de salida. Por otro lado, en el aprendizaje supervisado, al algoritmo se le indican las salidas, y éstas quedan incorporadas en los datos. Además, entre los sistemas supervisados se distinguen dos variantes: la clasificación y la regresión. En el aprendizaje supervisado además se precisa la separación del conjunto de datos en, normalmente, dos subconjuntos: el subconjunto de entrenamiento (o *training*), utilizado para ajustar los parámetros del algoritmo seleccionado; y el subconjunto de validación, para verificar la bondad del modelo, y que éste funciona con datos distintos a aquellos con los que fue entrenado. En la figura 1, tomada de la web de Matlab [Mat19], se muestra un esquema de los tipos de aprendizaje automático y los algoritmos más frecuentes.

Para finalizar la introducción al aprendizaje automático, existe una técnica dentro del aprendizaje automático que se denomina aprendizaje profundo (ampliamente conocido por su nombre en inglés: *deep learnig*). Esta técnica utiliza mayoritariamente redes neuronales para aprender del conjunto de datos. Como se indica en el curso de Métodos Cuantitativos para la Decisión, del Máster de Ingeniería Industrial del ICAI, las redes neuronales son modelos aditivos no lineales de regresión, y cuyo funcionamiento está basado en el funcionamiento del cerebro [SSR18]. Existen diversas configuraciones de redes neuronales, pero en la figura 2, tomada de una web específica sobre redes neuronales[vV19], se muestran las principales configuraciones básicas de estas redes así como las distintos clases de neuronas.¹

Por otro lado, como se dice en un artículo de *The Economist* [Eco19], la inversión en los mercados bursátiles de empresas y el comercio de las acciones ha sufrido grandes cambios en los últimos 50 años. Inicialmente se trataba de un trabajo realizado, casi exclusivamente, por las personas, con grandes incertidumbres en el mercado, dado que no existían los sistemas informático que de los cuales disponemos a día de hoy . No obstante, las transacciones y los procesos cada día están más ligados a la automatización, llegando al nivel de que el 35% de las acciones del mercado americano se gestionan de forma automática, como se puede ver en la figura 3, tomada de [Eco19].

¹Aunque en la imagen original se muestran 29 estructuras distintas, aquí tan solo se muestran 11

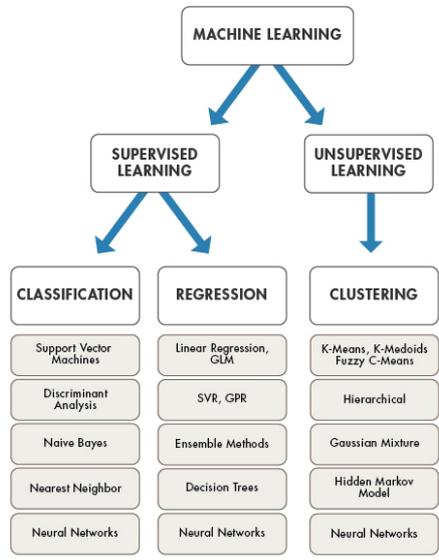


Figura 1: Tipos de aprendizaje automático y algoritmos utilizados

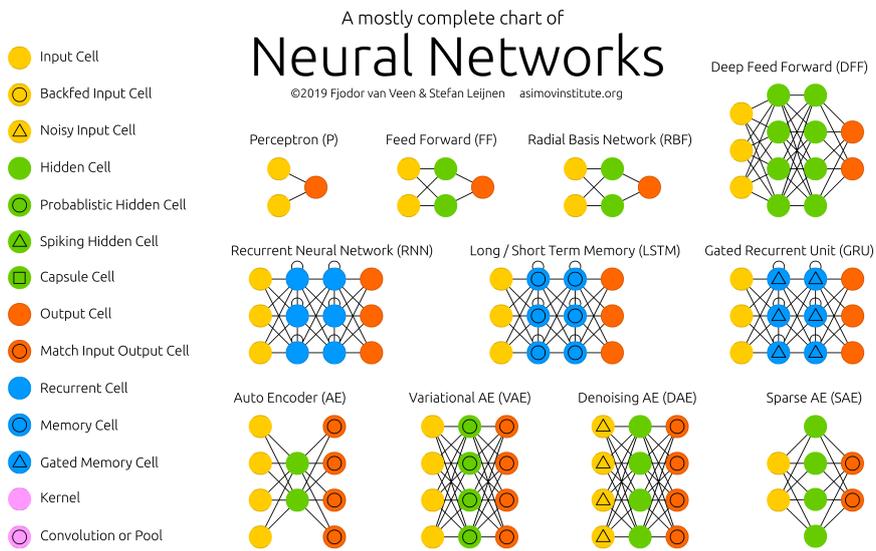
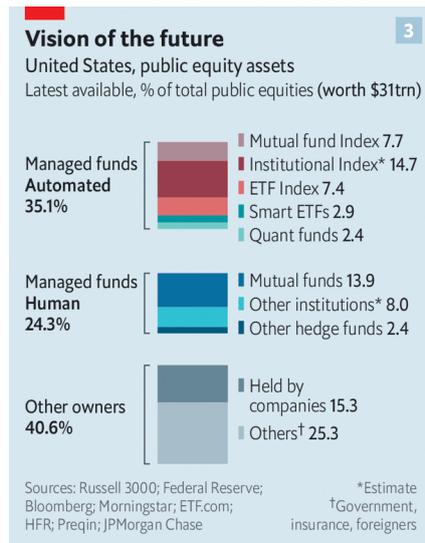


Figura 2: Principales tipos de redes neuronales y de neuronas



The Economist

Figura 3: Distribución de los tipos de empresas que gestionan acciones en el mercado americano

Este apogeo de la gestión automática de los mercados y carteras bursátiles es una tendencia que ha cogido fuerza en los últimos 9 años. Esto es debido al gran desarrollo de las herramientas de análisis cuantitativo avanzado, que es la herramienta básica que se utilizan en los fondos cuantitativos (*quant funds*), de acuerdo con las fuentes consultadas [qua19]. Como se puede apreciar en la figura 4, tomada de un artículo de *The Economist* [Eco19], este tipo de fondos es el más abundante en los Estados Unidos desde 2016. Esto es coherente con lo que se indica en el artículo del Instituto de ingenieros eléctricos y electrónicos (IEEE) [NPdO17] aproximadamente el 85 % de las transacciones bursátiles fueron realizadas por algoritmos.



The Economist

Figura 4: Distribución del comercio de acciones del mercado americano según el tipo de compañía que los gestionan

Por otro lado, en el mundo del comercio bursátil existe un problema denominado gestión de carteras (en inglés: *Portfolio Management*). La gestión de carteras se define en un artículo de Adam Hayes como el arte y la ciencia de seleccionar y supervisar un grupo de inversiones bursátiles para cumplir con los objetivos a largo plazo en términos financieros y de tolerancia al riesgo. El objetivo de un gestor de carteras es el de maximizar el rendimiento de los activos gestionados (en este caso, acciones de empresas) y cumpliendo con las distintas tolerancias de riesgo.

El problema que surge a la hora de tomar las decisiones de inversión es la gran cantidad de posibilidades que se abren y el gran número de valores en los que se puede invertir. El objetivo de este trabajo es el de utilizar las últimas técnicas desarrolladas en el aprendizaje automático para ayudar en estas decisiones relacionadas, de forma que se maximice el rendimiento económico. Por ello, se va a realizar una revisión del estado del arte actual de este tipo de técnicas aplicadas. Para ello, las series temporales relacionadas con los valores bursátiles se transformarán en una serie de características que representan dicho valor. Con estas características se buscarán relaciones ocultas o difícilmente alcanzables a través de otros análisis. Esta búsqueda se realizará a través de un análisis exploratorio de los datos, visualizando los datos sobre gráficos de puntos, con análisis de componentes principales y clusterización.

Tras el análisis exploratorio se procederá a aplicar las técnicas de aprendizaje supervisado sobre estrategias de inversión. Para ello se aplicará la metodología sobre un tipo de estrategia usado comúnmente en la actualidad. Se usarán las características de las series temporales y datos anteriores de rentabilidad para predecir la rentabilidad de las distintas empresas a lo largo de un periodo de tiempo determinado. Una vez realizada la predicción se deberá tomar una decisión de si se debe invertir en el valor. Se evaluará, para los distintos espacios de tiempo, la bondad de este tipo de estrategia, comparado varias estrategias distintas y con distintos modelos de predicción.

2. Estado del Arte

A día de hoy se está volviendo cada vez más común que distintas organizaciones hacen esfuerzos por obtener y acumular grandes cantidades de datos. A veces, incluso, se pueden recabar tantos datos que no se poseen (en la actualidad) los medios necesarios para procesarlos todos. Como se indica en el artículo de Rob J Hyndman et al. [Hyn15], el uso de técnicas de selección de características sobre series temporales permite identificar las series temporales cuyos comportamientos son anómalos en el rendimiento de servidores, de entre un total de 1500 series temporales.

También en esta línea se desarrolla el artículo de Yanfei Kang et al. [KHSM17], donde se usa el conjunto de más de tres mil series temporales del *International Institute of Forecasters* para la competición M-3. Estos datos incluyen datos sobre empresas, industrias, valores macroeconómicos y demográficos. En este artículo se trata de visualizar sobre 2 ejes todas las series temporales con las que se trabaja. Estos dos ejes representan las componentes principales de una serie de características de las series temporales. Sobre esta visualización se pueden evaluar y medir la eficacia de predicciones. También se propone un método para generar nuevas series temporales con características controladas, haciendo que las generalizaciones de la predicción sean más robustas.

Siguiendo esta línea, se escribe el artículo de [THA18], donde se propone la selección de modelos de predicción basada en las características de las series temporales o FFORMS por sus siglas en inglés (Feature-based FORecast-model Selection). En este artículo se usan algoritmos de tipo bosque aleatorio (random forest) para identificar el mejor método de predicción de series temporales haciendo uso exclusivo de las características de las series. La metodología propuesta (FFORMS) tiene la ventaja, respecto a otras metodologías revisadas en el artículo, de incluir la clasificación de las series antes de realizar la predicción.

Para terminar con esta línea, existe un paquete de R creado por Mitchell O’Hara-Wild (aunque otros autores han contribuido en el desarrollo del paquete) para la extracción de características y estadísticos de series temporales denominado FEASTS (*Feature Extraction And Statistics for Time Series*). Este paquete, como viene indicado en su descripción [OWHW⁺20], provee de una colección de características, métodos de descomposición, resúmenes estadísticos y funciones gráficas para el análisis de series temporales. Este paquete será utilizado en el desarrollo del presente trabajo.

Hasta ahora todos los artículos revisados tenían una remota relación con datos bursátiles. Sin embargo, en paralelo, se han desarrollado aplicaciones de aprendizaje automático basados en redes neuronales para predecir los precios de las acciones en el futuro. Concretamente, se van a presentar artículos publicados entre 2015 y 2019.

En 2015 se publicó un artículo en la Conferencia Internacional del Congreso de Big Data del Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) [CZD15] donde se desarrolla un sistema de predicción consiguió una precisión del 27 % en un caso de estudio de

predicción para acciones del mercado chino en Shanghai y Shenzhen. Este caso de estudio se realizó con los datos relativos al precio apertura, cierre, máximo, mínimo, volumen diarios y un índice denominado *SSE* y, utilizando una red neuronal con memoria al largo y corto plazo, también conocida como LSTM (*Long-short term memory*). Los datos de entrada son básicamente toda la información diaria (sin llegar a entrar en la evolución del mercado intra-diario) que se suele tratar. Además, se utilizaron aproximadamente un millón de series de 3049 valores entre los años 2013 y 2015. Pese a que en el artículo se indique que la precisión no es muy alta, predecir con una precisión del 27% en las predicciones puede ser un auténtico éxito dada la volatilidad que marca este tipo de mercados, además que puede reportar grandes beneficios si se consigue utilizar bien la predicción. Sin embargo, esto depende de cómo sea medida la precisión, dado que no se explica cómo se ha realizado el cálculo. Además, en el artículo no se explica si se ha estudiado que la precisión se debida a un posible sobre-ajuste de los parámetros.

Siguiendo la línea temporal, se publicó en 2016 un artículo en la Conferencia Internacional sobre la Ciencia Computacional y de la Información (ICIS) [AYMU16] donde se incluyen además, de los valores históricos, la influencia de las noticias a través del aprendizaje profundo (*Deep Learning*). Para ello se integraron a la red neuronal una serie de combinaciones de vectores de párrafos de un modelo de memoria distribuida (también denominado PV-DM por sus siglas en inglés) y de bolsa de palabras distribuida (conocido como PV-DBOW por sus siglas en inglés). Sin embargo, como no siempre se tienen noticias actualizadas sobre la empresa o, de haberlas, lo normal es que haya más de una, se decide rellenar con ceros los huecos y hacer una media cuando hay varias. Además, por simplicidad, en vez de analizar la totalidad de las noticias se analizan exclusivamente los titulares de dichas noticias. Se utiliza un modelo basado en redes neuronales de tipo LSTM y se indica que, al añadir información en forma de textos relativos a los titulares de las noticias a los datos numéricos, se mejoran los resultados respecto al caso del uso exclusivo de datos numéricos. Siendo crítico, este comportamiento se considera normal, dado que, al haber más variables, se permite ajustar mejor el modelo.

También en 2016 se publicó un artículo en el Periódico Internacional de Circuitos, Sistemas y Procesado de Señales [DPH16] donde se compara la bondad de varias arquitecturas distintas de redes neuronales, en concreto el perceptrón multi-capas, redes neuronales convolucionales, recurrentes y combinaciones de *wavelet* con redes neuronales convolucionales y, por último, una combinación de todos los modelos anteriores. A nivel individual, las redes que mejor se ajustan a este tipo de predicciones son las redes convolucionales, obteniendo una notable mejoría en la precisión si se combinan con los *wavelets*.

En 2017 se publica otro artículo en el IEEE [NPdO17] donde se realiza una aproximación distinta: hasta ahora todos los datos utilizados eran los del mercado diario, mientras que en este artículo se toman cada 15 minutos (mercado intra-diario) y se realiza un pre-procesado de la serie, con transformaciones de tipo logarítmicas y de tipo suavizados exponenciales de las series. Se estudia, por tanto, una estrategia de inversión a muy corto plazo y consigue una precisión del 56% indicando a la hora de predecir si el precio de una acción va a subir o a bajar en el periodo de estudio.

En 2018 se publica el artículo de la Conferencia Internacional de los Sistemas de Información y Comunicación (ICICS) [AEAM18] donde se toman datos bursátiles de los últimos 7 años y de las 500 empresas americanas más importantes referenciado al índice Standard and Poor's 500 (S&P500). Se utilizan redes neuronales de tipo LSTM apiladas (SLSTM por sus siglas en ingles *Stacked LSTM*) y de tipo bidireccional (BLSTM). Se predice con respecto a dos horizontes: al corto y al largo plazo. Además, se utilizan tres medidas de desempeño para evaluar el nivel de bondad de dichas predicciones. Se concluye que las redes de tipo SLSTM y BLSTM son mejores para el corto plazo pero que los resultados al largo plazo no son notablemente peores. En la figura 5, tomada de [AEAM18], se muestra la estructura de una red neuronal de tipo BLSTM.

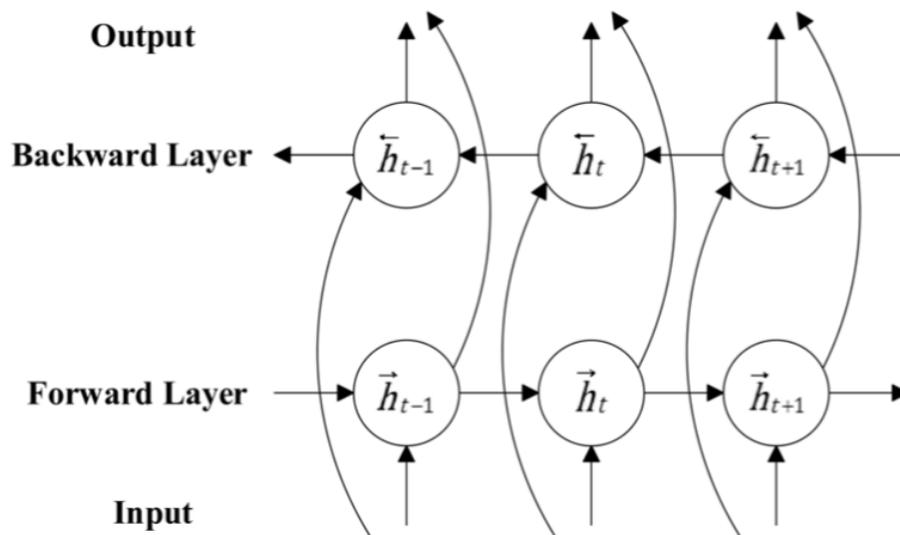


Figura 5: Estructura de una BLSTM

Por último, en 2019 se publicó un artículo de Ricardo Michel Reyes [Rey19], donde se dice que la inteligencia artificial (que se basa mayoritariamente en este tipo de sistemas de aprendizaje automático) no aplicaría en un entorno como es el de los mercados bursátiles. No obstante, se indica que estas técnicas sí que pueden ayudar en la selección de un portafolio más ajustado a las necesidades de un perfil, es decir, optimizar la gestión de carteras.

Siguiendo con esta última afirmación se presentan las principales técnicas para la gestión de carteras. En la bibliografía económica [Hay20] se distinguen dos formas de gestionar las carteras: la gestión activa y la gestión pasiva. La gestión pasiva consiste en comprar las acciones con una visión al largo plazo, dejando que éstas evolucionen según el mercado. Esta gestión suele significar la inversión simultánea en varias empresas. A este tipo de gestión se le suele denominar también indexación o inversión en índices (*index investing*). Algunos de estos índices son el Standard & Poor's 500 o el IBEX 35. Esta gestión de carteras suele ir guiada por la gestión de carteras moderna (*Modern Portfolio Management* - MLP). Estas siglas no deben confundirse con el perceptron multicapa (*Multi-Layer Perceptron* - MLP), que es un tipo de red neuronal.

Por otro lado, la gestión activa trata de vencer el rendimiento de un índice, de activos indexados o del mercado a través de una compraventa activa de valores individualizados. Este tipo de gestión suele ir acompañada de una gran variedad de modelos de tipo cuantitativo y cualitativo que permiten una mejor evaluación de las inversiones potenciales a las que se tiene acceso. Este trabajo iría enmarcado en este tipo de gestión.

Además en el artículo sobre la gestión de carteras con inteligencia artificial (gestión de carteras activa) [Hon19], Alexandr Honchar se centra en la predicción de las series temporales y la gestión de carteras. Realiza una aproximación a la inversión a través de un aprendizaje no supervisado, tratando de agrupar las distintas compañías según la rentabilidad que aportan. Posteriormente realiza un aprendizaje supervisado para optimizar los pesos que da a cada valor. Finalmente realiza un aprendizaje por refuerzo y evalúa la estrategia con una serie de valores bursátiles y criptomonedas.

Tras esta revisión del estado del arte se procede a presentar la metodología propuesta para ayudar en la selección de valores en los que invertir.

3. Metodología Propuesta

Los datos con los que se va a tratar son los valores de bolsa de varias compañías (1274) a lo largo de 10 años, de distintos mercados de valores y muchos sectores e industrias. Estos datos, provenientes de Yahoo! Finance y de Bloomberg, tienen para cada día varios indicadores distintos. Los indicadores que se tienen para cada compañía y cada día son los explicados (brevemente) a continuación.

- **Stock_quotes_open:** Valor de cotización de apertura.
- **Stock_quotes_close:** Valor de cotización de cierre.
- **Stock_quotes_high:** Máximo diario del valor de cotización.
- **Stock_quotes_low:** Mínimo diario del valor de cotización.
- **Stock_quotes_volume:** Volumen diario de acciones vendidas.
- **CUR_MKT_CAP** Valor de mercado de la empresa, igual al número de acciones existentes por el precio de cierre de las acciones.
- **BEST_TARGET_PRICE:** Mejor precio objetivo propuesto según el análisis de Bloomberg.
- **BEST_ANALYST_RATING:** Mejor r ating de analistas de Bloomberg.
- **APPLIED_BETA:** Factor de riesgo espec fico de la empresa respecto al mercado (Beta) aplicada para los c lculos de Bloomberg.
- **Currency_rates:** Valor de la divisa respecto al Euro
- **EBITDA:** Se refiere al beneficio de la empresa en los  ltimos 12 meses, antes de incluir los efectos de los intereses, impuestos, depreciaciones y amortizaciones. Se actualiza cada tres meses.
- **PE_RATIO:** Ratio entre el precio de la acci n y los ingresos de la compa n a por acci n. Los ingresos se refieren a los ingresos en los  ltimos 12 meses y se actualizan cada tres meses.
- **PX_TO_BOOK_RATIO:** Ratio entre el valor de mercado de la empresa y el valor contable.

Adem s, de cada empresa se conoce a qu  mercado pertenece, por ejemplo el mercado espa ol (MCE), el londinense (LSE) o el de neoyorquino (NYSE). Tambi n se tienen clasificadas las empresas por sector e industria. Por  ltimo, se tiene un indicador de la moneda en la que se comercializa la empresa. Cada empresa se identifica con un peque o texto llamado *Ticker*.

Se ha de indicar que este tipo de datos no est n disponibles en todas las fechas, ya que puede haber nuevas salidas a bolsa, quiebras y fusiones entre empresas. En total, combinando compa n as, d as cotizados e indicadores, se superan los 40 millones de datos.

En la figura 6 se muestra un ejemplo de las series temporales que componen un solo país (España) y un solo indicador (valor de cierre). Como se puede apreciar, resulta muy complicado encontrar así en qué valores invertir.

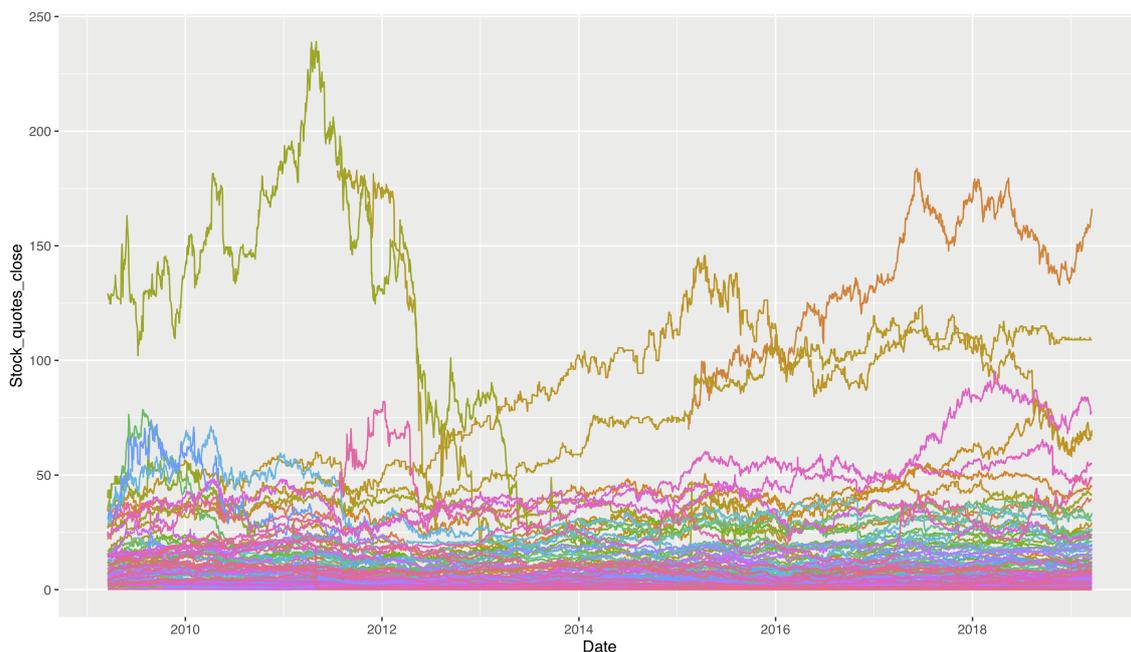


Figura 6: Evolución de los valores españoles a lo largo de 10 años

La metodología que se va a seguir es la de agrupar las series temporales a través de las principales características de las mismas. Estas características, tomadas de [HA19] se dividen en varios grupos:

1. **Características STL:** Calcula una serie de métricas de una descomposición STL para las series temporales. Esta descomposición separa la serie original (y_t) en tres sub-series: una de tendencia (T_t), una de estacionalidad (S_t) y una de resto (R_t). La expresión matemática de la descomposición se muestra en la ecuación 1.

$$y_t = T_t + S_t + R_t \quad (1)$$

Sin embargo, en este estudio estamos más interesados en incluir en los análisis la fuerza de estas tendencias o estacionalidades. Para ello, las características que más información dará es la representada en las ecuaciones 2 (fuerza de la tendencia) y 3 (fuerza de la estacionalidad).

$$T_s = \max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)} \right) \quad (2)$$

$$T_s = \max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)} \right) \quad (3)$$

Además, hay otras características que se pueden calcular tales como la linealidad, la curvatura o cómo de puntiaguda es la serie o *spikiness*. La característica de *linealidad*

mide la linealidad de la componente de tendencia (T_t). Se basa en el coeficiente de regresión lineal aplicado a la componente de tendencia. La característica de *curvatura* mide la curvatura de la componente de tendencia. Se basa en el coeficiente de regresión para una proyección ortogonal cuadrática de la componente de tendencia. Por último, el *spikiness* mide la prevalencia de las puntas en la componente de resto (R_t). Se calcula como la varianza de las varianzas excluyendo un dato. Este último cálculo será más largo cuanto más larga sea la serie.

Por último, se proporciona también el día dentro de la estacionalidad (en este caso, la semana) en el que se encuentran los picos y los valles. Estos se representan en las características `seasonal_peak_week` y `seasonal_trough_week` respectivamente.

Las características de este tipo se denominarán `trend_strength_week`, `seasonal_strength_week`, `spikiness`, `linearity`, `curvature`, `seasonal_peak_week` y `seasonal_trough_week`

2. **Características ACF:** Se refiere a las funciones de auto correlación de la serie (Auto Correlation Function). La auto correlación muestra la dependencia lineal de una variable respecto a sí misma n instantes de tiempo anteriores. La auto correlación se calcula según la ecuación 4.

$$\rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \quad (4)$$

En el caso de estudio, se va a calcular el valor de la ACF para un salto de un instante de tiempo (un día hábil) así como la ACF para el periodo de estacionalidad y la suma del cuadrado de dicha función para los primeros 10 días. Los mismos cálculos (ACF y suma) se realizan para la serie diferenciada una y dos veces. De esta forma, las características ACF se resumen en la tabla 1

Indicador	Fórmula
acf1	$\frac{Cov(y_t, y_{t-1})}{Var(y_t)}$
acf10	$\sum_{k=1}^{10} \left(\frac{Cov(y_t, y_{t-k})}{Var(y_t)} \right)^2$
season_acf1	$\frac{Cov(y_t, y_{t-s})}{Var(y_t)}$
diff1_acf1	$\frac{Cov(y'_t, y'_{t-1})}{Var(y'_t)}$
diff1_acf10	$\sum_{k=1}^{10} \left(\frac{Cov(y'_t, y'_{t-k})}{Var(y'_t)} \right)^2$
diff2_acf1	$\frac{Cov(y''_t, y''_{t-1})}{Var(y''_t)}$
diff2_acf10	$\sum_{k=1}^{10} \left(\frac{Cov(y''_t, y''_{t-k})}{Var(y''_t)} \right)^2$

Tabla 1: Métricas para la función de auto correlación

3. **Características PACF:** la función de auto correlación parcial (Partial Auto Correlation Function) representa la auto correlación entre dos instantes de tiempo quitando los efectos de los instantes intermedios. Se tomará la función de auto correlación parcial para un salto de 5 días (1 semana laborable) para la serie original y la serie diferenciada 1 y 2 veces. También se calculará la función de auto correlación parcial para el periodo de estacionalidad.
4. **Característica de Guerrero:** Se calcula el valor de λ para la transformación de Box-Cox [BC64], utilizada para normalizar (en forma) variables dependientes no normales, a través del método de Guerrero [Gue93]. Este método minimiza el coeficiente de variación de las sub series de la serie original. Esta característica se denominará `lambda_guerrero`.
5. **Raíces unitarias:** Una raíz unitaria [Mr.08, Ste16] es una tendencia estocástica en series temporales por la cual, si dicha serie posee una raíz unitaria, tendrá un patrón sistemático que será impredecible. La existencia de raíces unitarias es una causa que provoca que un proceso no sea estacionario. Existen varios métodos para comprobar la existencia de dichas raíces unitarias, como el KPSS (desarrollado por Kwiatkowski et al.) que contrasta con una tendencia lineal, el PP (desarrollado por Phillips y Perron) que contrasta con una tendencia constante y un salto de 1 [KS92] o el `ndiffs` (número mínimo de diferenciaciones requeridas para que el proceso sea estacionario). Las dos primeras tienen asociadas un valor del estadístico y un p-valor.
6. **Características de los Tests de Portmanteau:** Consiste en un test cuyo estadísticos sirven para evaluar la hipótesis nula de independencia en una serie temporal. Este test es el denominados Ljung-Box o Box-Pierce. Se obtendrá el valor de cada estadístico y su p-valor. Se denominarán `lb_stat`, `lb_pvalue`, `bp_stat` y `bp_pvalue`. Los resultados pueden ser muy similares.
7. **Características de las ventanas sin solape:** Se mide la estabilidad y el abultamiento de la serie. La estabilidad se mide como la varianza de las medias de las ventanas mientras que el abultamiento se mide como la varianza de las varianzas de las ventanas. El tamaño de la ventana utilizada será el de la estacionalidad encontrada. Las variables se llamarán `var_tiled_var` y `var_tiled_mean`.
8. **Características de las ventanas deslizantes:** Se trata del cálculo de la mayor variación de la media, de la varianza y de la divergencia de Kullback-Leibler entre dos ventanas consecutivas. La divergencia de Kullback-Leibler sirve para medir cómo de similares o diferentes son dos distribuciones de probabilidad. Los dos indicadores que se dan para cada variación es el nivel de dicha variación y la posición (dentro de la serie temporal) en la que ocurre. Estas características se denominarán `shift_level_max`, `shift_level_index`, `shift_var_max`, `shift_var_index`, `shift_kl_max` y `shift_kl_index`.
9. **Característica de entropía espectral:** Esta característica mide la dificultad para predecir una serie. Esto se puede medir a través del cálculo de la entropía espectral de Shannon. Una serie que tenga una tendencia y estacionalidad fuertes, por lo que será más fácil de predecir, tendrá una entropía baja, cercana a cero; mientras que una serie que sea ruidosa, será difícil de predecir y tendrá una entropía alta, es decir, cercana a 1. Este indicador se denominará `spectral_entropy`
10. **Características de puntos:** Se refieren al número de veces que la serie cruza su

mediana y el mayor número de puntos constantes consecutivos. Estas características se denominarán `n_crossing_points` y `n_flat_spots`.

11. **Coficiente de Hurst:** Una estimación del coeficiente o exponente de Hurst, que es una medida de la memoria a largo plazo de una serie temporal [Sub12]. Valores entre 0 y 0.5 indican que la serie es anti persistente, un valor cercano a 0.5 indica un comportamiento Browniano o de camino aleatorio, mientras que valores entre 0.5 y 1 indican que la serie es persistente. El parámetro se denominará `coef_hurst`.
12. **Estadístico ARCH LM:** Se refiere al estadístico basado en el test del multiplicador de Lagrange (Lagrange Multiplier - LM) para la heterocedasticidad condicional autorregresiva (Autoregressive conditional heteroscedasticity - ARCH). Se denominará `stat_arch_lm`

En la tabla 2 se muestran todos los indicadores, que puede servir de referencia a lo largo del trabajo.

<code>trend_strength</code>	<code>seasonal_strength_week</code>	<code>spikiness</code>
<code>linearity</code>	<code>curvature</code>	<code>seasonal_peak_week</code>
<code>seasonal_trough_week</code>	<code>acf1</code>	<code>acf10</code>
<code>diff1_acf1</code>	<code>diff1_acf10</code>	<code>diff2_acf1</code>
<code>diff2_acf10</code>	<code>season_acf1</code>	<code>pacf5</code>
<code>diff1_pacf5</code>	<code>diff2_pacf5</code>	<code>season_pacf</code>
<code>lambda_guerrero</code>	<code>kpss_stat</code>	<code>kpss_pvalue</code>
<code>pp_stat</code>	<code>pp_pvalue</code>	<code>ndiffs</code>
<code>nsdiffs</code>	<code>bp_stat</code>	<code>bp_pvalue</code>
<code>lb_stat</code>	<code>lb_pvalue</code>	<code>var_tiled_var</code>
<code>var_tiled_mean</code>	<code>shift_level_max</code>	<code>shift_level_index</code>
<code>shift_var_max</code>	<code>shift_var_index</code>	<code>shift_kl_max</code>
<code>shift_kl_index</code>	<code>spectral_entropy</code>	<code>n_crossing_points</code>
<code>n_flat_spots</code>	<code>coef_hurst</code>	<code>stat_arch_lm</code>

Tabla 2: Características de las series temporales utilizadas

A través de todos estos parámetros, se pueden transformar las series en puntos que pueden ser agrupados, transformados y dar ideas sobre el funcionamiento de las dinámicas internas de las series, ayudando a identificar, por ejemplo, qué valores pueden ser más o menos arriesgados o más o menos rentables. Un esquema de esta transformación se muestra en la figura 7. Con esta metodología se va a realizar un primer análisis exploratorio de los datos.

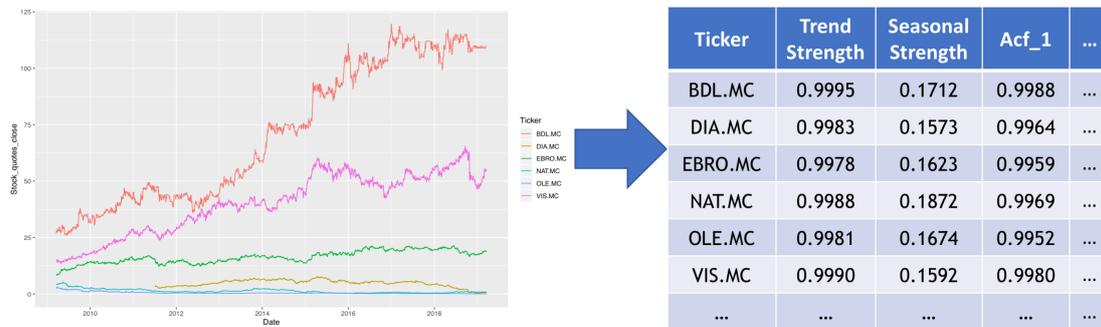


Figura 7: Esquema de la transformación realizada

4. Análisis Exploratorio de los Datos

El objetivo de esta sección es el de realizar un análisis exploratorio de estas series, realizando una serie de análisis para tratar de encontrar patrones en los datos y que éstos ayuden a entender mejor las dinámicas internas que gobiernan cada una de las series o agrupar las series que tengan dinámicas similares. Para ilustrar mejor el conjunto de datos inicial y así poder entender mejor los siguientes gráficos, se muestra el gráfico de la figura 8. En este gráfico aparece representado el número de compañías dentro de cada mercado, y divididas por sector al que pertenece cada compañía. Se puede apreciar que el mercado americano (NYQ) es el que más empresas tiene, seguido del Australiano (ASX). El mercado español (MCE) se encuentra en cuarta posición.

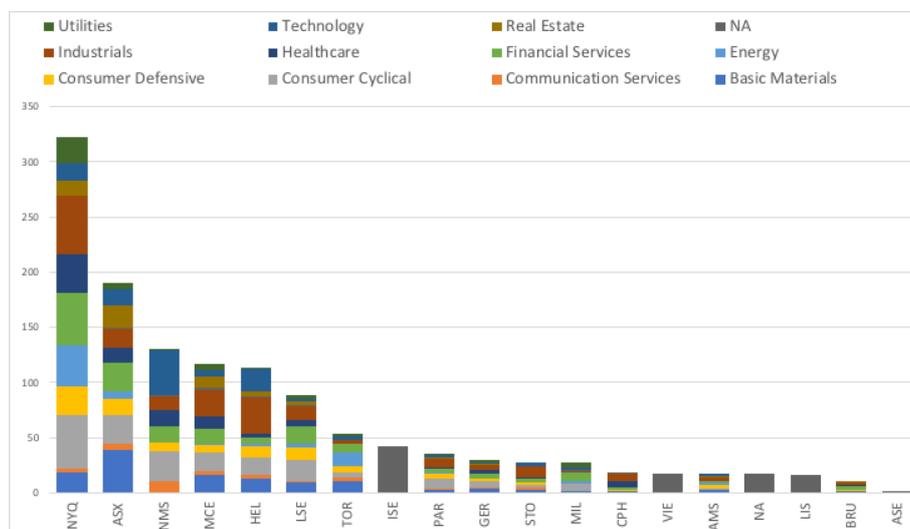


Figura 8: Distribución de las empresas según mercado y sector

4.1. Métricas principales

En primer lugar se van a calcular todos los parámetros o características citadas en la sección anterior para todos los indicadores. Para esto se van a analizar las series a lo largo de los 10 años con el paquete FEASTS de R. Estas series, como se dijo en el apartado anterior, son de los datos de cierre, apertura, máximo y mínimo de los valores, pero también indicadores como el EBITDA o ratios. Cada serie de empresa-indicador (por ejemplo ACS y valor de cierre sería una serie) es representado por un punto. En este contexto se muestra la figura 9, donde aparecen representados dos de los parámetros.

Se han representado la fuerza de la tendencia en el eje x y la fuerza de la estacionariedad semanal en el eje y. Estas características se han seleccionado a título ilustrativo. Además, los puntos representan las series que se refieren al volumen de acciones de una empresa vendidas en el día, y el color del punto indica a qué mercado pertenece dicha empresa. Se puede apreciar que las tendencias de los volúmenes negociados en el mercado

neoyorquino son, en general, más fuertes que en el resto de países, mientras que en Helsinki las tendencias son más flojas. En cuanto a la fuerza de la estacionalidad semanal, se puede ver que los valores son bajos (la gran mayoría inferiores a 0.25) y que no existe un mercado que tenga mayor fuerza estacional.

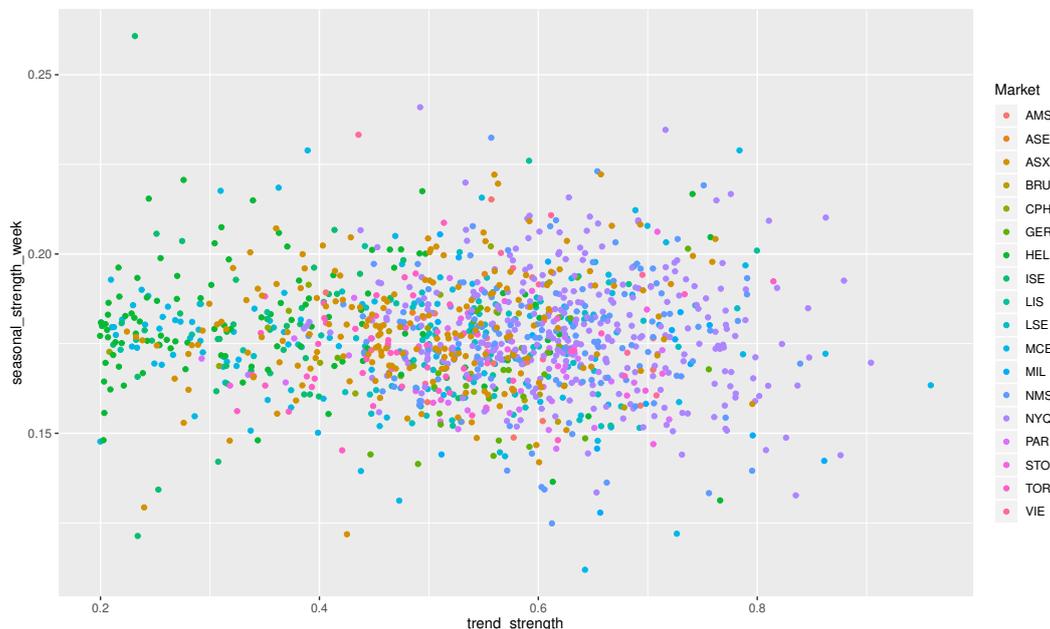


Figura 9: Fuerza de la tendencia y la estacionalidad para el volumen de acciones

Siguiendo este esquema, se presenta la figura 10, donde se representan los valores de trend_strength y seasonal_strength_week (al igual que en la figura anterior), pero esta vez con todos los indicadores, no sólo el volumen vendido del día. Cada indicador es representado separado del resto.

Se puede apreciar que en términos de APPLIED_BETA los valores son, en general, todos muy parecidos, con fuertes tendencias pero bajas estacionalidades. También se pueden apreciar ciertos datos raros, que pueden resultar en valores atípicos. Esto puede tener cierto interés a la hora de decidir en qué valores invertir. Esto ocurre con casi todos los indicadores. El único que tiene un gran abanico de fuerzas de tendencia es el volumen de transacciones, comentado anteriormente. Parece ser que hay mercados, como Helsinki, donde las series de volumen tienen menor tendencia.

Como lo que nos interesa es sacar rentabilidad a los valores, parece razonable investigar más a fondo los datos relativos al valor de dichas acciones. Para ello, se va a analizar más en profundidad el precio de apertura de los valores (Stock_quotes_open).

Con este objetivo, se presenta la figura 11, donde se muestran las características anteriores (fuerza de tendencia y de estacionalidad) organizadas por sector. Se han quitado los valores que no estuvieran clasificados en algún sector.

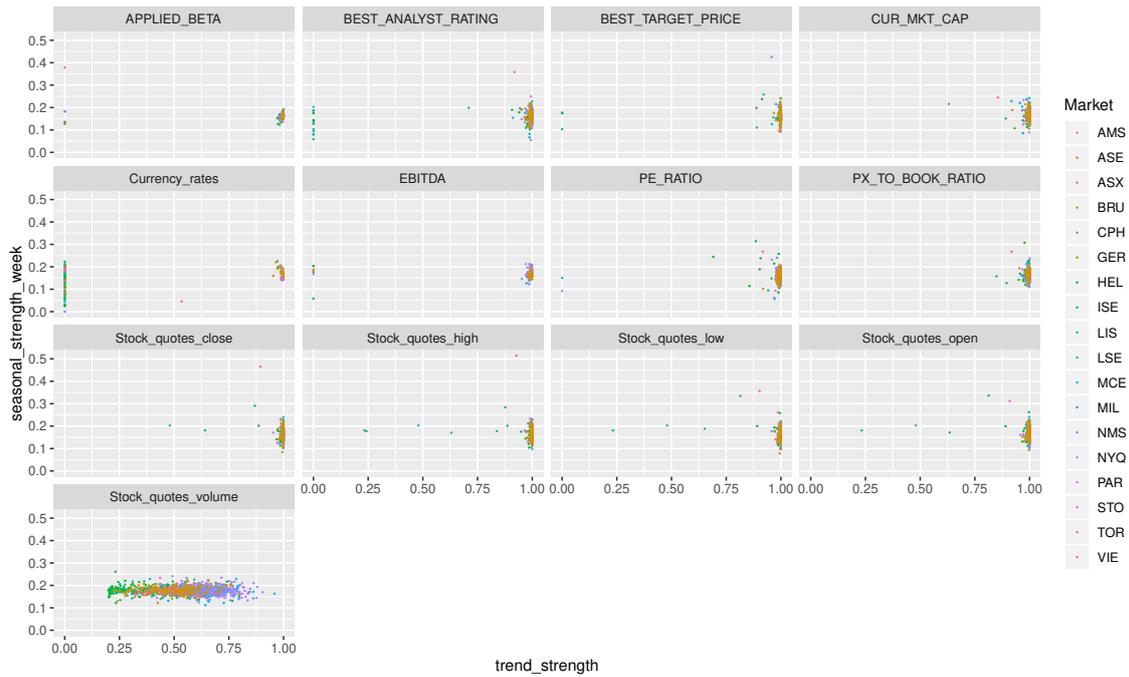


Figura 10: Fuerza de la tendencia y la estacionalidad de los indicadores

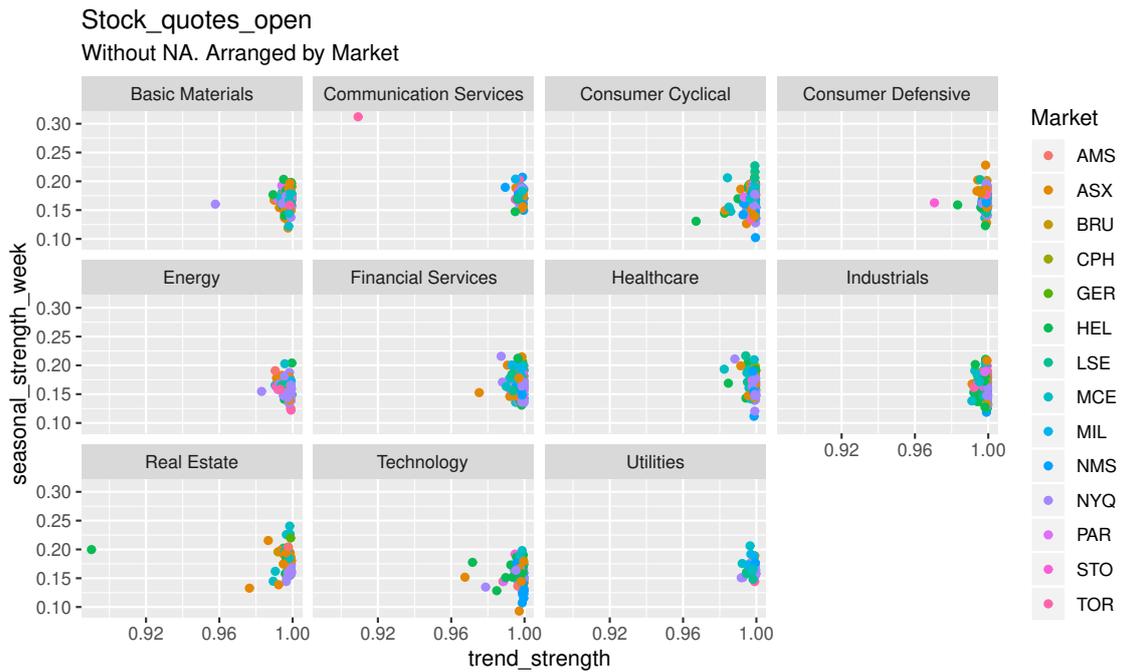


Figura 11: Fuerza de la tendencia y la estacionalidad del valor de apertura

Esta figura nos muestra que algunos sectores tienen mayores variaciones en la fuerza de su estacionalidad y tendencia que otros. Por ejemplo, el sector de las *Utilities* tiene tendencias y estacionalidades muy concentradas, mientras que el sector de las tecnologías tiene un abanico mayor de fuerza de la estacionalidad. Además, se siguen localizando valores raros, como en el sector de los servicios de comunicación, donde un valor (del

mercado suizo) tiene un comportamiento muy distinto al del resto del sector.

Resulta interesante ver cómo afectan estos dos indicadores (fuerza de la tendencia y estacionalidad) a las series temporales. Por esta razón, se ha decidido mostrar las series que representan percentiles 10, 50 y 90 de cada uno de estos indicadores. Los valores que representan estos percentiles se muestran en la tabla 3 y las series se muestran en las figuras 12 a 17. En estas figuras se puede ver como se han clasificado series con distinta longitud basándose en los indicadores.

Indicador	Percentil 10	Percentil 50	Percentil 90
<i>Trend_strength</i>	IPH.AX	STAN.L	MCD
<i>Seasonal_strength_week</i>	KSS	MTS.MC	CAF.MC

Tabla 3: Valores que representan los percentiles de cada indicador

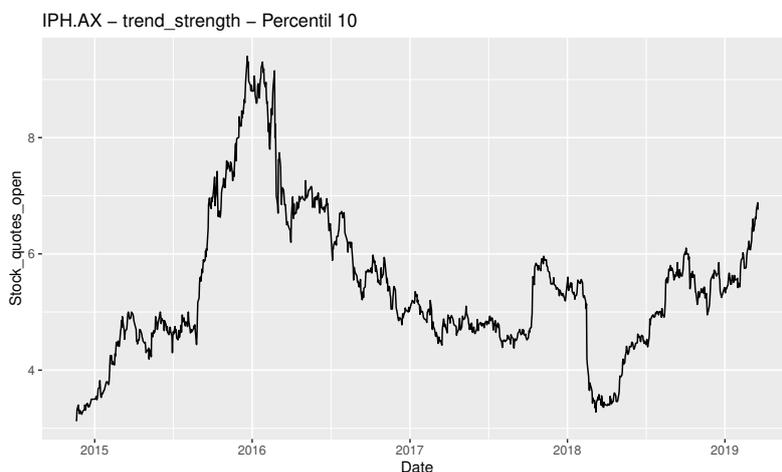


Figura 12: Valor correspondiente a percentil 10 en trend_strength



Figura 13: Valor correspondiente a percentil 50 en trend_strength

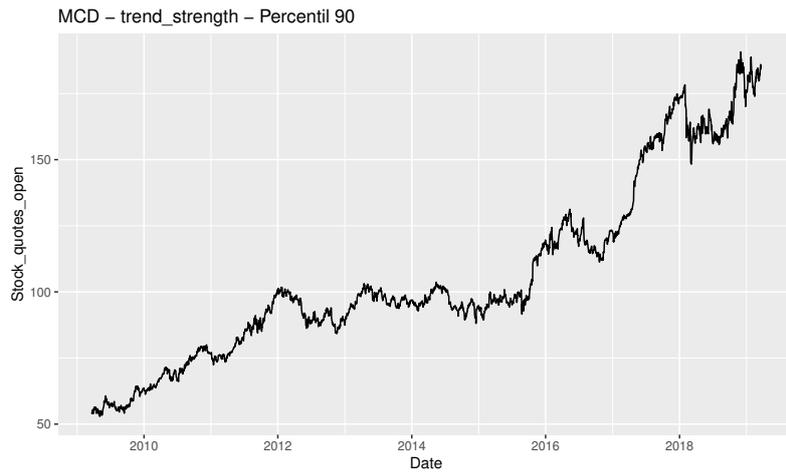


Figura 14: Valor correspondiente a percentil 90 en trend_strength

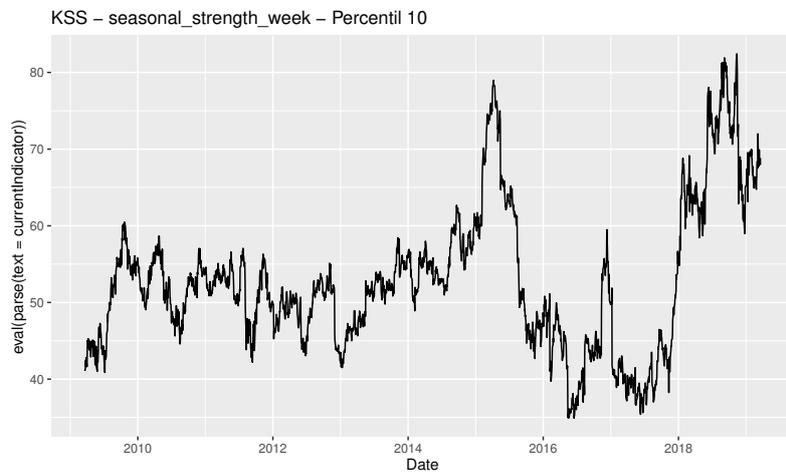


Figura 15: Valor correspondiente a percentil 10 en seasonal_strength_week



Figura 16: Valor correspondiente a percentil 50 en trend_strength

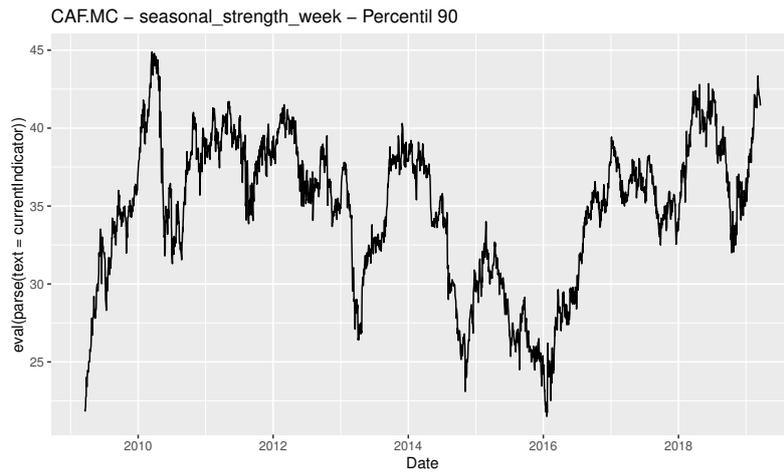


Figura 17: Valor correspondiente a percentil 90 en trend_strength

En la figura 12 se puede observar cómo no existe una clara tendencia. Sin embargo, ya en la figura 13 se nota una mayor fuerza de la tendencia, pese a que ésta sea negativa. Por último, se nota una fuerte tendencia en la figura 14. Estos resultados eran esperables dado que se tratan de percentiles crecientes, y a mayor percentil, mayor fuerza de la tendencia.

En cuanto a la fuerza de la estacionalidad semanal, se puede ver que no se nota apenas diferencias entre los percentiles 10, 50 y 90, mostrados en las figuras 15, 16 y 17. Esto nos indica que, por la naturaleza de las series, no hay efectos estacionales en periodos semanales. Para futuros desarrollos, puede ser interesante analizar periodos de estacionalidad mayores, ya que el valor de algunas compañías pueden depender de la estación en la que se encuentre, como por ejemplo, el consumo de crema solar es mayor en verano que en invierno, por lo que es posible que las compañías relacionadas con estos productos sufran estacionalidades de periodo un año.

Nótese que, hasta ahora, hemos estado tratando con solo dos indicadores de los 42 calculados. Para ver más indicadores en el mismo gráfico se presenta la figura 18, donde se muestra la matriz de correlación entre las principales características STL: fuerza de la tendencia y de la estacionalidad, el *spikiness*, la linealidad y la curvatura. Se ha evitado dividir los gráficos por sector o mercado, o añadir más variables dado que el gráfico era más difícil de leer.

De este gráfico, se pueden sacar varias conclusiones. La primera es que los indicadores están, en general, poco correlados entre sí. El mayor valor de correlación es de 0.5, entre la linealidad y el *spikiness*. También se puede ver que hay ciertas características que tienen valores muy dispares. Por ejemplo, las fuerzas de la tendencia y estacionalidad tienen valores entre 0 y 1 mientras que el *spikiness* tiene valores de entre 0 y 400, estando la mayoría comprendidos entre 0 y 20, y solamente 4 valores son superiores a 100. Algo parecido ocurre con la linealidad y la curvatura, donde hay valores que se vuelven muy extremos respecto al grueso de datos. Esto es importante dado que las “unidades” de los

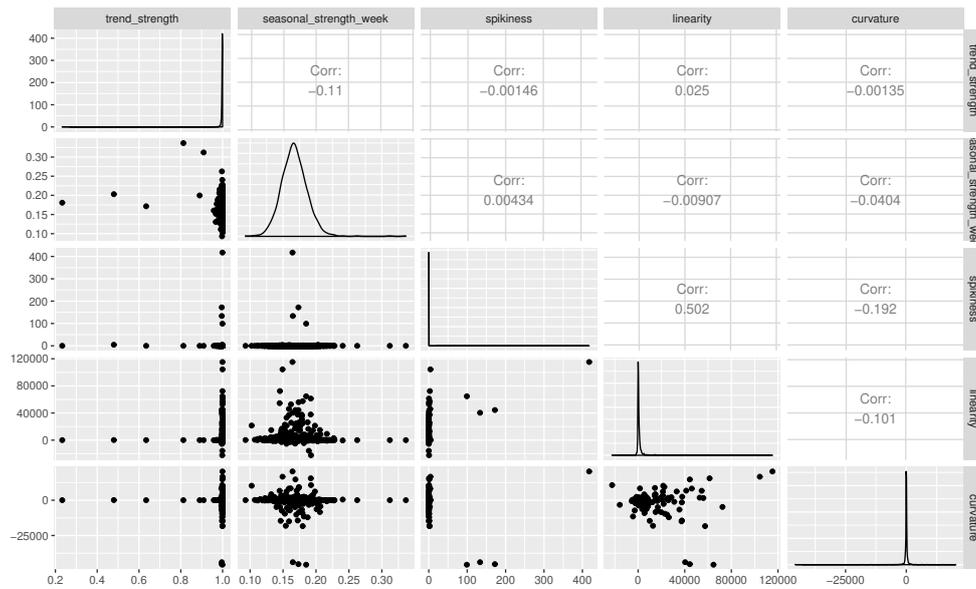


Figura 18: Matriz de correlación de las principales características STL

atributos son distintas y, por tanto, no son comparables.

De la figura 18 parece interesante analizar más en profundidad la relación entre las variables de linealidad y curvatura. Con este objetivo se muestra la figura 19. Se puede apreciar que la mayoría de las compañías de Londres tienen valores muy extremos de linealidad y curvatura. Esto se puede deber a su fuerte dependencia de una moneda que sufre inestabilidades, especialmente desde el Brexit, y, por lo cual, el precio de venta de las acciones se ve distorsionado. Por esta razón, puede ser interesante eliminar estas series de los análisis futuros o normalizar las características antes de realizar otros análisis, para evitar que sesguen los resultados obtenidos por estas grandes variaciones. Además, resulta interesante ver cómo elevados valores de curvatura (forma parabólica) y de linealidad son compatibles.

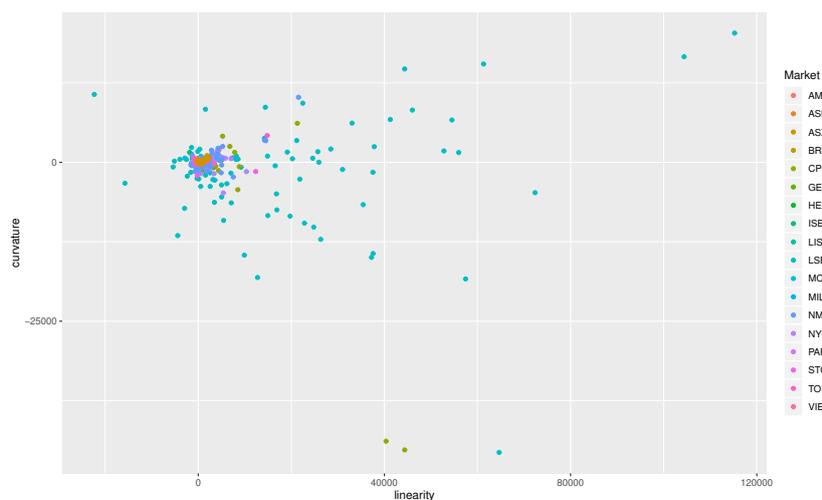


Figura 19: Relación entre curvatura y linealidad según los mercados

Se ha de añadir que esta matriz está incompleta dado que solo representa 5 atributos de los 42 disponibles. La matriz completa se muestra en la figura 22. De esta figura se puede tratar de enfocar algunos resultados, como por ejemplo la relación cuadrática entre las características de auto-correlación para un día y para la suma cuadrática de la función en los diez días anteriores para la serie diferenciada. Resulta sorprendente que esta relación no exista en misma la serie sin diferenciar. Esta relación se puede ver en la figura 20.

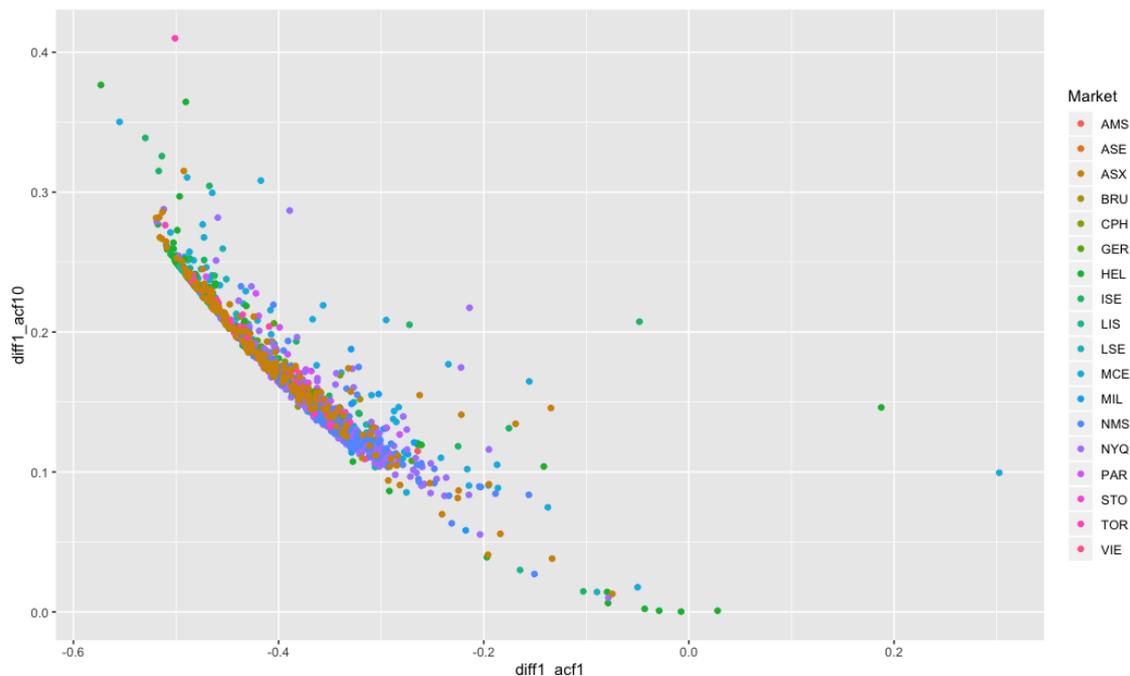


Figura 20: Comparativa de las variables diff1_acf y diff10_acf

En esta figura también se puede apreciar que hay dos compañías que tienen tres funciones de auto-correlación positivas. Si se analizan estas compañías en particular se descubre que son Altia, Gestamp y Valoe. Estas compañías se muestran en la figura 21.

Si se observa Valoe, resulta que tiene muchas oscilaciones diarias que pueden provocar ese valor positivo. Las otras dos compañías tienen pocos datos y puede que por ello tengan valores raros respecto al resto de compañías. Estas dos últimas podrían ser objeto de descarte a la hora de invertir dado el poco recorrido de la compañía en los mercados bursátiles, pero más por la ausencia de datos que por su novedad. Valoe puede ser también objeto de descarte dado que los patrones que han podido provocar la rareza de sus valores han dejado de tener lugar. Cabe indicar que un comportamiento así se puede deber a poco volumen de ventas, lo cual supone una falta de liquidez de las acciones de la compañía. Esto puede suponer ciertas dificultades a la hora de comerciar con ellas y, por tanto, puede restar interés a las acciones para determinados perfiles.

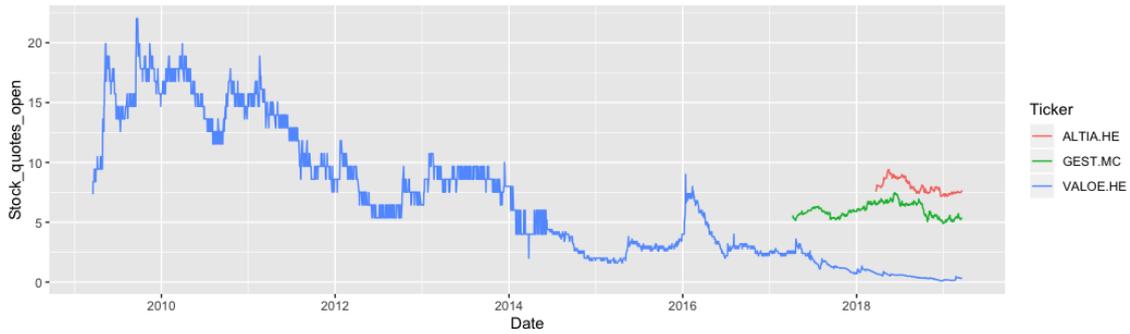


Figura 21: Compañías con diff_acf1 positivo (outliers)

Se considera que toda la información que se está aportando, procedente de análisis de tipo genérico para series temporales, ayuda en la toma de decisiones a la hora de invertir. En estos casos puede servir para señalar casos raros, que pueden sacar a la luz relaciones no tan sencillas de encontrar a través de análisis fundamental del mercado. Estos casos raros pueden señalar compañías que deben excluirse de los análisis o señalar compañías que pueden tener un especial interés económico. Sin embargo, existen tantas relaciones por analizar que resulta necesario realizar un análisis de componentes principales. No obstante, se presenta la matriz de correlación en la figura 22.

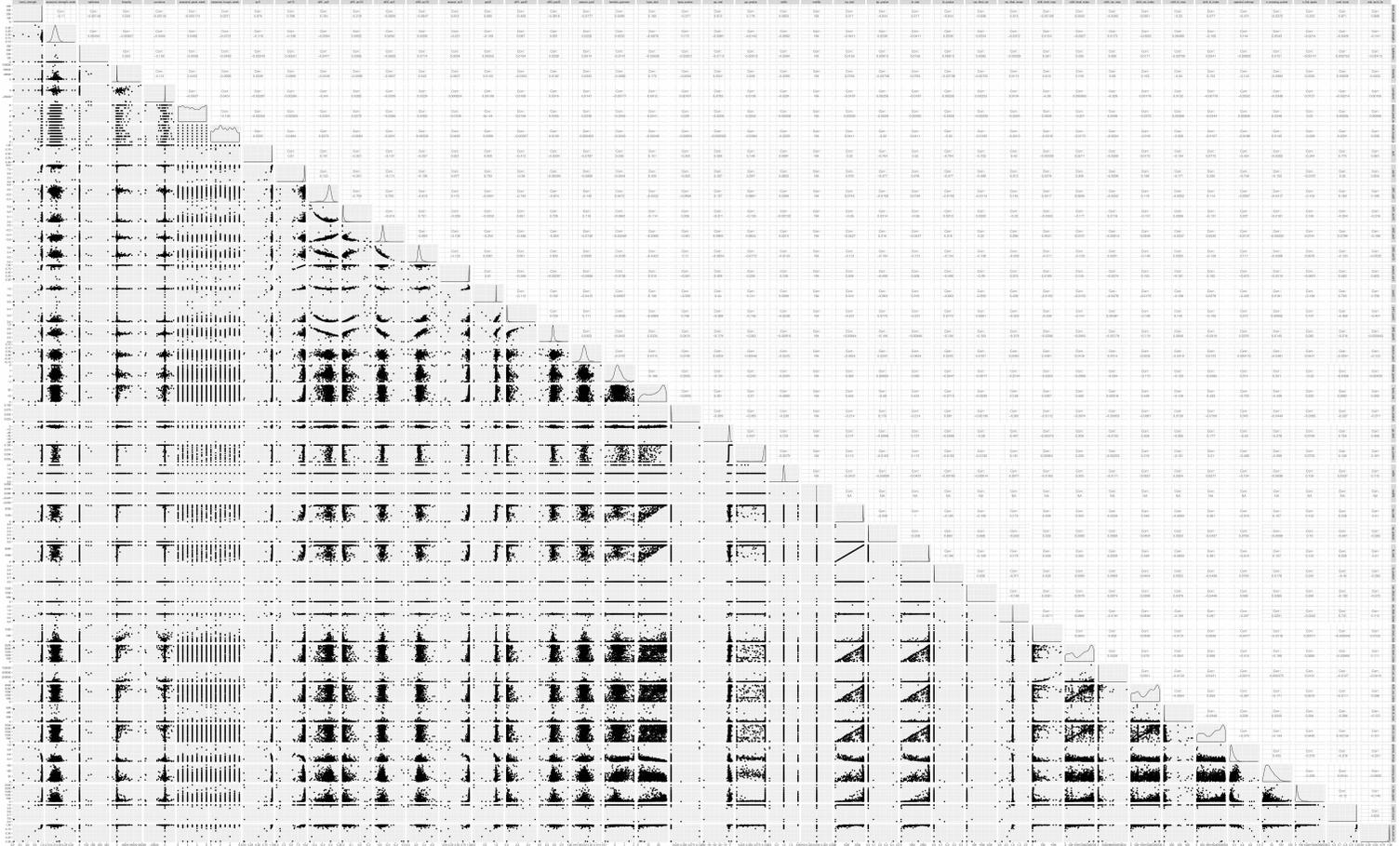


Figura 22: Matriz de correlación completa

4.2. Análisis de componentes principales

Dado el gran número de indicadores que hay, puede resultar conveniente realizar un análisis de componentes principales (PCA por sus siglas en inglés, *Principal Component Analysis*). Este análisis es una forma de identificar patrones en los datos, de forma que se puedan resaltar las similitudes y diferencias entre los datos. Esto resulta en la posibilidad de reducir el número de dimensiones con las que se trabaja (ahora mismo se tienen 42 dimensiones), ya que se eliminarían las variables que resultan menos explicativas. Esta técnica es muy utilizada cuando el número de variables de entrada es alto (nuestro caso de estudio) o cuando éstas están correlacionadas (aún se desconoce si aplica a este caso).

En este caso, se ha realizado este análisis, pero ha sido necesario quitar una serie de variables que dificultaban este análisis, como por ejemplo `arch_lm`, `pp_pvalue` y `nsdiffs` dado que daban problemas al algoritmo utilizado. Tras realizar este análisis se ha decidido mostrar las dos principales componentes en un gráfico, que se encuentra en la figura 23. Se han eliminado los valores que tenían unos valores excesivos en las componentes principales respecto al grueso de puntos (se ha puesto un límite de 20 en las primeras dos componentes)

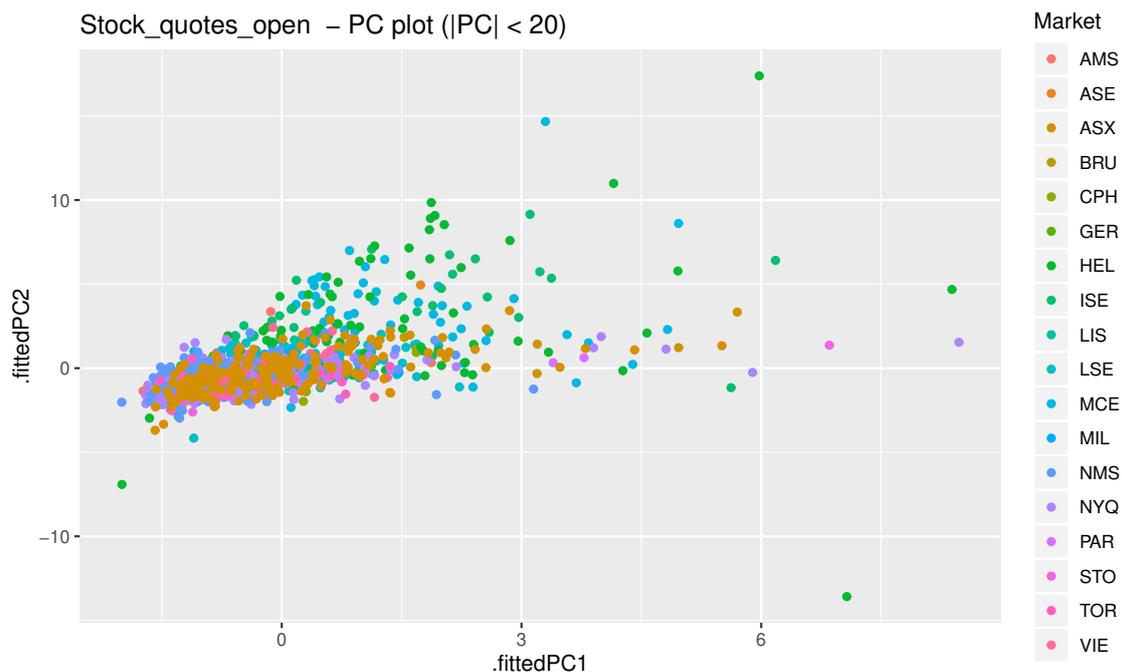


Figura 23: Valores proyectados sobre las componentes principales

En este caso, parece que en la representación sobre las dos componentes principales no se distingue zonas en las que se concentre algún mercado. Resulta interesante ver cuánta varianza explica cada componente principal. Para ello se muestra el gráfico de la figura 24 donde las barras representan el porcentaje de varianza explicada por cada componente. Este valor se refleja en el eje y de la izquierda. Por otro lado, la línea representa el porcentaje de la varianza explicado por todas las componentes principales hasta la actual (incluida). En este caso, su valor se ve reflejado en el eje y de la derecha.

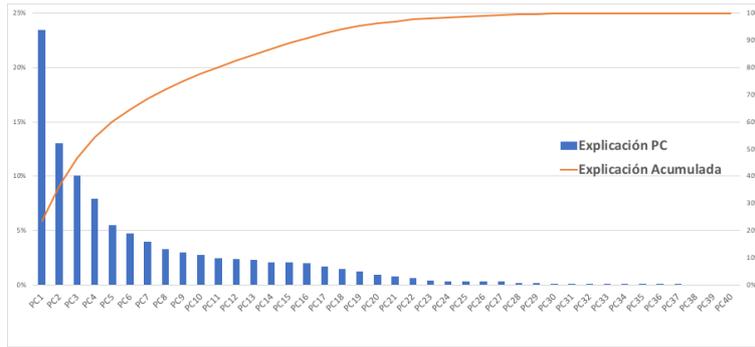


Figura 24: Explicación de la varianza de cada PC y acumulada

Se puede apreciar que entre las dos primeras componentes principales explican menos del 40% de la varianza de todas las variables. Esto puede parecer poca explicación, pero si se compara con el caso de que este análisis no aportara valor, estas dos variables explicarían tan solo el 5% de la varianza. También se puede ver que para explicar el 80% de la varianza sólo son necesarias 11 componentes. Por último, también se aprecia que, a partir de la componente 22, cada componente explica menos de un 1%. Esto nos estaría indicando que no es necesario trabajar con las 40 dimensiones que se han propuesto, sino que once combinaciones lineales de ellas bastarían para explicar al 80% la varianza.

Por otro lado, en la figura 25 se muestra la influencia que tiene cada indicador para cada una de las primeras dos componentes principales. En la figura 26 se muestra la influencia acumulada de las distintos indicadores en las primeras cinco componentes. Se puede ver que, en estas primeras cinco componentes, hay algunas variables que tienen prácticamente nula importancia, concretamente `seasonal_strength_week`, `seasonal_peak_week`, `seasonal_trough_week` (estas dos tienen de interés su naturaleza similar), `season_pacf` y `shift_ld_index`.

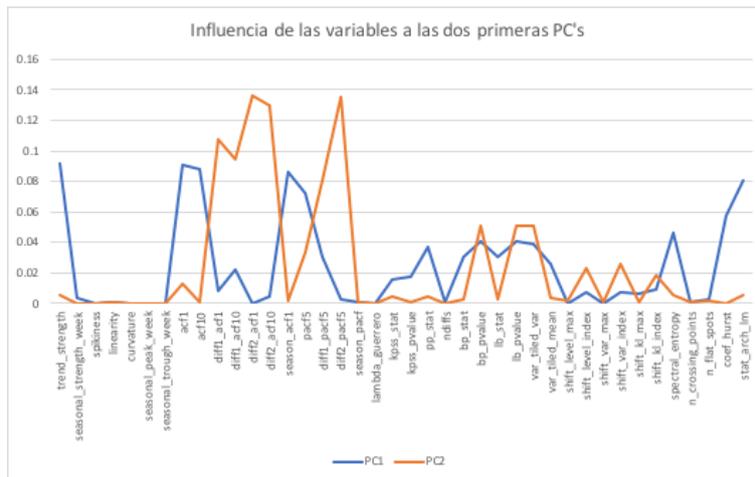


Figura 25: Influencia de cada indicador en las primeras dos PCs

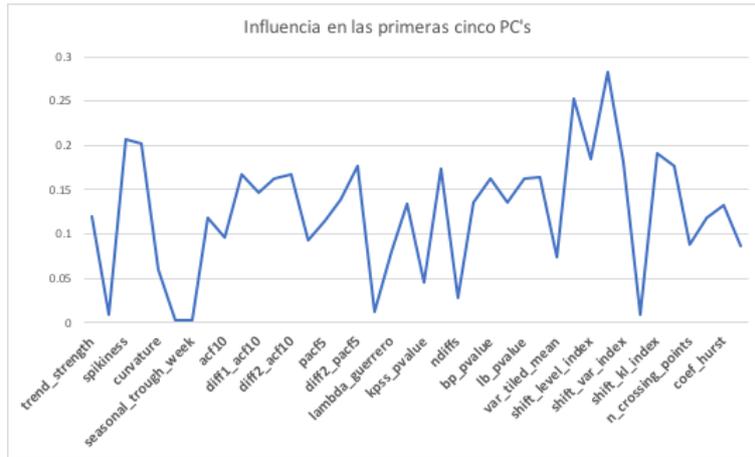


Figura 26: Influencia acumulada de cada indicador en las primeras cinco PCs

Un último aspecto para el PCA sería el gráfico biplot, que muestra en el mismo gráfico los puntos con las principales componentes, así como la proyección de los indicadores sobre las componentes. El gráfico biplot se muestra en la figura 27. En este gráfico se ven varios outliers, que parece que, pese a las componentes principales, se siguen desmarcando del resto de datos. Podría ser interesante hacer de nuevo este análisis con los datos normalizados.

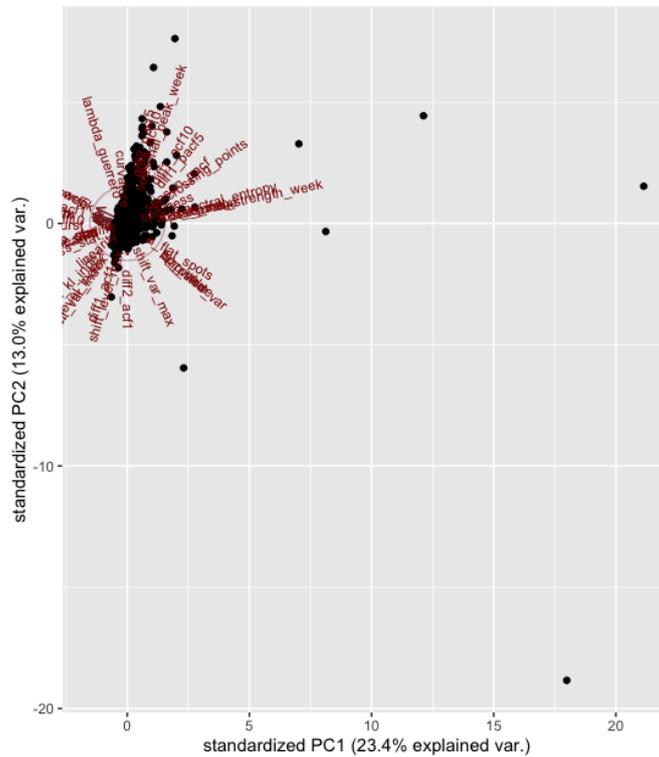


Figura 27: Gráfico biplot

Se considera de interés mostrar los percentiles extremos e intermedio de estas primeras dos componentes principales. Para ello se muestran las figuras 28 a 33.

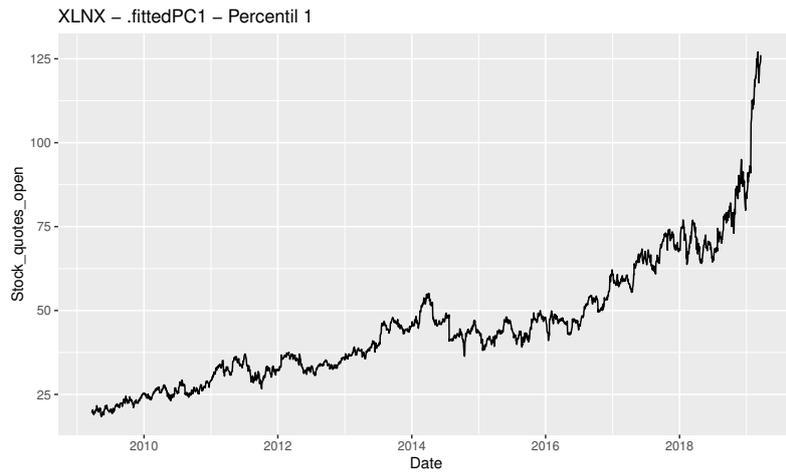


Figura 28: Compañía correspondiente al percentil 1 de la primera PC

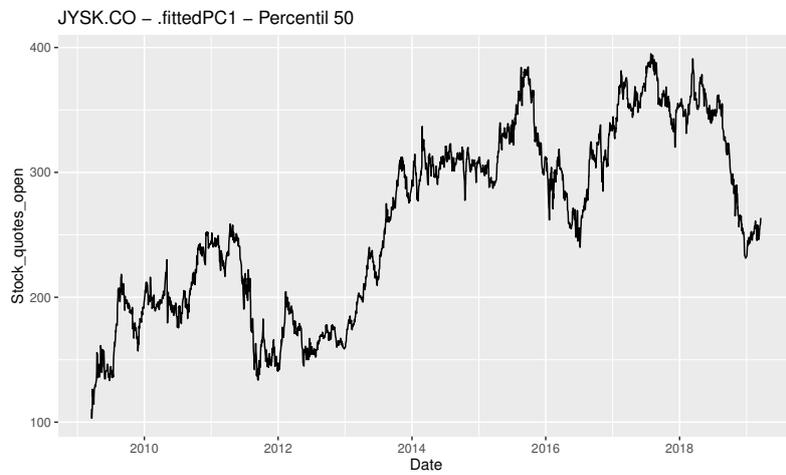


Figura 29: Compañía correspondiente al percentil 50 de la primera PC

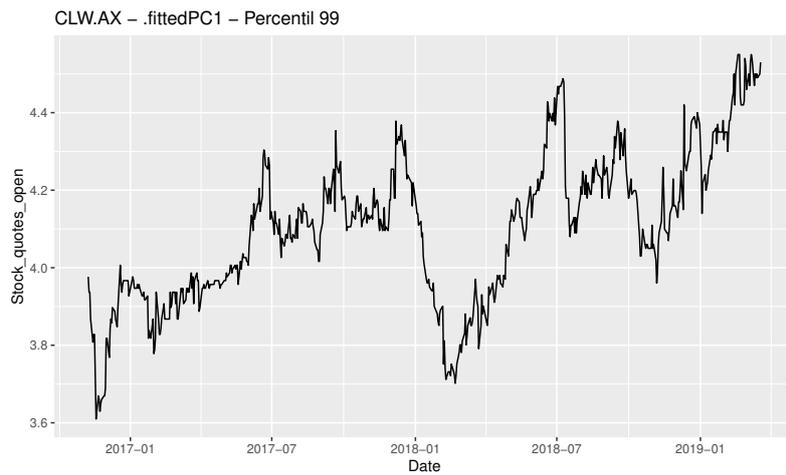


Figura 30: Compañía correspondiente al percentil 99 de la primera PC

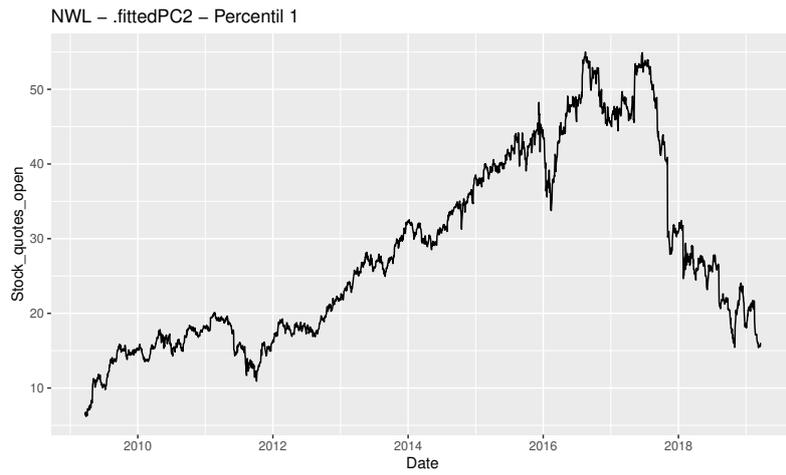


Figura 31: Compañía correspondiente al percentil 1 de la segunda PC

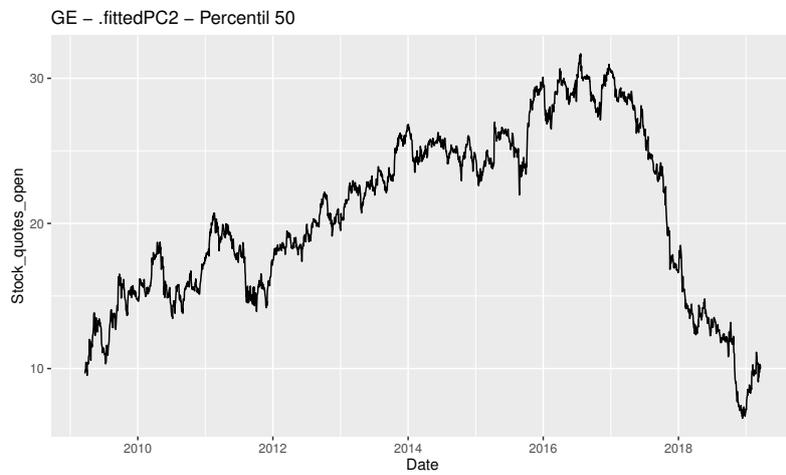


Figura 32: Compañía correspondiente al percentil 50 de la segunda PC

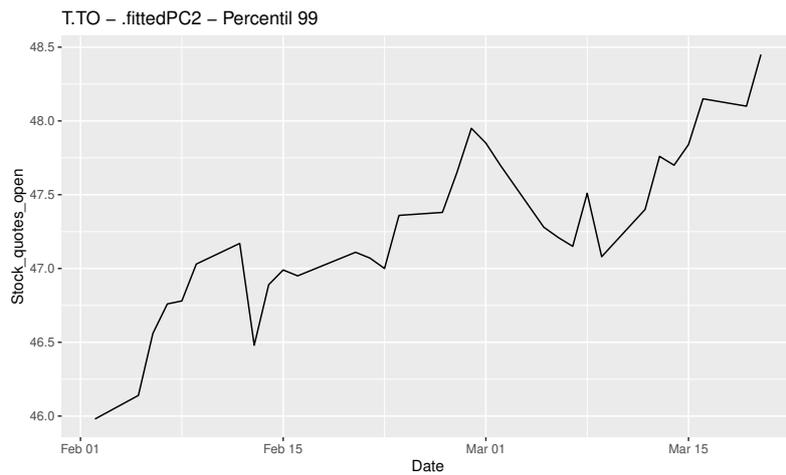


Figura 33: Compañía correspondiente al percentil 99 de la segunda PC

En la figura 28 se puede apreciar que la evolución de la acción tiene una fuerte tendencia. Si se vuelve a mirar la figura 25 se observa que la primera componente está muy influida por la fuerza de la tendencia y por el estadístico de PP (de Phillips y Perron). Estos dos tienen en común que están relacionados con la tendencia. Además, cabe destacar que se trata del percentil 1, por lo que corresponde a los valores más bajos. Esto se debe a que estas características tienen un valor negativo en la componente, como se puede intuir en el gráfico biplot. Sin embargo, si se observan las figuras 29 y 30 se puede apreciar que la componente de tendencia se ve notablemente reducida, aunque siempre se puede apreciar una tendencia positiva. Esto se puede deber al crecimiento generalizado del mercado por ciertas características macroeconómicas, tales como la inflación y el crecimiento económico del país.

En cuanto a la segunda componente principal se puede apreciar que ambas compañías mostradas en las figuras 31 y 32 tienen fuertes caídas a partir de la segunda mitad de 2017. Sin embargo, esta característica se trata de algo trivial, ya que no todos los percentiles intermedios muestran esta característica, mostrándose un ejemplo en la figura 34. Por otro lado, en la figura 33 lo que se puede observar es que se trata de una compañía con insuficientes datos. Por esta razón se presenta el percentil 95 en la figura 35.

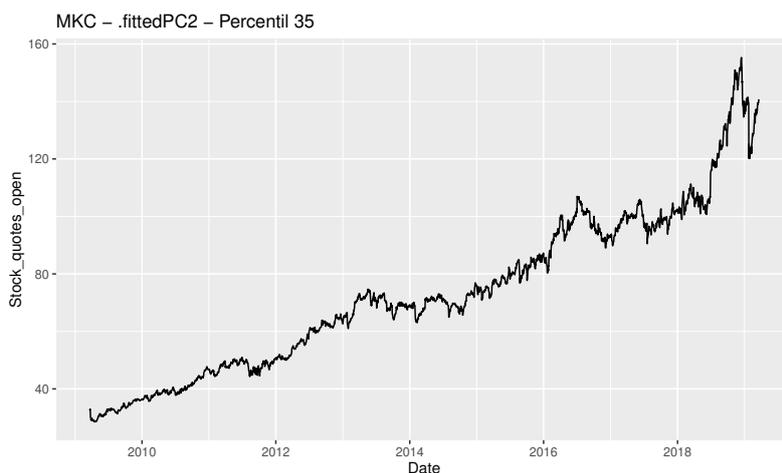


Figura 34: Compañía correspondiente al percentil 35 de la segunda PC

Continuado el análisis, si se vuelve a observar la figura 25 se observa que esta componente principal tiene grandes relaciones con la serie diferenciada. Por esta razón, se ha considerado interesante mostrar las series diferenciadas una y dos veces para tratar de descubrir la relación que existe entre estas series. Estas series diferenciadas se muestran en las figuras 36 a 41.

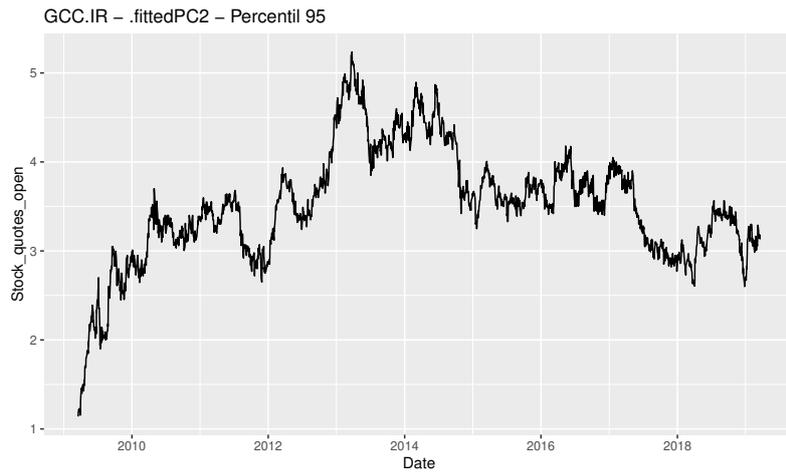


Figura 35: Compañía correspondiente al percentil 95 de la segunda PC

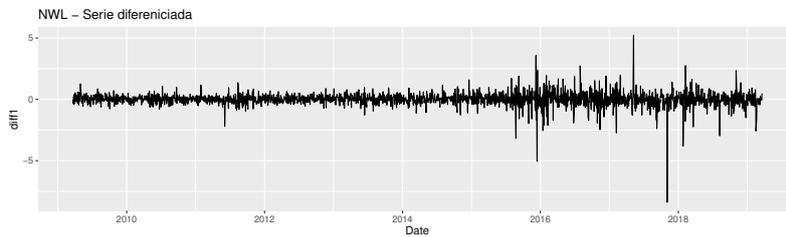


Figura 36: Compañía diferenciada correspondiente al percentil 1 de la segunda PC

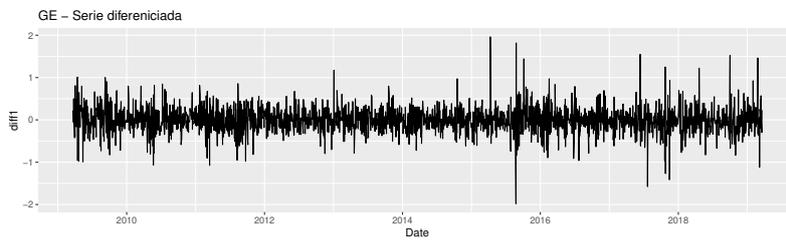


Figura 37: Compañía diferenciada correspondiente al percentil 50 de la segunda PC

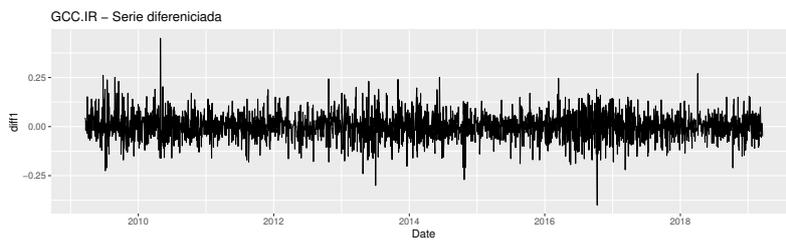


Figura 38: Compañía diferenciada correspondiente al percentil 95 de la segunda PC

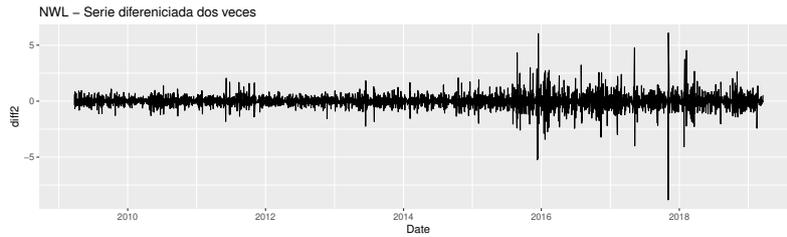


Figura 39: Compañía doblemente diferenciada correspondiente al percentil 1 de la segunda PC

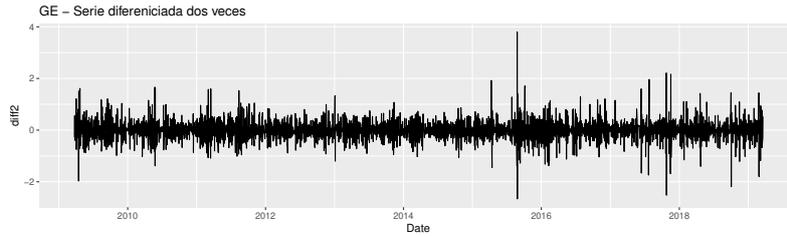


Figura 40: Compañía doblemente diferenciada correspondiente al percentil 50 de la segunda PC

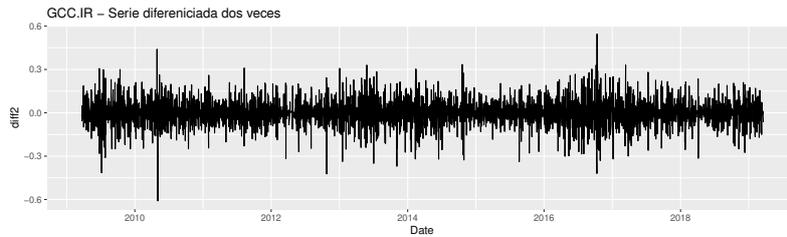


Figura 41: Compañía doblemente diferenciada correspondiente al percentil 95 de la segunda PC

En estas figuras se puede observar que, a mayor percentil, mayor es el ruido de la serie diferenciada (una o dos veces). Además, dado que se sabe esta gran correlación, se puede intuir que los percentiles más bajos tienen mayores auto-correlaciones en estas series diferenciadas. Esto es muy difícil de apreciar en una observación directa de la serie original y esta metodología lo saca a relucir. Estas medidas pueden estar relacionadas con el riesgo. Resulta interesante ver que las dos principales conclusiones sacadas por el análisis de componentes principales son las relaciones de tendencia, que pueden relacionarse con el rendimiento económico; y la relación con el nivel de ruido de la serie, que puede relacionarse con el riesgo. Esto resulta muy interesante dado que son las principales métricas utilizadas a la hora de invertir.

4.3. Clusterización

Como se indica en la figura 1, la clusterización (o *clustering*) es una forma de aprendizaje automático no supervisado. Este clustering se puede hacer con los principales indicadores que se han calculado. Se ha realizado este clustering con todas los indicadores a

través del algoritmo *kmeans*. En la figura 42 se muestra las series sobre las componentes principales. Se definieron 3 clústers siguiendo el método del codo, mostrado en la figura 43.

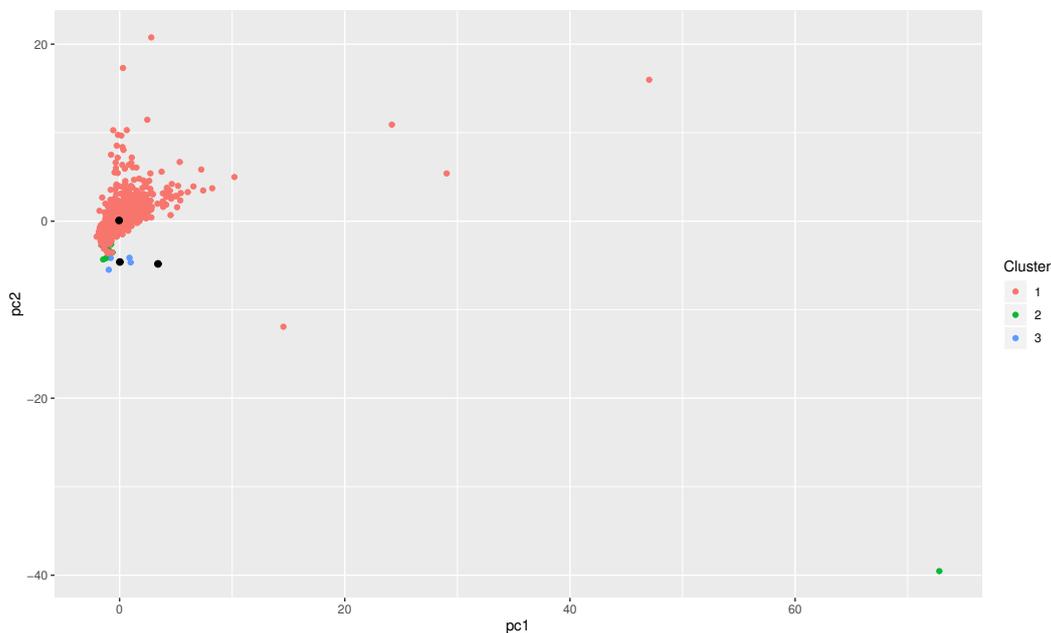


Figura 42: Clustering de las series mostrado sobre las componentes principales

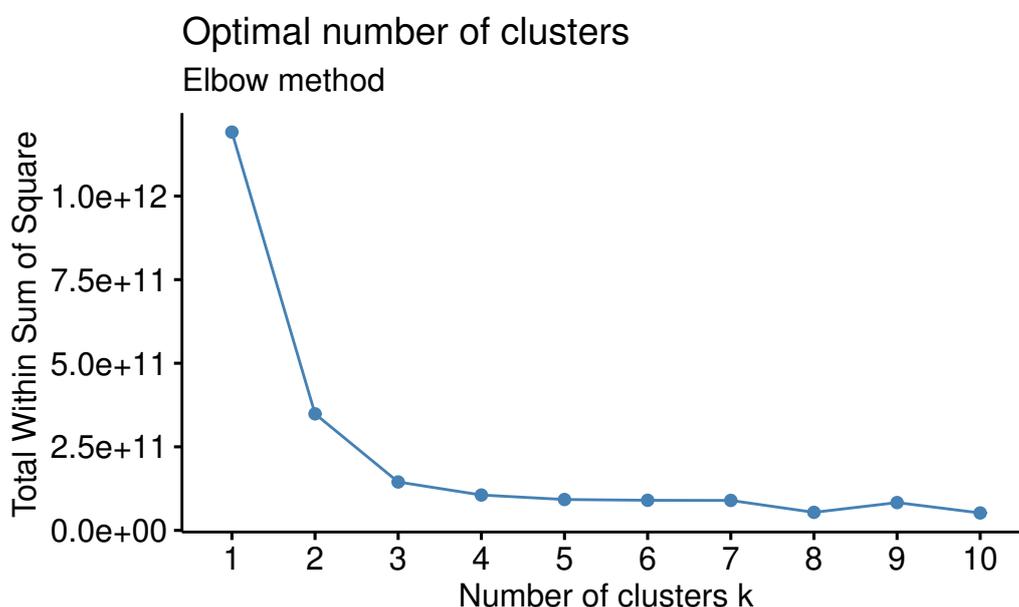


Figura 43: Método del codo para definir el número de clústers

Se puede apreciar en la figura que la mayoría de las series se encuentran en el clúster #1. Para mostrar este fenómeno, se muestra un histograma de escala logarítmica en la figura 44. Esto nos lleva a pensar que puede haber una serie de datos excesivamente raros o destacados que hacen que lleve los centroides de los clústers a los extremos. Por ello

se realiza una representación de los centroides, mostrada en la figura 45. En esta figura se corrobora que estos clústers agrupan compañías con valores “raros”. También resulta difícil ver sobre estas componentes qué centroides (en negro) pertenecen a los clústers 2 y 3. Otro dato curioso es que el valor más extremo en la PC1 pertenece al clúster 2 en vez de al 1 (como lo hace una compañía que se encuentra a “mitad de camino” entre la mayoría de los datos y este *outlayer*).

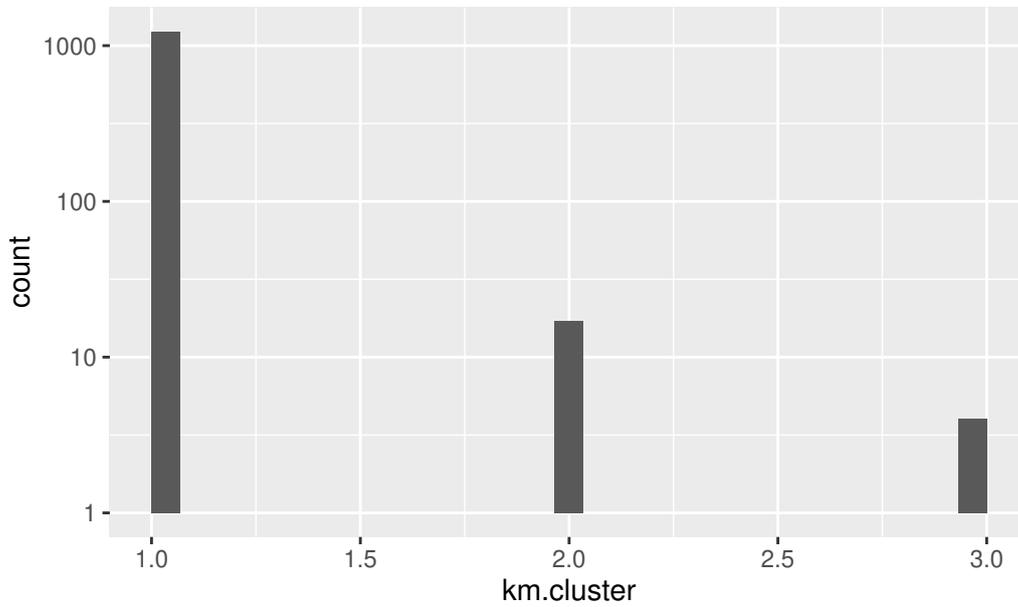


Figura 44: Número de compañías dentro de cada clúster

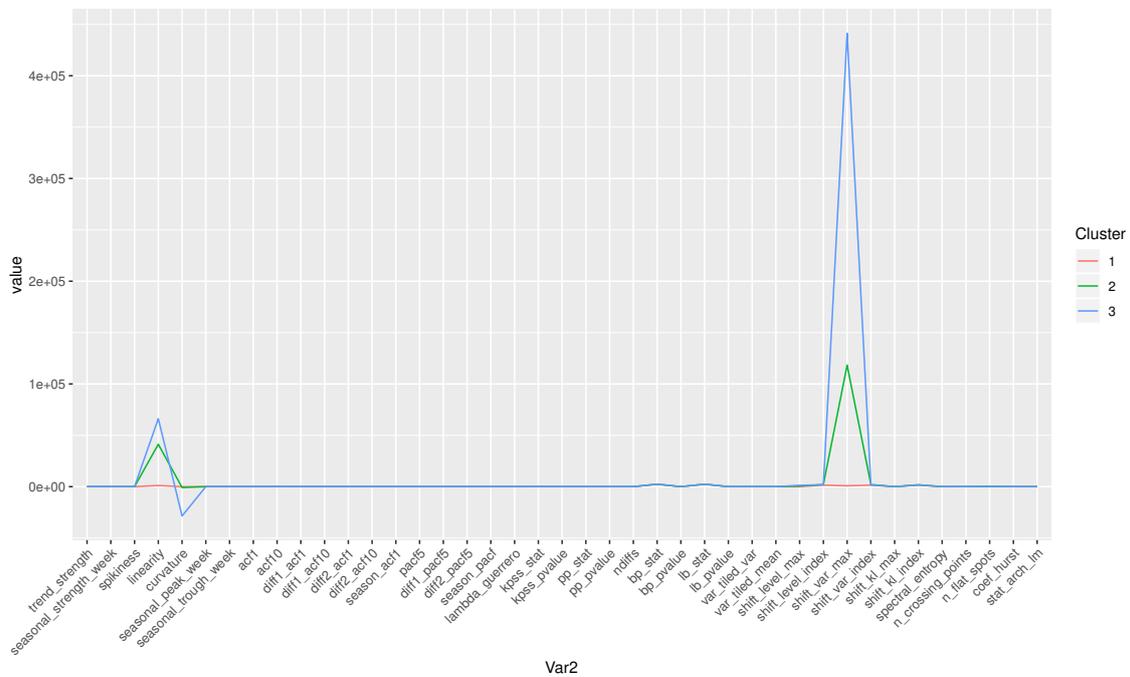


Figura 45: Coordenadas de los centroides

Con esta información, se decide investigar qué valores son éstos y si siguen algún patrón. Para ello, se suponen dos hipótesis: o bien se tratan de mercados raros o de sectores raros. La distribución de los clúster por mercado y sector se muestran en las tablas 4 y 5 respectivamente. En estas tablas se puede apreciar que todas las compañías de los clústeres 2 y 3 pertenecen a los mercados de Londres (LSE) y Copenhage (CPH). También se puede ver que el sector influye poco en que se esté en un clúster u otro. En la primera se puede observar cómo en las componentes principales 1 y 2 sí influyen en la clusterización, siendo los valores de mayor componente 1 los pertenecientes al clúster #3. Sin embargo, en la segunda figura se puede ver que aún quedan compañías con valores raros, recogidas en el clúster #1. Es muy probable que estas fueran excluidas de los primeros clústeres “*extraños*” (2 y 3 en la primera clusterización) debido a que sus valores no son tan extremos y caían en el clúster principal.

Mercado\Cluster	1	2	3	Grand Total
AMS	17			17
ASE	2			2
ASX	190			190
BRU	11			11
CPH	16	2		18
GER	30			30
HEL	113			113
ISE	41			41
LIS	16			16
LSE	70	17	2	89
MCE	117			117
MIL	27			27
NMS	130			130
NYQ	323			323
PAR	35			35
STO	27			27
TOR	53			53
VIE	17			17
Grand Total	1235	17	4	1256

Tabla 4: Distribución de los clúster por mercado

Sector\Cluster	1	2	3	Grand Total
Basic Materials	118	3		121
Communication Services	35			35
Consumer Cyclical	179	5	2	186
Consumer Defensive	94	2		96
Energy	67	1		68
Financial Services	155	2		157
Healthcare	97	1		98
Industrials	185	1	2	188
NA	84	1		85
Real Estate	58			58
Technology	113	1		114
Utilities	50			50
Grand Total	1235	17	4	1256

Tabla 5: Distribución de los clúster por sector

Antes de realizar ninguna acción con estas compañías puede resultar interesante observar la evolución de estas compañías en el tiempo. Por esta razón se muestra la figura

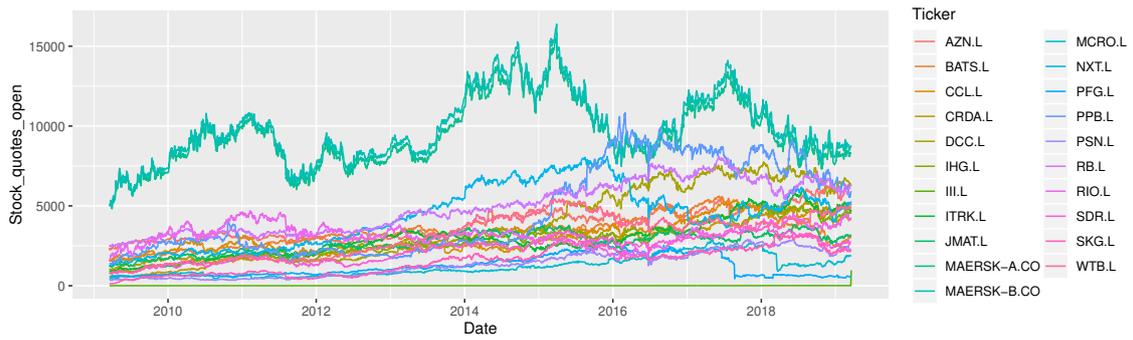


Figura 46: Evolución de los valores de apertura de las compañías de los clústers 2 y 3

46. En esta figura se puede ver cómo estas compañías no parecen tener nada en común al analizarse visualmente. Cabe destacar que los precios son muy elevados. También existen bajadas más abruptas (en ciertas ocasiones). Será interesante analizar las rentabilidades que puedan dar estas compañías.

Para evitar la influencia de estas compañías en la clusterización, se quitan estas compañías y se vuelve a clusterizar. Se obtienen la distribución mostrada en la figura 47 y los valores de los centroides se muestran en la figura 48.

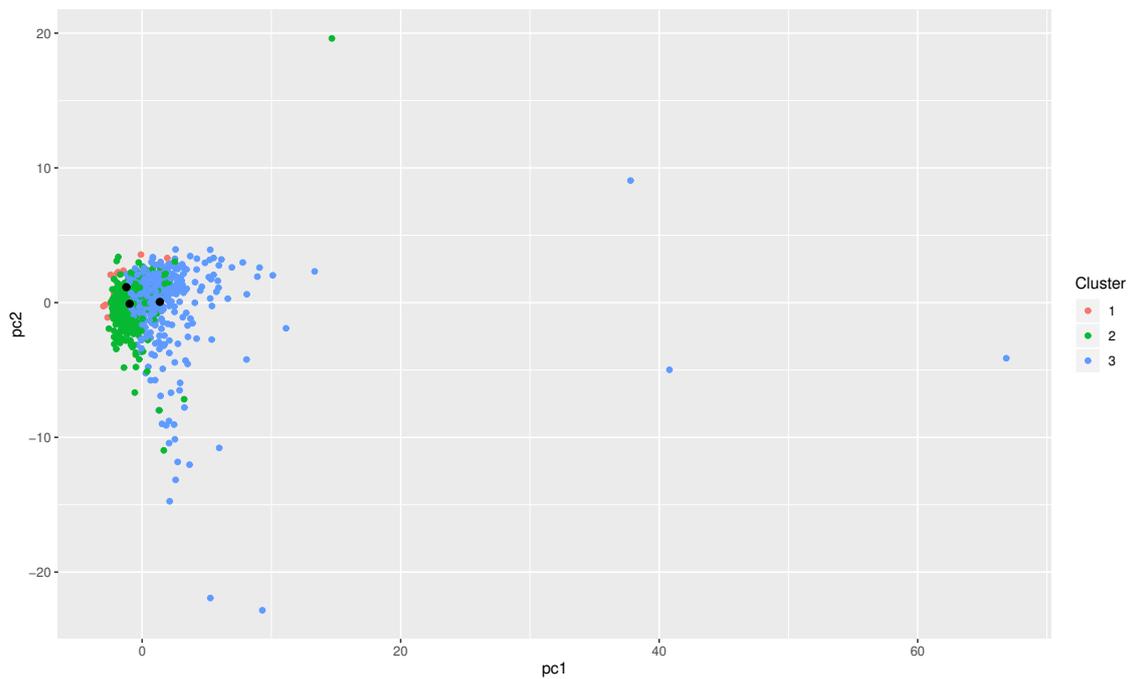


Figura 47: Nueva clusterización para el valor de apertura

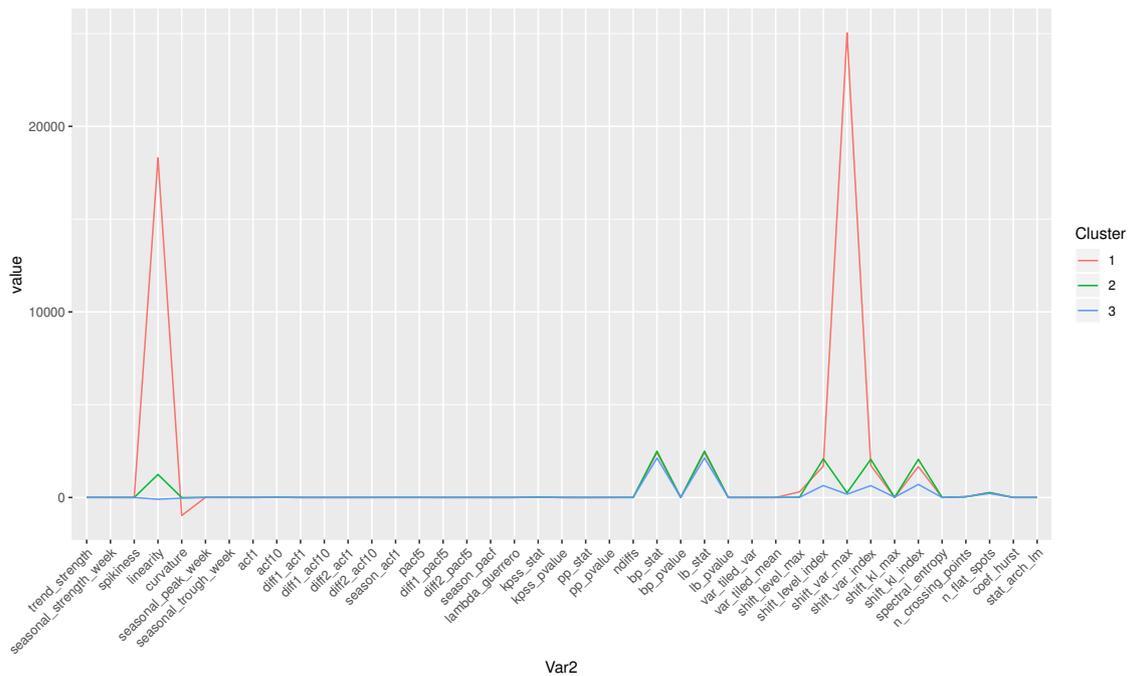


Figura 48: Centroides de la nueva clusterización para el valor de apertura

Pese a haber quitado las compañías con los valores más extremos, se puede observar que las características siguen teniendo valores muy extremos. Por esta razón se ha decidido que la clusterización se realice con los datos normalizados. De esta forma, se unificarán las unidades de todas las características.

Con estos datos normalizados, se vuelve a calcular el número de clústers necesarios a través del método del codo, resultado éste en nueve clústers. Los resultados de esta clusterización se muestran en la figura 49.

En estas figuras se observa que hay unos clústers que abarcan la mayoría de los valores que antes se calificaron de *raros*. Se puede observar que hay un clúster, el 7, con valores realmente extremos respecto al resto. Esto se observa claramente en las coordenadas del centroide. Por otro lado, hay algunos clústers que también tienen valores extremos, como el 6, el 8 y el 9. Si además se observa la tabla 6, se aprecia que estos clústers son precisamente los que menos integrantes tienen. Se ha decidido mostrar unass muestras (20 como máximo para que las gráficas fueran legibles) de cada uno de estos clústers con pocos integrantes. Más adelante se discutirá si estos clústers (6, 7, 8 y 9) merecen un trato especial.

<i>Clúster</i>	1	2	3	4	5	6	7	8	9
<i>Integrantes</i>	166	195	176	360	180	99	3	60	17

Tabla 6: Número de integrantes de cada cada clúster

En la figura 51 se aprecian algunas rarezas en las series, como que son cortas o con unos saltos muy grandes. En la figura 52 se aprecia claramente que tienen unos valores

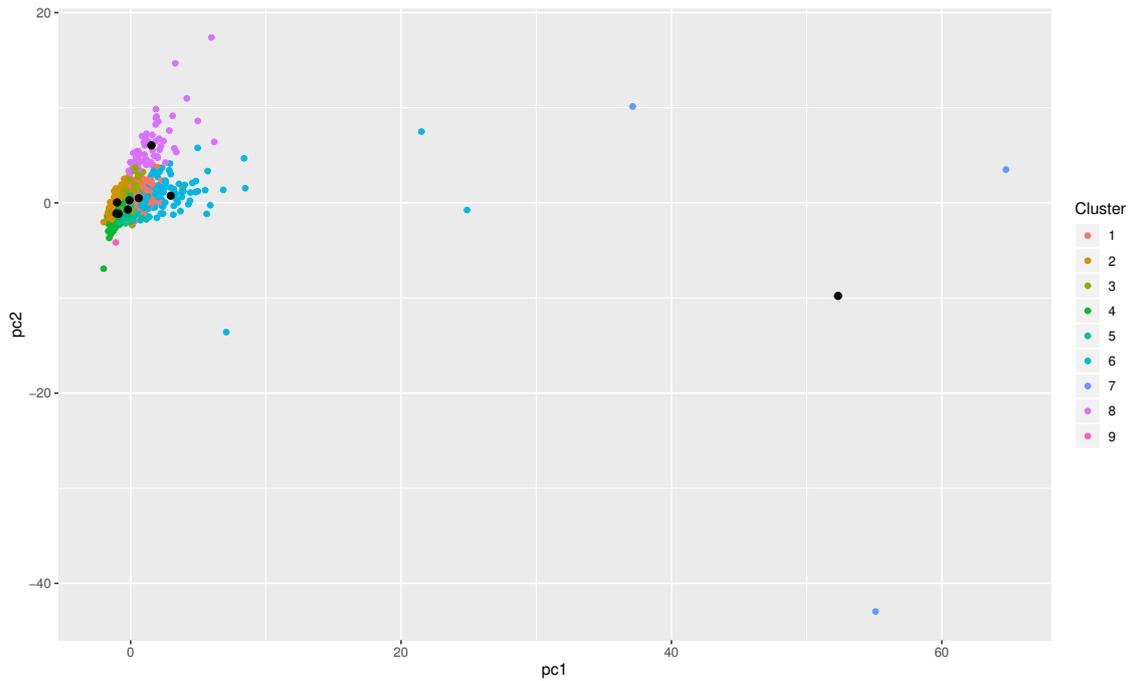


Figura 49: Clústers normalizados

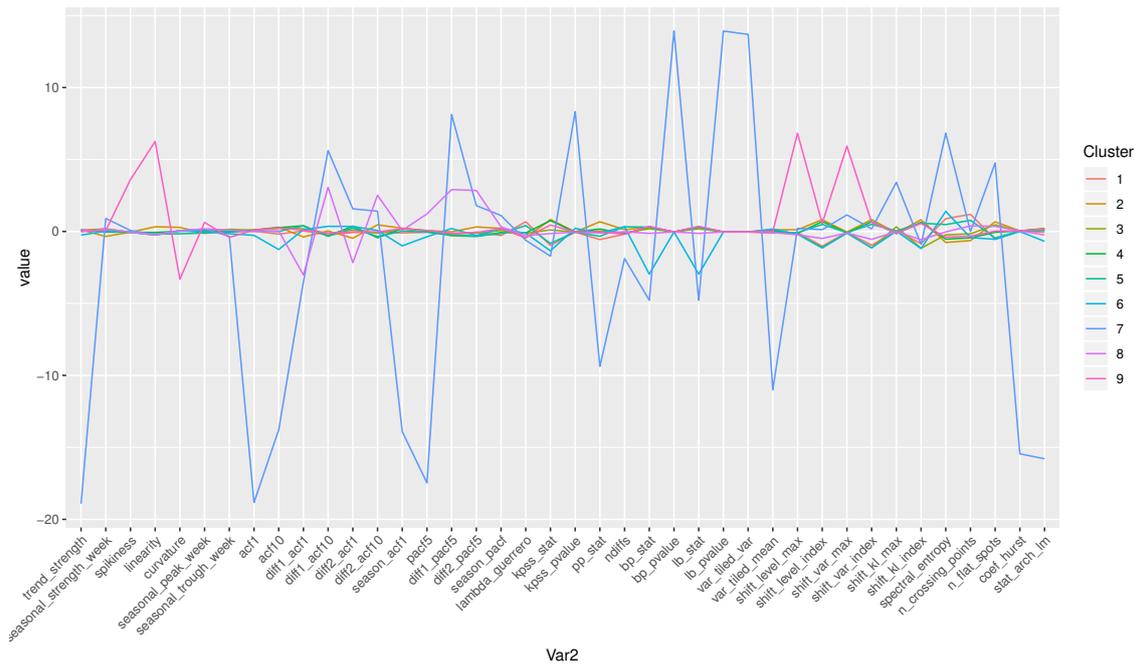


Figura 50: Centroides de los clústers normalizados

anómalos, muy superiores al resto de la serie. En las figuras 53 y 54 no se aprecian características especiales de las series a simple vista, ni similitudes claras entre ellas. Es probable que las características y el algoritmo de clusterización vean cosas que nosotros no somos capaces de ver, dado que se centran en las dinámicas internas de las series.

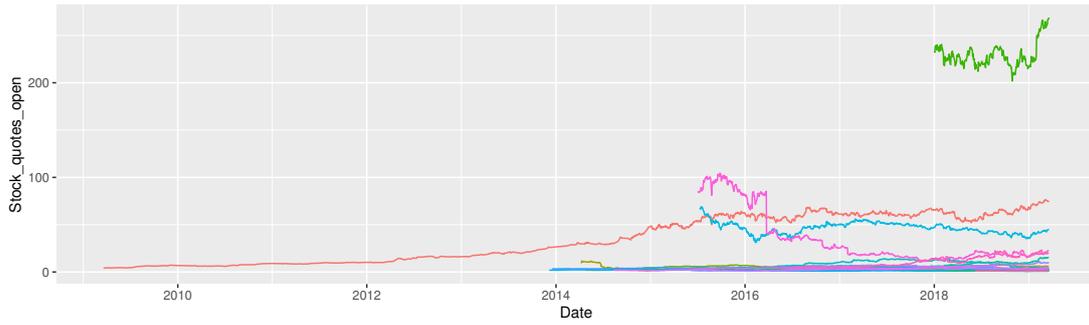


Figura 51: Muestra de curvas del clúster 6

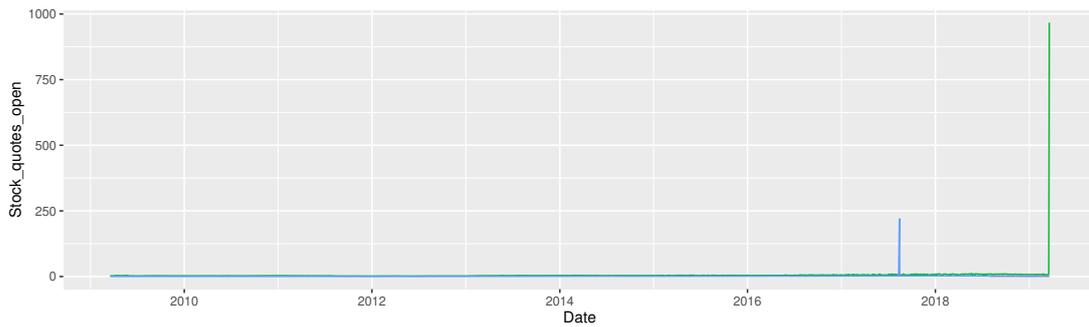


Figura 52: Muestra de curvas del clúster 7

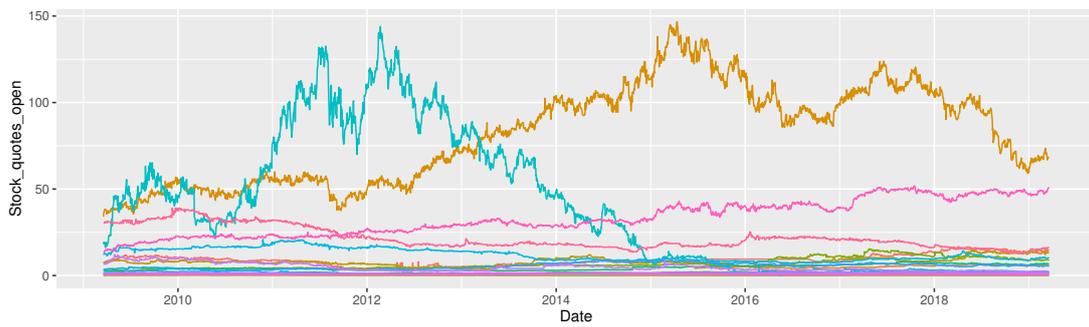


Figura 53: Muestra de curvas del clúster 8

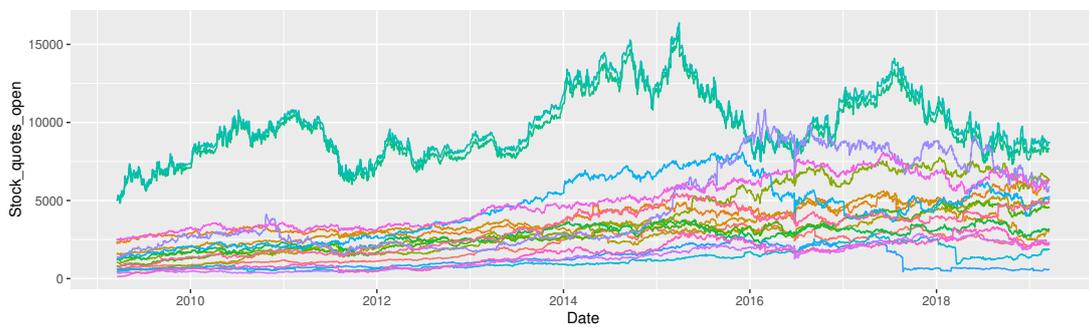


Figura 54: Muestra de curvas del clúster 9

Por otro lado, puede resultar de interés evaluar los rendimientos a lo largo de todo el periodo. Para ello se va a seguir una estrategia que se mostrará más adelante, llamada RSI. Se evaluarán los rendimientos económicos obtenidos de estas compañías a lo largo de todo el periodo para tres estrategias, de más a menos agresivas. La estrategia más agresiva se representa con el valor 70 y la menos agresiva se representa con el valor 90. El rendimiento se calcula como el incremento de capital a lo largo de la estrategia. Los resultados de estas estrategias se muestran en la figura 55, en la forma de un gráfico de cajas y bigotes. Se han quitado del gráfico las compañías con rendimientos superiores a 10.

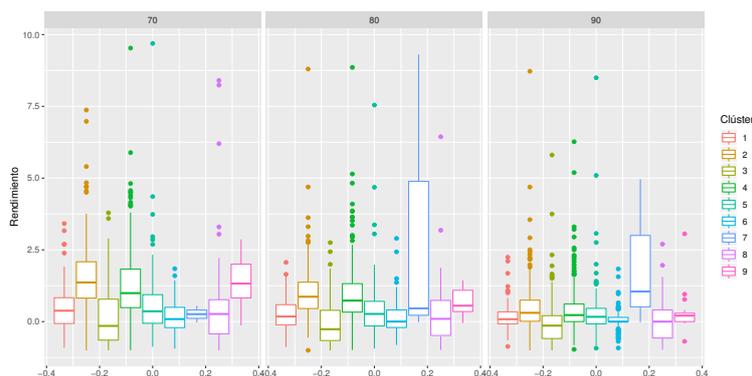


Figura 55: Rendimientos de cada clúster

De este gráfico se pueden extraer algunas conclusiones. La primera es que los clústers 2 y 4 tienen rendimientos superiores, en general, que los clústers 1 y 3, sin importar la estrategia. El clúster 5 también da mejores resultados que el 1 y el 3, pero éstos no son tan claros. Respecto a los clústers 6 al 9, cabe destacar que son los más sensibles al nivel de agresividad. Con esto se quiere decir que son los que más varían sus rendimientos al cambiar la agresividad de la estrategia. Un caso significativo respecto a estos clústers es el de clúster 7, el cual, independientemente de la agresividad, no obtiene rendimientos negativos. Con estos resultados, y con vistas a futuro, parece que sería interesante invertir en compañías cuyas características caigan sean más similares a las de los centroides de los clústers 2 y 4. Analizando las principales diferencias entre los centroides de los clústers 1 y 3 con los de los clústers 2 y 4, las principales diferencias encontradas son los valores de la estadística KPSS, del estadístico de Phillips Perron, de la fuerza de la tendencia y del `shift_var_index`.

Por otro lado, las parejas de compañía y estrategia que han dado rendimientos superiores a 10 se muestran en la tabla 7. En esta tabla se aprecia que una compañía del clúster 8 (*NEO1V.HE*) da unos rendimientos desorbitados independientemente del nivel de agresividad de la estrategia. Las otras dos compañías que hay en esta tabla dan rendimientos de 27 y 14, perteneciendo a los clústers 7 y 5. Resulta interesante ver que, de 5 entradas en la tabla, 4 pertenecen a los clústers cuyos centroides tenían valores extremos.

Ticker	uSup	uInf	Rendimiento	nCluster
III.L	70	30	27.3	7
NEO1V.HE	70	30	17815.6	8
NEO1V.HE	80	20	93881.2	8
IIA.VI	80	20	14.6	5
NEO1V.HE	90	10	5093.1	8

Tabla 7: Compañías y estrategias con un rendimiento superior a 10

4.4. Conclusiones

En este capítulo se ha realizado un análisis exploratorio de los datos que ha permitido entender los patrones que tienen estos datos, descubriendo algunas dinámicas internas de los datos, independientemente de las importantes diferencias de nivel que tengan las series entre sí. Gracias a la metodología propuesta se han podido observar que el conjunto de datos contiene algunos patrones.

En primer lugar se ha descubierto que las compañías del mercado inglés tienen unos valores de curvatura y linealidad mucho más extremos que el resto de mercados. Esto les puede dotar de algunas características que harán que los valores se comporten de forma distinta, por lo que merecerán un análisis individualizado con respecto al resto de datos. También se ha descubierto que algunas series pueden presentar valores distintos del conjunto por tener pocos datos o comportamientos anómalos por reducida liquidez en el mercado.

En segundo lugar, gracias al análisis de componentes principales se ha podido apreciar que, en este tipo de series, existen grandes correlaciones entre las características de fuerza de la tendencia y el estadístico de Phillips y Perron, y entre las funciones de auto-correlación de las series diferenciadas una y dos veces. Por esta razón puede tener sentido, en el futuro, agrupar estas variables. También se ha descubierto que hay algunas características que aportan muy poco valor. En concreto resulta que las que tienen que ver con la estacionalidad de la serie aportan especialmente poco en este análisis. Lo cierto es que sería curioso encontrar estacionalidades en los valores bursátiles a lo largo de una semana. Es por ello que se debería considerar usar un periodo de estacionalidad más amplio, como por ejemplo un año, antes de descartar este tipo de indicadores. También resulta interesante que los indicadores de las dos primeras componentes pueden ser relacionados, de forma independiente, con las dos principales métricas consideradas a la hora de invertir en los mercados bursátiles: rendimiento y riesgo.

Por último, se ha descubierto que hay una serie de compañías que tienen valores muy distintos al resto de las series. Resulta, además, que coinciden con las compañías inglesas que tenían valores extremos de linealidad y curvatura. Por otro lado, se ha realizado una clusterización con los valores de los indicadores normalizados, encontrando que en estos clústers se encuentran diferencias significativas entre los clústers en relación a los rendimientos obtenidos a través de distintas estrategias.

Esta sección se ha abordado el análisis exploratorio de los datos, acompañado de un aprendizaje no supervisado de los mismos, encontrando patrones que puedan ayudar a su correcta interpretación, entendiendo mejor las dinámicas internas que marcan el comportamiento de las series. En la siguiente sección se desarrollará un aprendizaje supervisado, en la que se buscará obtener rédito de los análisis, por lo que se aplicará a estrategias específicas de inversión.

5. Modelo de inversión

Un problema que surge del análisis realizado anteriormente es que no existe una relación directa de las características con la rentabilidad que se puede obtener de las acciones. Es por ello que se propone relacionar estas características con la rentabilidad. En este contexto se propone encontrar estas relaciones a través de un aprendizaje supervisado, aplicado sobre una estrategia de inversión. Para ello se propone una transformación que permite definir la estrategia: el índice de fuerza relativa (RSI por sus siglas en inglés: *Relative Strength Index*).

5.1. Cálculo y uso del RSI

Como indica James Chen en [Jam20], el RSI es un indicador de momento y es ampliamente utilizado en el análisis técnico. Este indicador mide la magnitud de las variaciones de precio para determinar si una acción está sobre-vendida o infra-vendida. Este indicador fue introducido por Welles Wilder (Junior) en el 1978. El objetivo de este indicador es dar señales de compra (cuando una acción está infra-vendida) y de venta (cuando está sobre-vendida). El valor del RSI oscila entre los valores de 0 y 100. El valor 100 implica que el valor del mercado es superior al de la acción y por lo tanto se debe vender (estaría sobre-vendido).

Para el cálculo del RSI se definen primero dos funciones U y D en función de los precios de cierre (PC), donde U representa los cambios al alza y D los cambios a la baja. Estas funciones se representan en la ecuación 5.

$$\begin{cases} U_i = \max(PC_i - PC_{i-1}, 0) \\ D_i = \max(PC_{i-1} - PC_i, 0) \end{cases} \quad (5)$$

Tras el cálculo de estas funciones, se propone una fórmula para el cálculo de la fuerza relativa (RS) que varía según la fuente consultada. La estructura general es la mostrada en la ecuación 6, donde N es un periodo de referencia, normalmente 14 días. Las funciones f_N más utilizadas son la media móvil y la media móvil exponencial. La media móvil es la utilizada por algunos programas de cálculo como Matlab y es, por tanto, la que se utilizará en este caso

$$RS = \frac{f_N(U)}{f_N(D)} \quad (6)$$

Finalmente, conocido el RS, el cálculo del RSI es bastante directo y está consensuado en todas las fuentes consultada, a través de la ecuación 7.

$$RSI = 100 \left(1 - \frac{1}{1 + RS} \right) \quad (7)$$

De esta forma, todas las series, independientemente de sus valores, se transforman en series que oscilan entre valores fijos. Para ilustrar el funcionamiento de este indicador, se muestra en las figura 56-60 todas las funciones, iniciales, intermedias y finales, para el cálculo del RSI para la compañía ACS en el año 2018.

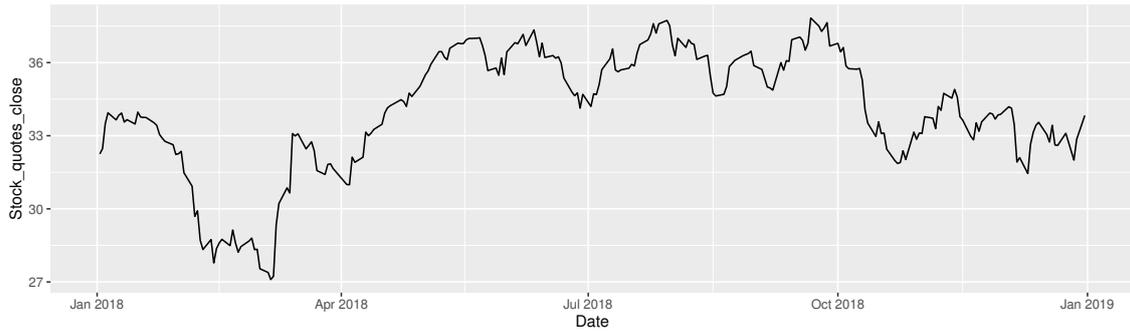


Figura 56: Valores de cierre de ACS

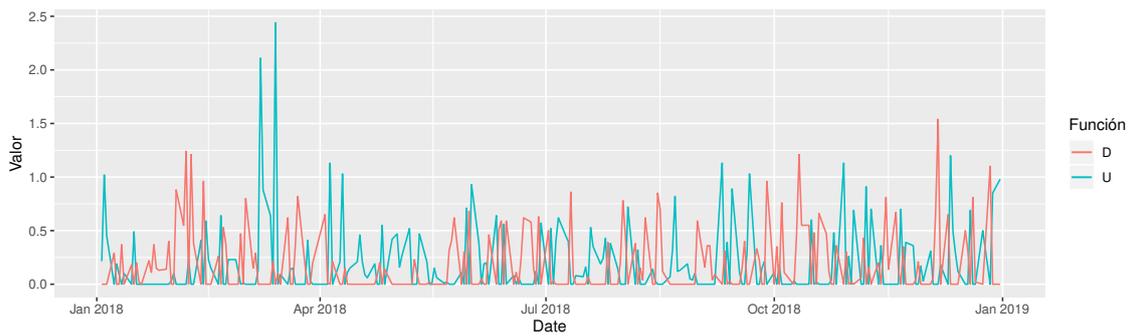


Figura 57: Cambios al alza y baja de ACS



Figura 58: Medias móviles de los cambios al alza y baja de ACS

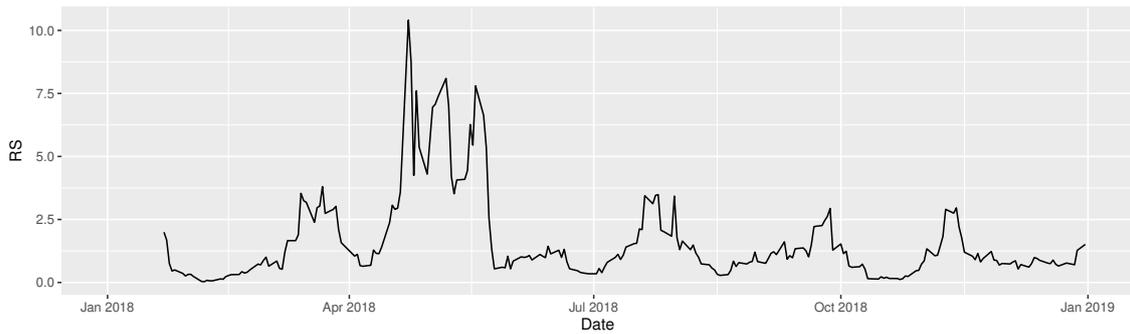


Figura 59: Fuerza relativa de ACS

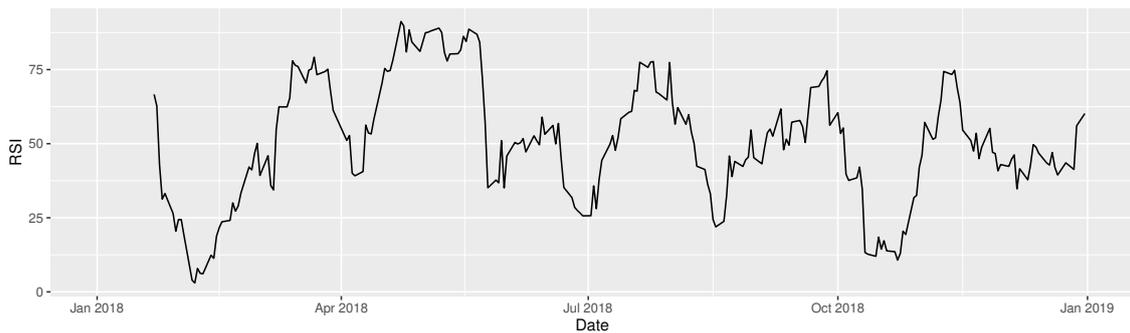


Figura 60: Índice de fuerza relativa de ACS

El RSI se suele utilizar de la siguiente forma. Se definen dos umbrales a partir de los cuales se lanza una señal de compra y de venta. Normalmente estos umbrales se suelen considerar de 70 (umbral superior) y 30 (umbral inferior), pero pueden tomarse los que uno considere. La señal de compra se lanza cuando el RSI baja del umbral inferior mientras que la señal de venta se lanza cuando el RSI supera el umbral superior. Siguiendo el ejemplo de ACS en 2018, se muestra en la figura 61 se muestran los límites con la evolución del RSI. En la figura 62 se incluyen además las señales de compra (verde) y venta (rojo), siendo ocurriendo la compra o venta cuando la línea se encuentra en el valor 100.

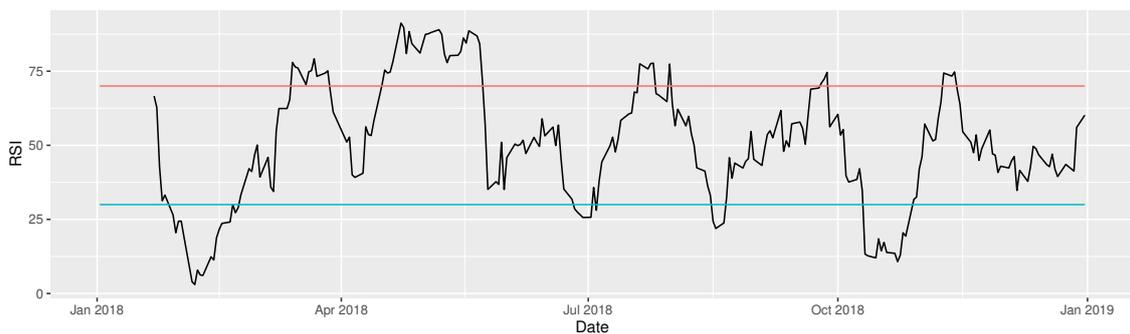


Figura 61: RSI de ACS con sus límites

Resulta interesante analizar estas señales de compra y venta con respecto a la evolución bruta (de los valores de cierre, no del RSI) del valor. Se puede apreciar en la figura 63 que este tipo de estrategia da resultados positivos en este caso.

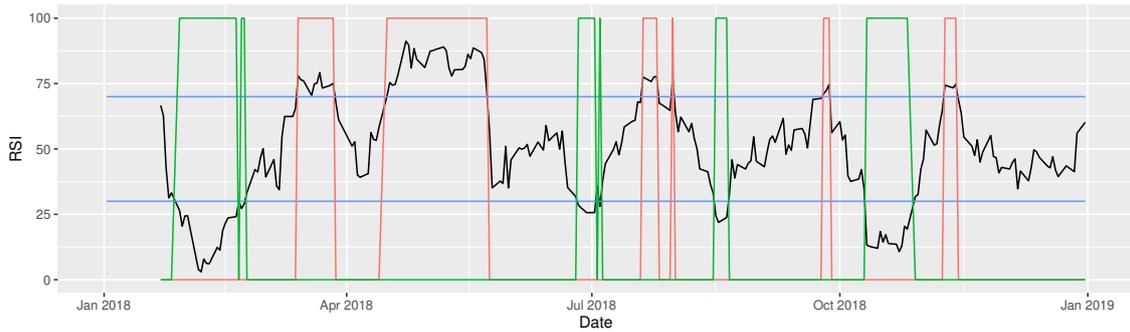


Figura 62: Señales de compra y venta para el RSI

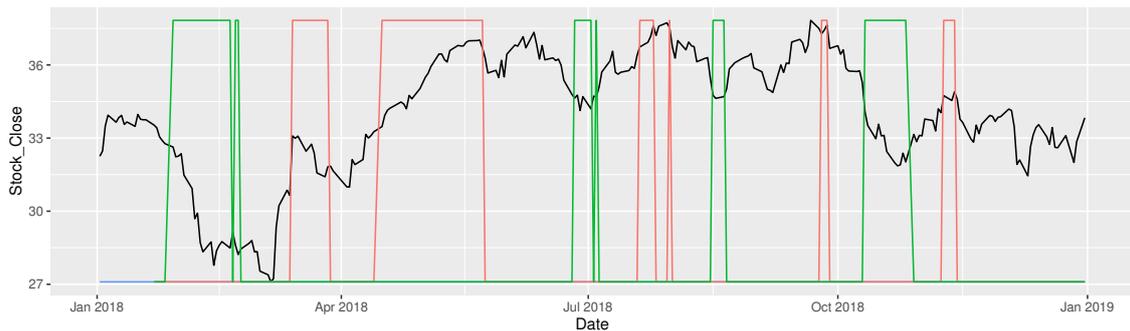


Figura 63: Señales de compra y venta con la evolución del precio de cierre

5.2. Aplicación de la metodología a la estrategia RSI

El siguiente paso que se desea tomar es el de aplicar la metodología propuesta a este tipo de series. El objetivo es analizar si la información relativa al año anterior, debidamente procesada, ayuda a obtener mayores rendimientos en el año siguiente. Para ello, el set de datos se divide por años y se calculan los parámetros para cada nueva serie. Por otro lado, para poder realizar la evaluación del rendimiento es necesario seguir una estrategia. Esta estrategia se evaluará en el año analizado (Y). La estrategia quedará definida por los umbrales de compra y de venta y se escogerá la pareja de umbrales que mayor rendimiento dé en el año Y.

En la definición del rendimiento se tomarán varias suposiciones. En primer lugar, se supondrá que cuando hay una señal de compra o de venta, se realizará la operación con todo el capital disponible (sistema todo o nada). Para simular esto se supondrá un capital inicial de valor la unidad, y ninguna acción en la cartera. Se ignorarán todas las señales de venta iniciales hasta que no haya una ocurrido una compra. Cuando se realice la compra, se convertirá todo el capital en acciones. Se desprecia el efecto que tiene no poder comprar fracciones de acciones. Una vez realizada una compra, se ignorarán todas las señales de compra posteriores mientras no se produzca una venta, ya que no quedaría capital para invertir. En las ventas se vende la totalidad de las acciones que se tienen en cartera. En este momento, las acciones se convierten en capital. El valor del nuevo capital será el ratio entre el precio de la acción el día de la venta y el día de la compra.

En segundo lugar, se supondrá que la totalidad de las acciones que se tengan en cartera se venden al acabar el año analizado. Esto, aunque no es el comportamiento fidedigno que seguiría un fondo, es necesario para evaluar correctamente el rendimiento de la estrategia en un año. Esto no debería afectar negativamente ya que, si se ha realizado alguna compra es bajo la premisa de que la acción estaba infra-vendida y deberá reportar al menos un cierto beneficio, aunque éste no sea el máximo posible.

En tercer lugar, simulando el comportamiento real, se añade una componente de comisión a la hora de realizar cada transacción. Esto se debe a que las entidades bancarias, a través de las cuales se realizan las operaciones, cobran una comisión por operación. En los casos estudiados se supondrá una comisión del 0.5 % por operación, lo que implica una comisión de aproximadamente 1 % por pareja de compra-venta.

En último lugar, para encontrar los rendimientos máximos, se probarán un total de 12 valores en cada umbral (superior e inferior). En el caso del umbral de compra, se probarán valores entre 2.5 y 30, mientras que en los umbrales de venta, se probarán valores de entre 70 y 97.5. Ha parecido razonable no usar saltos inferiores a 2.5.

Siguiendo con el ejemplo de ACS entre 2018 y 2019, se ha decidido ilustrar la función de rendimiento según los umbrales elegidos. Dicha función se muestra en la figura 64, donde se puede apreciar que el óptimo de rendimiento se encuentra en umbrales entre 75-77.5 para la venta y 5-7.5 en la compra. También se puede apreciar que, cuanto más extremos son los valores, menor es el riesgo ya que éste está asociado a la variabilidad de la rentabilidad.

Rendimiento según umbrales para ACS 2018–2019

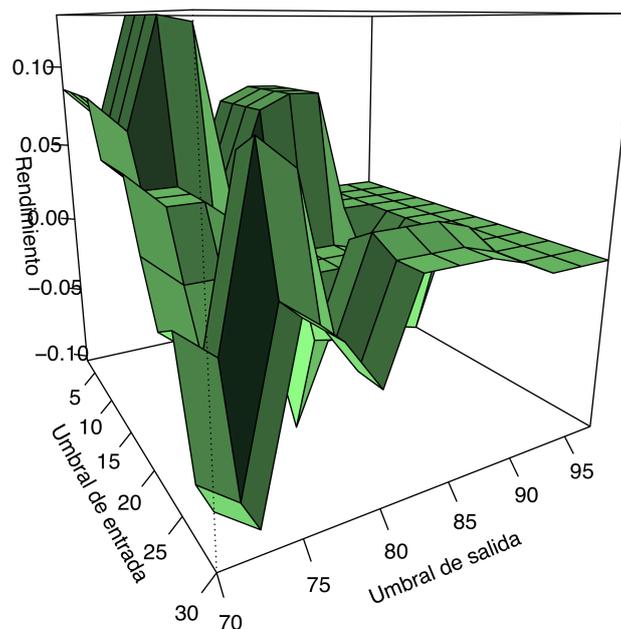


Figura 64: Rendimiento en función de los umbrales

De este análisis obtenemos, para cada pareja de empresa-año tres parámetros: rendimiento máximo que se puede obtener siguiendo la estrategia RSI y los umbrales óptimos en los que se obtiene dicho rendimiento.

El objetivo es utilizar estos datos de cada serie (pareja empresa-año) junto con todas las características de la serie temporal (especificados en las secciones anteriores) para predecir el rendimiento que puede tener la empresa en el próximo año. Para esto, se supondrá que se usan los mismos umbrales que en el año anterior.

Con este objetivo, se tratará de predecir el rendimiento que dará una empresa el año siguiente. Para ello, se compararán tres modelos de predicción: modelo *dummy* o *sin*, modelo lineal y modelo de red neuronal. En el modelo *dummy* se supondrá que el rendimiento del año $Y+1$ será el mismo que en el año Y (el analizado) si se usan los mismos umbrales. En el modelo lineal se usarán todos los parámetros de la serie para el año Y , el rendimiento óptimo en ese año y los umbrales óptimos para realizar una regresión lineal de la rentabilidad del año siguiente. En el caso del modelo de red neuronal, se usará una red neuronal con los mismos parámetros.

Para este tipo de aprendizaje se ha de definir un conjunto de datos de entrenamiento y de test. En este caso, el conjunto de test será el año tratado, mientras que el de entrenamiento serán los años anteriores. Se harán varias pruebas combinando el año predicho y el número de años utilizados como conjunto de entrenamiento. Además, se evaluará la bondad de la predicción de la siguiente manera: se propondrá un rendimiento mínimo objetivo y se invertirá en todas las empresas cuyo rendimiento sea superior al mínimo objetivo. Como se desea evitar el efecto de las empresas en un auge extremo, se quitarán aquellas que tengan un rendimiento superior al 250% en un año. Esto se debe a que estas empresas sesgan las regresiones y las redes neuronales.

Se va a realizar paso a paso un ejemplo de este análisis. Se va a tomar como ejemplo el año cinco, con los cuatro primeros años como conjunto de entrenamiento para una red neuronal. Para cada uno de estos cinco años se tiene la rentabilidad del año siguiente al seguirse la misma estrategia que el año pasado. Con estos datos (además de las características de la serie temporal), se entrena la red. En la figura 65 se muestran los rendimientos de los conjuntos de entrenamiento y test en el año analizado y en el siguiente. Se ha de indicar que, si tuviéramos esta información antes de invertir, la decisión sería fácil, ya que sólo habría que invertir en los valores que vayan a dar una mayor rentabilidad.

Para verificar la bondad del modelo, se analiza cómo de bien predice las rentabilidades para el año siguiente. Para ello se muestran sobre el eje x la rentabilidad predicha mientras que en el eje y se muestra la rentabilidad real que se habría obtenido. Si los resultados fueran exactos, los puntos estarían sobre la recta negra.

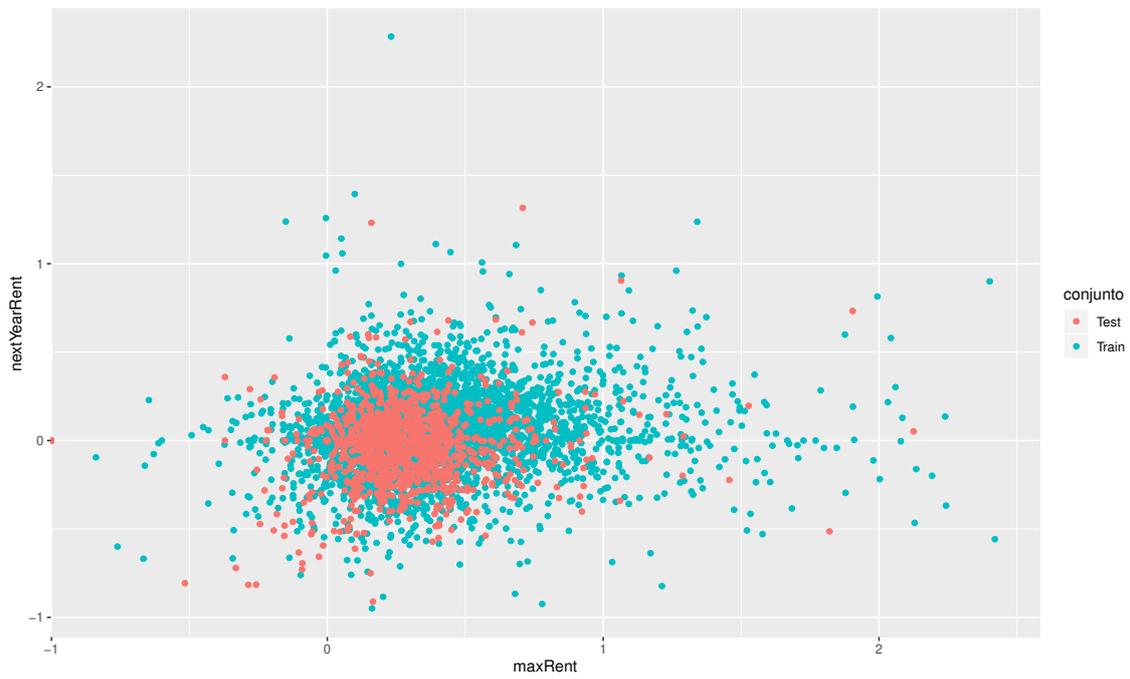


Figura 65: Relación entre las rentabilidades reales de un año y el siguiente

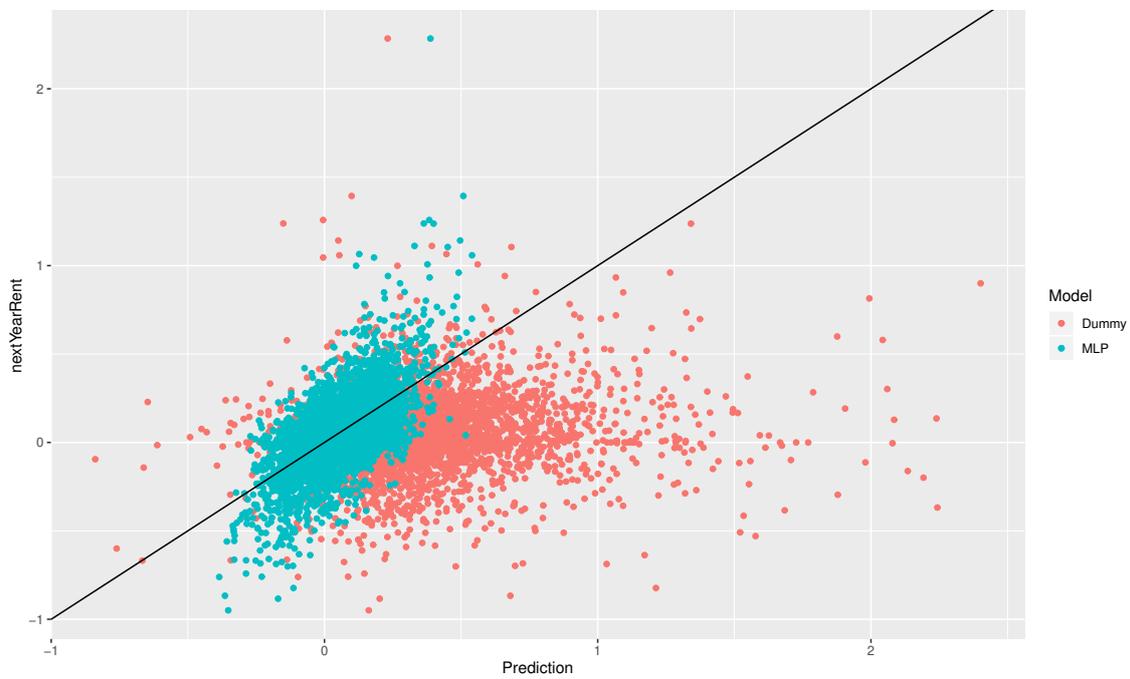


Figura 66: Comparativa dummy vs red neuronal - Conjunto de entrenamiento

El siguiente paso es decidir sobre qué empresas se va a invertir. Se va a realizar una prueba con el conjunto de entrenamiento. Con este objetivo se ha de establecer un filtro. En este caso se propone como filtro aquellos que tengan una rentabilidad predicha mayor al 10%. Esto quiere decir que no nos queremos arriesgar a invertir en aquellas empresas que no prevemos que vayan a dar una rentabilidad mayor al 10%. En la figura 67 se muestran las empresas seleccionadas. Para entender mejor dónde se realizan las inversiones y la rentabilidad de los países se muestra también la figura 68, donde aparecen unos diagramas de cajas y bigotes de las rentabilidades obtenidas por país.

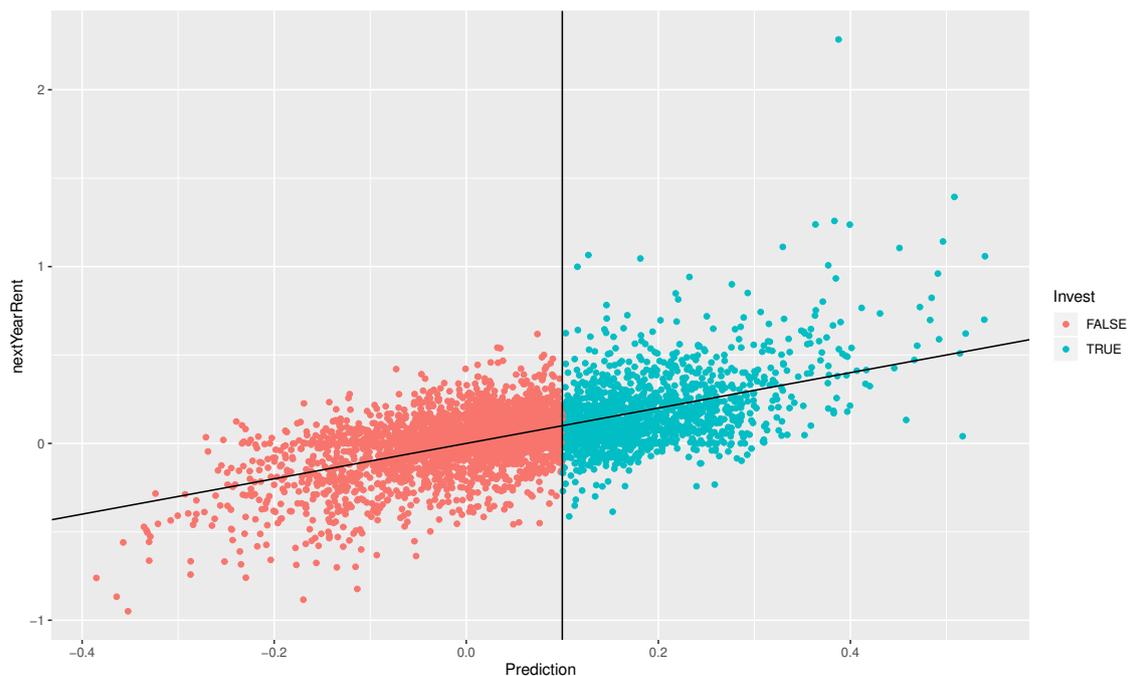


Figura 67: Predicción vs realidad de las empresas seleccionadas - Red Neuronal

En esta última figura puede parecer que la estrategia es bastante buena dado que tiene una rentabilidad bastante alta. Sin embargo, se ha de comparar con la estrategia con un modelo tomando los datos del año anterior como referencia para verificar si realmente es el modelo o que ese año fue un buen año para las inversiones bursátiles.

Se ha de verificar, también, si existe un efecto que se repite para los distintos umbrales con los que se trabaje (de -25% a +25% con incrementos del 1%). Con estos objetivos, se muestra la figura 69 donde se muestra el rendimiento para el umbral seleccionado entre los dos modelos: con una red neuronal y con la previsión del año anterior. En este año y con estos datos se puede observar que la red neuronal (MLP) da mejores resultados que la estrategia “dummy”. Sin embargo, resulta necesario verificar si esto se repite para el conjunto de test, para otros años y/o cogiendo un número distinto años para entrenar la red. También se debe verificar si las pérdidas se deben al uso de las técnicas incorrectas y se deben probar otras, tales como la regresión lineal.

Sin embargo, antes de realizar este tipo de pruebas, resulta interesante realizar un análisis de la sensibilidad de la red neuronal para ver así qué parámetros son los que

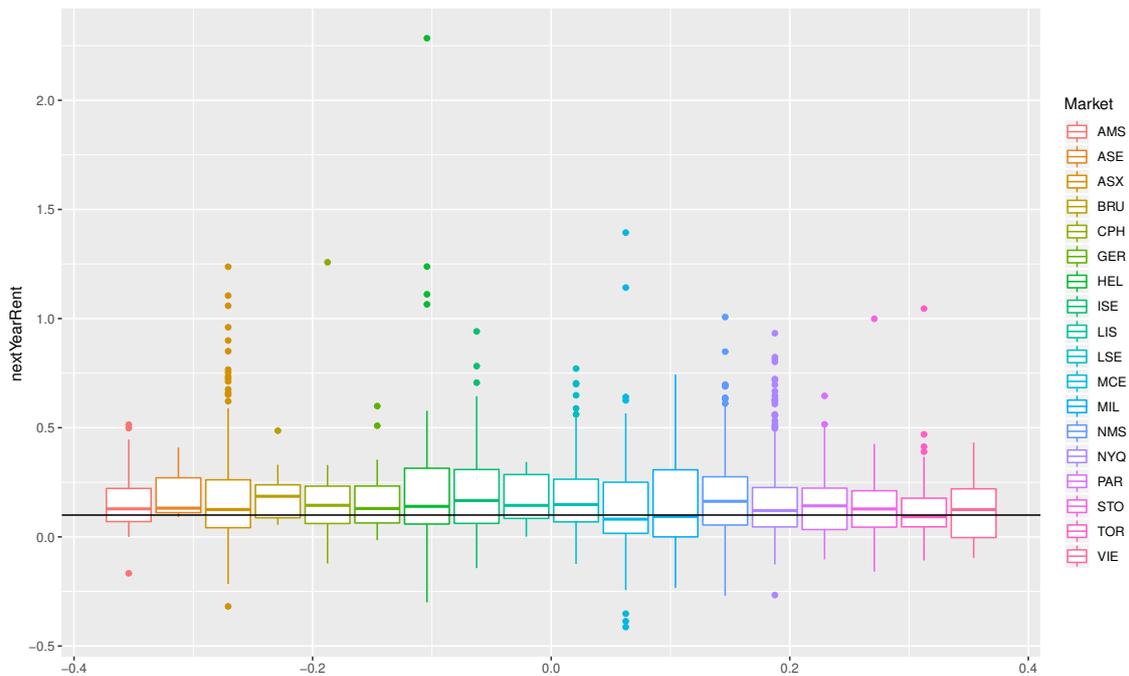


Figura 68: Rendimiento de la cartera por países

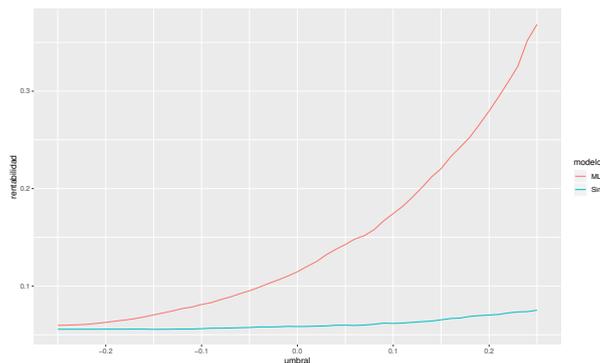


Figura 69: Comparativa de rentabilidad según los umbrales escogidos

más afectan. Por ello, se muestra la figura 70. En ella se puede observar que las variables más influyentes en esta red son el estadístico de Phillips Perron (relacionado con las raíces unitarias y las tendencias constantes), la fuerza de la tendencia, la función de autocorrelación parcial para la serie diferenciada y la entropía espectral. Resulta interesante resaltar que los valores que menos importancia tienen son los p-valores del estadísticos de Phillips Perron, de Ljung-Box y de Box-Pierce. Resulta curioso ver que el rendimiento del año anterior es la variable número 22 en influencia, mientras que los umbrales de entrada y de salida ocupan la posición sexta y duodécima.

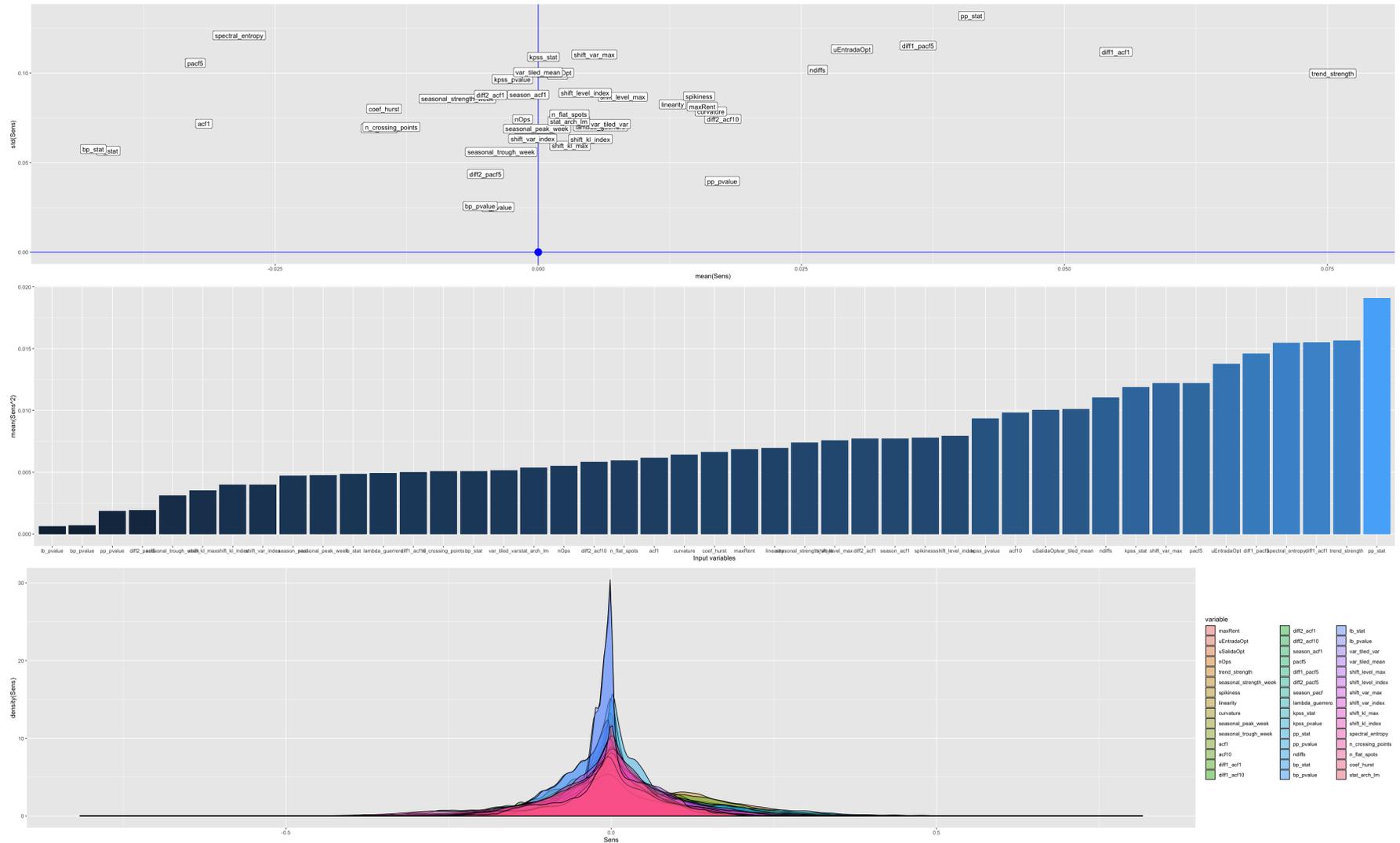


Figura 70: Análisis de sensibilidad de la red neuronal

Una vez verificado que el sistema funciona correctamente para el conjunto de entrenamiento, se debe verificar que este método sirve para invertir en el futuro. Para ello los únicos datos que tenemos disponibles son los del conjunto de test dado que los de entrenamiento se refieren al pasado y ya no podemos invertir. Se espera que los resultados sean notablemente peores, dado que hasta ahora se estaba trabajando conociendo lo que iba a pasar.

Con el objetivo de verificarlo se van a utilizar los 4 años anteriores y evaluando el rendimiento en el siguiente año. Al tener 10 años de datos, en este caso habrá 5 años de pruebas. Para ello se trata de predecir los valores de los años 5, 6, 7, 8 y 9 con los datos de los años 1-4, 2-5, 3-6, 4-7 y 5-8 respectivamente. El año 10 no se puede analizar porque no se tienen datos de las rentabilidades futuras. En las figuras 71 y 72 se muestran las predicciones realizadas por cada modelo respecto al valor real para los conjuntos de entrenamiento y de test. Se ha incluido el modelo lineal.

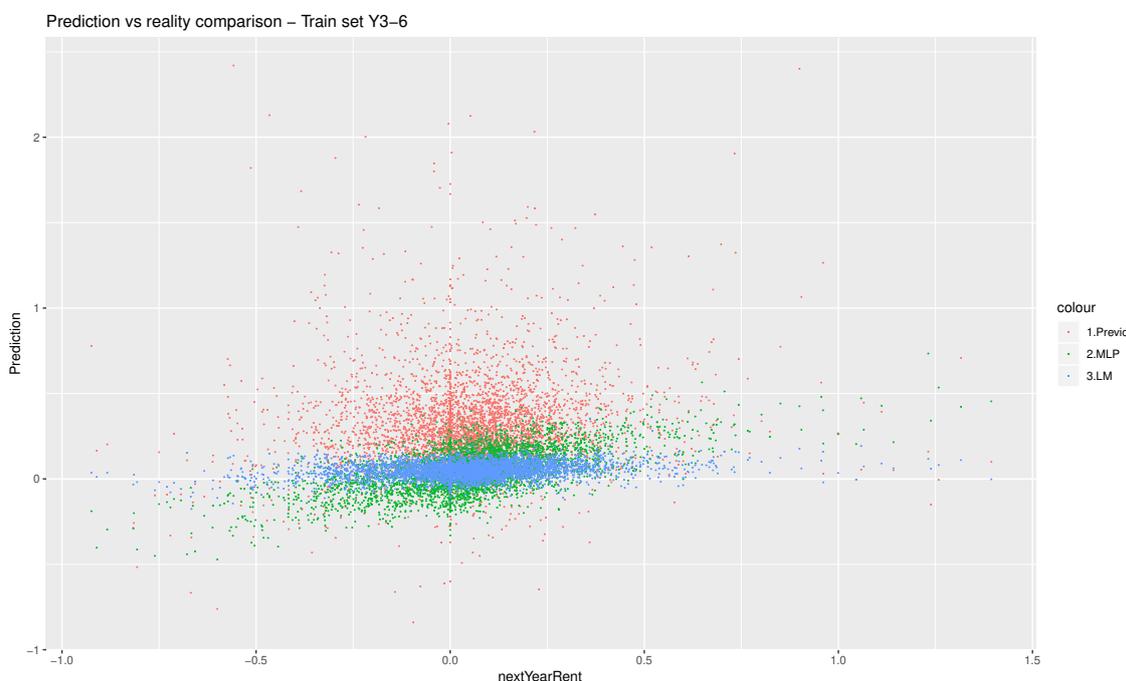


Figura 71: Predicción para el año 7 tomando los anteriores 4 años - Entrenamiento

Por otro lado, los resultados de la prueba se muestran en la figura 73. En esta figura se puede ver que la estrategia basada en la predicción de redes neuronales tiene una eficacia mayor a la del sector de una forma consistente respecto a la estrategia de usar los valores de un año anterior (estrategia *dummy*). La estrategia basada en la regresión lineal se puede considerar arriesgada dado que puede dar grandes beneficios o grandes pérdidas respecto al caso base. Se puede afirmar que la predicción con 4 años de experiencia y con filtros de rentabilidad esperada entre el 10% y el 20% resultan, de manera consistente, en mayores beneficios que las otras estrategias. Los resultados son claramente inferiores a los obtenidos por el conjunto de entrenamiento, como se esperaba.

Estos resultados son consistentes tomando como conjunto de entrenamiento los datos

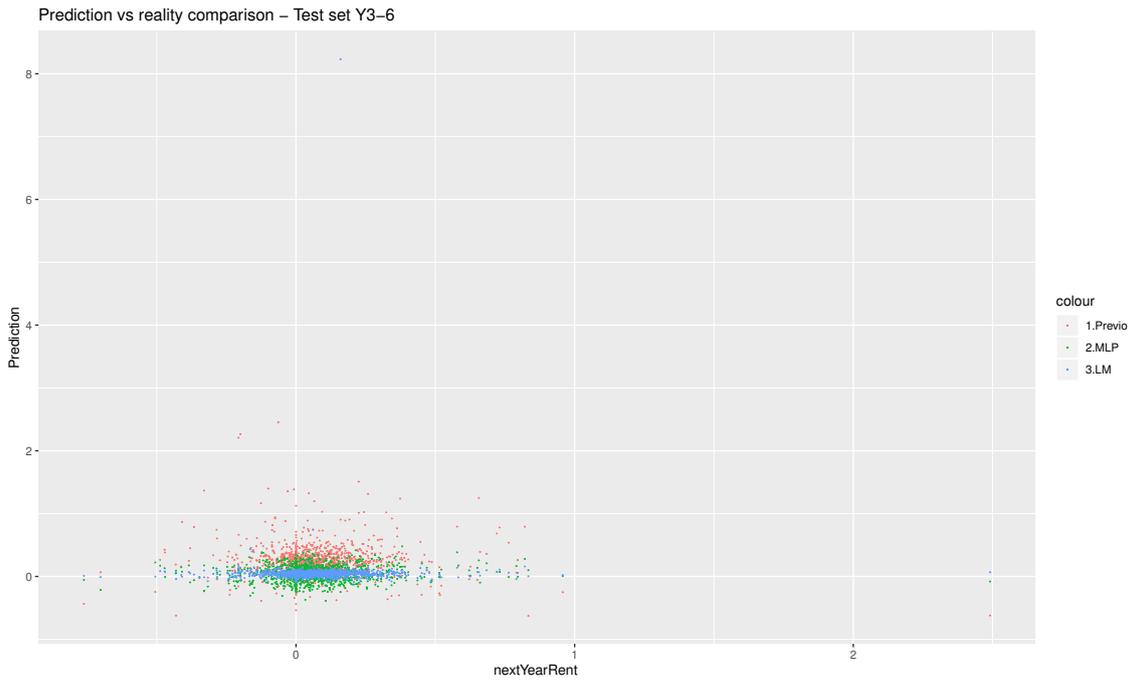


Figura 72: Predicción para el año 7 tomando los anteriores 4 años - Test

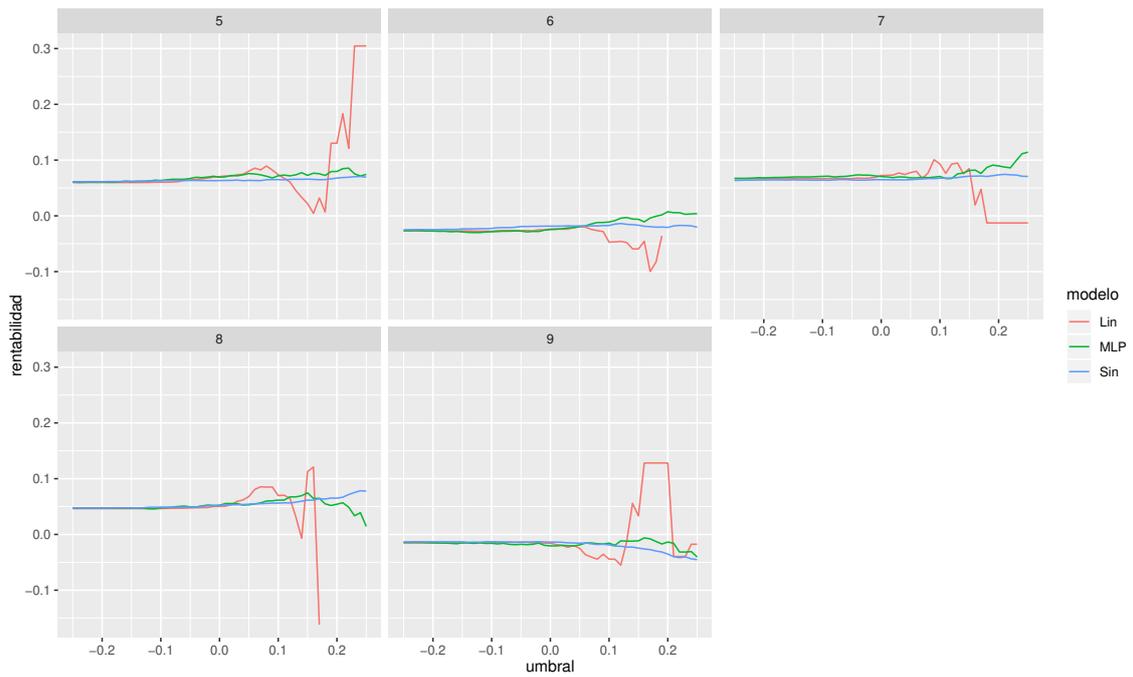


Figura 73: Rendimientos de los distintos modelos

de 3 ó 5 años anteriores. Esto indica que la metodología sirve para ganarle al mercado. Se ha de añadir que ganar aunque sea sólo un 1% respecto al mercado es todo un éxito ya que puede suponer un gran fuente de ingresos y una reducción importante de riesgos. Se ha de enfatizar la diferencia de esta metodología respecto a la que se ha utilizado en las estrategias presentadas en el estado del arte. En esas metodologías se trataba de

predecir el valor de las acciones al corto o medio plazo, mientras que en este caso se trata de predecir el rendimiento de una estrategia prefijada y decidir sobre qué valores invertir basado en esa predicción.

Un análisis que surge tras decidir usar una red neuronal es el de encontrar la arquitectura óptima para este tipo de predicciones. Hasta ahora se había estado trabajando con una red de 1 capa intermedia de 15 neuronas. Ahora se propone realizar pruebas con distinto número de neuronas. Para ello se realizan las mismas pruebas que con el modelo de 15 neuronas pero variando el número de neuronas entre 5 y 20. No se ha cogido un número de neuronas superior a 20 dado que el número de pesos es excesivo con respecto a los datos. Los resultados de este análisis se muestran en la figura 74.

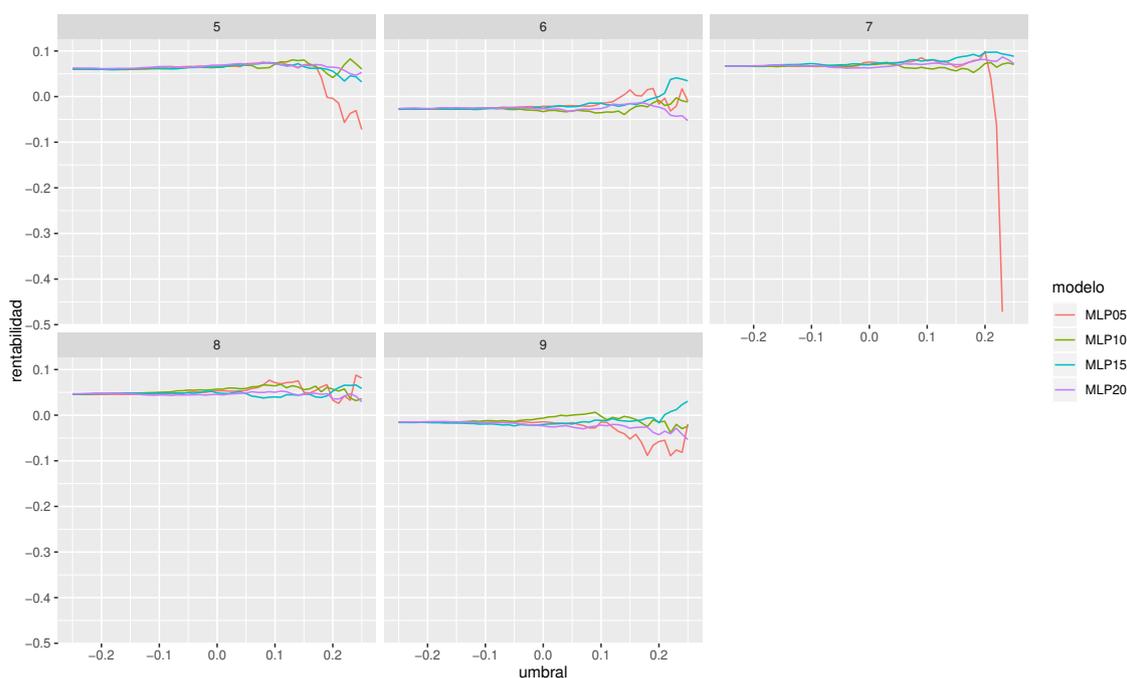


Figura 74: Comparativa entre diferentes tamaños de redes neuronales

En esta figura se puede observar que usar el modelo de 5 neuronas resulta contraproducente en la mayoría de los casos. Para umbrales inferiores al 0% los modelos funcionan prácticamente igual. Sin embargo, entre los umbrales del 10% y el 20% el que mayores rentabilidades da, y de manera consistente, es la red de 15 neuronas. Es por ello se puede decir que el modelo usado en el desarrollo es el óptimo entre los analizados.

Viendo que el modelo de 15 neuronas es el óptimo, se va a proponer una estrategia ligeramente distinta. Dado que la rentabilidad y el riesgo suelen ir de la mano (a mayor rentabilidad, mayor riesgo), parece razonable evitar invertir en las compañías que dan elevados niveles de rentabilidad si se busca un nivel moderado. Por tanto, la estrategia que se propone es la de invertir en valores que den una rentabilidad aproximadamente igual a la predicha, con un umbral de, por ejemplo, $\pm 2,5\%$.

Siguiendo con el ejemplo explicativo, se muestra en la figura 75 el equivalente a la

figura 67 para esta estrategia. Se elige como objetivo un rendimiento del 10%, y una tolerancia del 2.5%. Por tanto, se tomarán los valores cuyo rendimiento esperado esté entre 7.5% y 12.5%. También se muestra la distribución de los rendimientos obtenidos por país a través de una gráfico de cajas y bigotes en la figura 76.

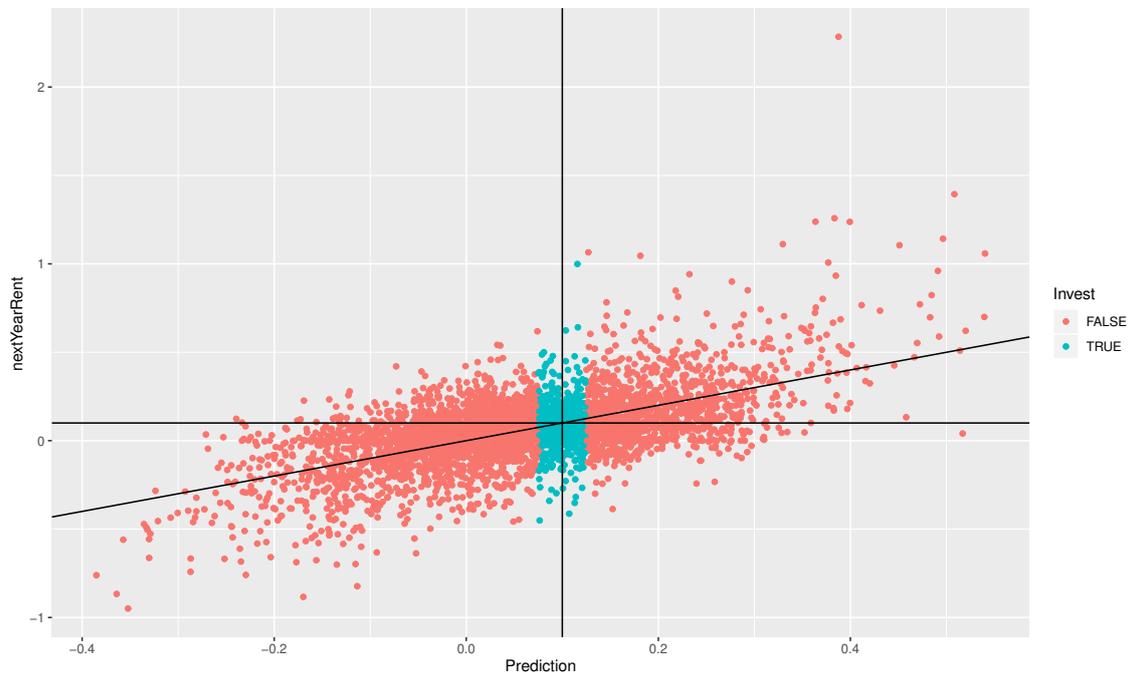


Figura 75: Empresas seleccionadas según la segunda estrategia- Red Neuronal

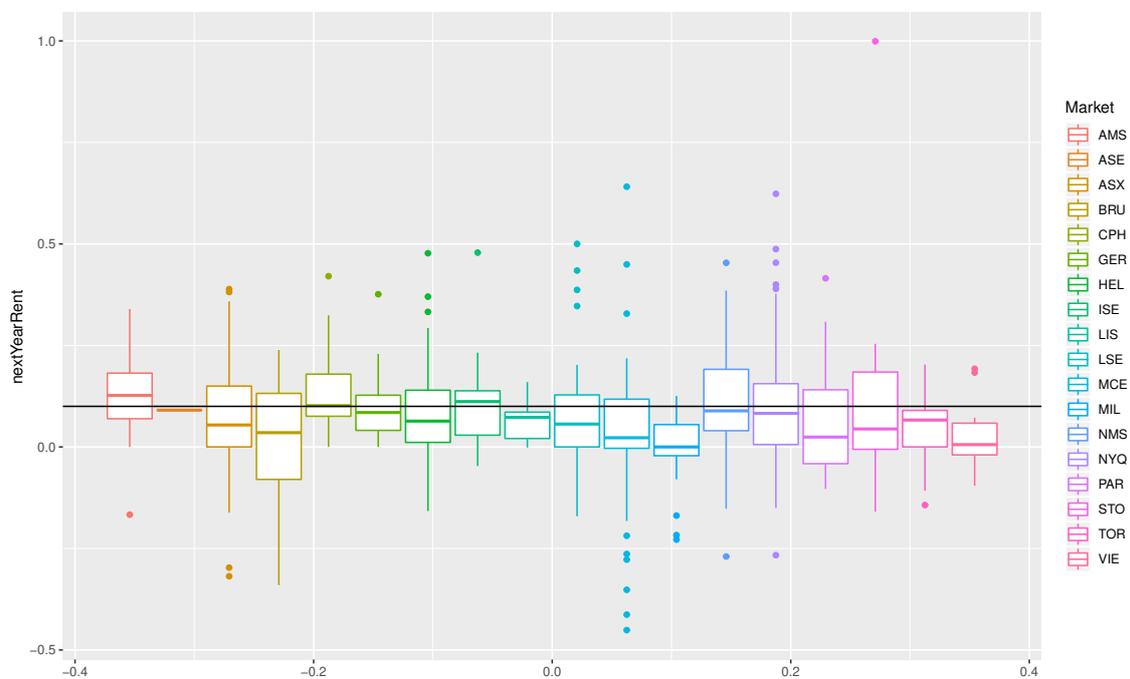


Figura 76: Empresas seleccionadas según la segunda estrategia- Red Neuronal

En estas figuras se puede apreciar que los resultados son claramente inferiores a los de

la otra estrategia en el conjunto de entrenamiento. Sin embargo, el objetivo perseguido es el de probar esta estrategia en el conjunto de test. Con este objetivo se ha construido la figura 77, donde se muestran los resultados de la estrategia propuesta en los mismos casos que se mostraron para la estrategia inicial. Además, al haber cambiado la estrategia, se evalúan de nuevo las predicciones con el valor del año anterior y con el modelo lineal. Las curvas pertenecientes a esta nueva estrategia son las que en el nombre acaban en 2.

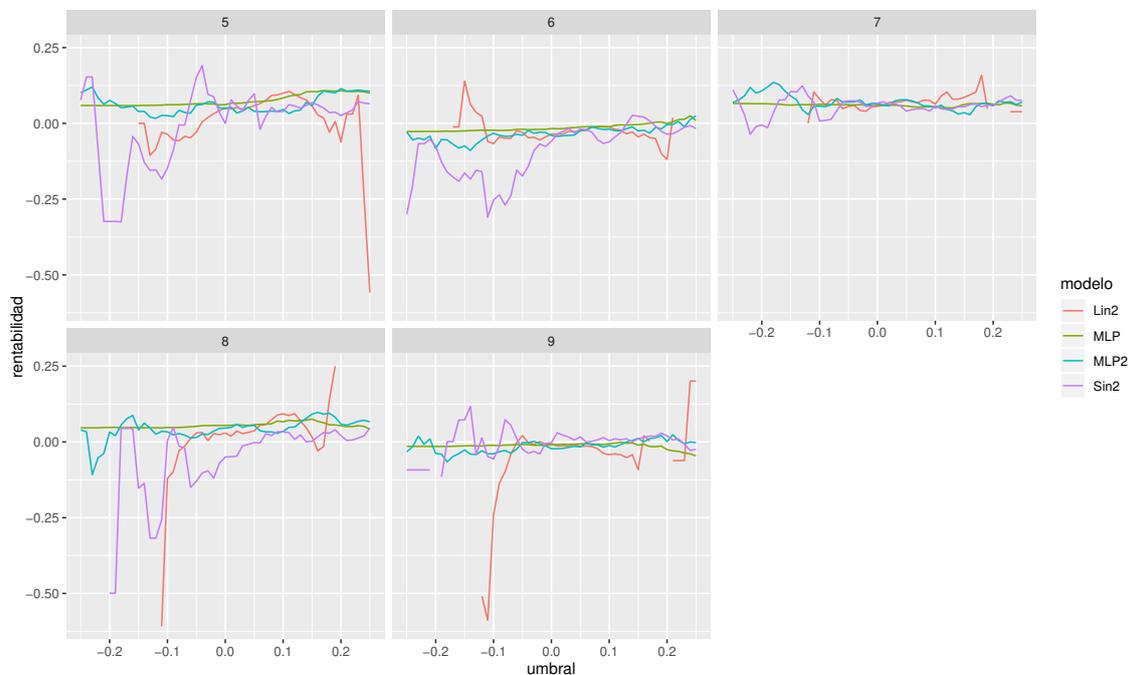


Figura 77: Resultados de la nueva estrategia

En esta figura se observa que ninguna de estas estrategias dan mayores beneficios que la estrategia original de manera consistente. En cambio, existe un patrón que sí puede ser de utilidad y es la siguiente: Esta estrategia aplicada sobre la predicción de la red neuronal funciona igual o mejor que la anterior para umbrales superiores al 20%. Con respecto al modelo lineal, a veces tiene un comportamiento mejor que el la estrategia original, pero no resulta consistente a lo largo del tiempo. Respecto a este modelo lineal resulta interesante ver que hay zonas en los que la serie simplemente desaparece. Esto se debe a que no existen compañías cuyo rendimiento predicho por el modelo lineal sea de los valores especificados.

Se considera, por tanto, que el mejor modelo posible de entre los calculados es un modelo que invierta en las compañías cuyo rentabilidad predicha sea mayor al umbral deseado salvo cuando este umbral sea superior al 20%, que entonces se deberá invertir en las compañías cuya predicción se encuentre entre $\pm 2,5\%$ del umbral seleccionado.

5.3. Conclusiones

En esta sección se ha abordado el aprendizaje supervisado aplicado a una estrategia de inversión. Se han propuesto tres modelos de predicción, y sobre estos modelos dos formas de elegir las compañías en las que invertir. De todas las posibilidades se comprueba que el sistema que mejor funciona es el de la predicción con un modelo de red neuronal de 15 neuronas, y tomando como conjunto de entrenamiento los 4 años anteriores. Además, la estrategia que mejor funciona es la de invertir en las empresas que tengan un rendimiento esperado mayor al umbral seleccionado, salvo para los rendimientos superiores al 20%. En estos últimos casos es conveniente utilizar la estrategia de invertir en aquellos valores cuyo rendimiento se encuentre en el entorno del $\pm 2,5\%$ alrededor del umbral seleccionado.

Se ha observado que este sistema da mejores rendimientos que la estrategia básica. Con esto se verifica que el sistema ayuda a dar mayores rentabilidades a la hora de invertir en los mercados financieros.

6. Conclusiones y Futuros Desarrollos

El objetivo del trabajo era el de conseguir rendimientos económicos de los mercados bursátiles aplicando técnicas de aprendizaje automático. Se ha aplicado una metodología de análisis genérico de series temporales para ayudar en la elección de valores en los que invertir. Las principales conclusiones obtenidas son:

- Esta metodología puede ser utilizada para distinguir unas series de otras, acelerando el proceso de elección de compañías en las que invertir.
- No todas las características resultan igual de útiles a la hora de diferenciar o predecir estas series. Algunas características presentan aportaciones marginales a la diferenciación o predicción.
- Existen relaciones entre las características y algunos mercados. Esto implica que, a través de esta metodología, no solo se pueden distinguir compañías, sino mercados enteros, pudiendo evaluar características globales los mismos, tales como el riesgo.
- Existen relaciones entre las características de las series temporales y los rendimientos que han dado durante el periodo analizado
- Esta metodología permite obtener eficazmente mejores resultados económicos respecto a una estrategia inicial. Esta metodología ha sido probada a largo plazo, pues se usaba la información para predecir los rendimientos a lo largo de un año completo.

Además, a lo largo del trabajo se ha considerado que puede ser interesante continuar el trabajo en las siguientes líneas, que se proponen como futuros desarrollos:

- Eliminación de ciertas características para la clusterización y predicción para observar si se aprecian diferencias notables en el análisis.
- Aumento de la frecuencia del análisis. Ahora mismo los datos de un año analizado no se solapa con otro año. Puede ser interesante analizar, manteniendo fijo el espacio de un año, ventanas que se solapen. Por ejemplo, Si se tuvieran solo los datos entre el 01/01/2015 y el 31/12/2016, en la metodología utilizada se tomarían dos espacios: 01/01/2015-31/12/2015 y 01/01/2016-31/12/2016. Sin embargo, con lo que se propone, se podría añadir el periodo de 01/07/2015-30/06/2016.
- Replicar la metodología de predicción e inversión con las características de la serie de valores de cierre (la original) en vez de la serie RSI como valores de entrada, conservando el valor de salida del rendimiento de dicha estrategia.
- Utilizar las redes neuronales para definir los umbrales que deben ser utilizados el año siguiente para obtener un combinación óptima entre rendimiento y riesgo.
- Evaluación de otras métricas, de tipo económicas y relacionadas con el riesgo, para ser utilizadas como entradas a las redes neuronales. Se sugieren: pérdidas y ganancias máximas a lo largo del periodo estudiado y la estabilidad de los rendimientos en otros años.

- Aprovechamiento de las métricas no utilizadas en este estudio para la predicción, tales como la valoración del riesgo de Bloomberg o el EBITDA de la compañía.
- Uso de estrategias para la inversión alternativas al RSI, tales como la media móvil convergente-divergente [Ada20] (MACD) y evaluación de cuál puede ser más adecuada para cada valor a través de aprendizaje supervisado de tipo clasificación.
- Uso de redes neuronales recurrentes para la gestión de carteras con capital limitado y dinámica. Esto quiere decir que la actualización de las compañías a invertir es actualizada cada día, semana o mes, según la información más reciente. El resultado podría ser una ordenación de las compañías según su atractivo.

Referencias

- [Ada20] Adam Hayes. Moving Average Convergence Divergence – MACD. *Investopedia*, 2020.
- [AEAM18] Khaled A. Althelaya, El-Sayed M. El-Alfy, and Salahadin Mohammed. Evaluation of Bidirectional LSTM for Short- and Long-Term Stock Market Prediction. *International Conference on Information and Communication Systems (ICICS)*, 9, 2018.
- [AYMU16] Ryo Akita, Akira Yoshihara, Takashi Matsubara, and Kuniaki Uehara. Deep learning for stock prediction using numerical and textual information. *ICIS Okayama, Japan*, June 2016.
- [BC64] G. E. P. Box and D. R. Cox. An analysis of transformations. *JRSS B*, 26(6):211–246, 1964.
- [CZD15] Kai Chen, Yi Zhou, and Fangyan Dai. A LSTM-based method for stock returns prediction : A case study of China stock market. *IEEE International Conference on Big Data*, 2015.
- [DPH16] Luca Di Persio and Oleksandr Honchar. Artificial Neural Networks architectures for stock price prediction: comparisons and applications. *International Journal of Circuits, Systems and Signal Processing*, 10, 2016.
- [Eco19] The stockmarket is now run by computers, algorithms and passive managers. *The Economist*, October 2019.
- [Gue93] V.M. Guerrero. Time-series analysis supported by power transformations. *Journal of Forecasting*. *Journal of Forecasting*, 12:37–48, 1993.
- [HA19] R.J. Hyndman and G Athanasopoulos. *Forecasting: principles and practice*. OTexts: Melbourne, Australia. OTexts.com/fpp3, 2019.
- [Hay20] Hayes, Adam and Scott, Gordon. Portfolio Management. *Investopedia*, 2020.
- [Hon19] Alexandr Honchar. AI for portfolio management: from Markowitz to Reinforcement Learning. *The Startup*, 2019.
- [Hyn15] Hyndman, Rob J. and Wang, Earo and Laptev, Nikolay. Large-Scale Unusual Time Series Detection. *ICDM*, 2015.
- [Jam20] James Chen. Relative Strength Index (RSI). *Investopedia*, 2020.
- [KHSM17] Yanfei Kang, Rob Hyndman, and Kate Smith-Miles. Visualising forecasting algorithm performance using time series instance spaces. *International Journal of Forecasting*, 33:345–358, 04 2017.

- [KS92] Phillips P.C.B. Schmidt P. Kwiatkowski, D. and Y. Shin. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Econometrics*, 54(6):159–178, 1992.
- [Mat19] Mathworks. *Aprendizaje automático en MATLAB*, 2019.
- [Mr.08] Mr. R. Tsay. Unit Root and Unit-Root Test. Lecture, The University of Chicago School of Business, 2008.
- [NPdO17] David M. Q. Nelson, Adriano C. M. Pereira, and Renato A. de Oliveira. Stock Market’s Price Movement Prediction With LSTM Neural Networks. *IEEE*, 2017.
- [OWHW+20] Mitchell O’Hara-Wild, Rob Hyndman, Earo Wang, Di Cook, Thiyanga Talagala, and Leanne Chhay. Package ‘feasts’. *CRAN*, Junio 2020.
- [qua19] Hedge Funds. *Investopedia*, September 2019.
- [Rey19] Ricardo Michel Reyes. El sueño de una máquina que gana dinero por ti mientras duermes. *El financiero*, June 2019.
- [SSR18] Eugenio Sánchez, Pedro Sánchez, and Andrés Ramos. Regression. *Quantitative Decision Methods*, September 2018.
- [Ste16] Stephanie. Unit Root: Simple Definition, Unit Root Tests. *statisticshowto.com*, 2016.
- [Sub12] Subir Mansukhani. The Hurst Exponent: Predictability of Time Series. *Analytics Magazine*, 2012.
- [THA18] Thiyanga S Talagala, Rob J Hyndman, and George Athanasopoulos. Meta-learning how to forecast time series. Monash Econometrics and Business Statistics Working Papers 6/18, Monash University, Department of Econometrics and Business Statistics, 2018.
- [Uni20] United Nations. About the Sustainable Development Goals. 2020.
- [vV19] Fjodor van Veen. *The Neural Network Zoo*, October 2019.

A. Alineación del Proyecto con los ODS

Los objetivos de desarrollo sostenible (ODS) de las Naciones Unidas son una serie de objetivos, 17 concretamente, que se fijaron en 2015 para abordar los retos globales. Estos retos incluyen aquellos relacionados con la pobreza, el cambio climático, la degradación del medio ambiente, la paz y la justicia. La idea es cumplir estos objetivos en el año 2030. En la figura 78 se muestra cada uno de estos 17 objetivos. Toda esta información está tomada de la página web de las Naciones Unidas [Uni20].



Producido en colaboración con TROLLBÄCK + COMPANY | TheGlobalGoals@trollback.com | +1.212.529.1010
Para cualquier duda sobre la utilización, por favor comuníquese con: dpicampaign@un.org

Figura 78: Objetivos de desarrollo sostenible

Este trabajo se puede considerar alineado con tres de estos objetivos. Se tratan del objetivo 8, trabajo decente y crecimiento económico; del objetivo 9, industria, innovación e infraestructura; y del objetivo 12, producción y consumo responsable.

La relación con el objetivo 8 es la siguiente: este trabajo desarrolla una metodología que mejora la eficiencia de las inversiones bursátiles, lo cual estimulará la economía. Esto se debe a que las propias empresas poseen y negocian con las acciones de la propia u otras empresas. El hecho de que estas empresas (o incluso particulares) obtengan mayores beneficios de los mercados bursátiles es lo que impulsará la economía.

En relación con el objetivo 9, este trabajo propone aplicar una técnica de análisis de datos de una forma de la que no se tiene constancia de que haya sido utilizada. Esto es, por definición, una innovación. Esto resulta interesante dado que puede inspirar a otras personas a desarrollar trabajos que potencien este mismo aspecto u otros alineados con los ODS.

Por último, existe una alineación con el objetivo 12. Se ha de recordar que las empresas invierten parte de su capital en otras empresas, por lo que pueden considerar estas acciones como activos propios de la empresa, al igual que puede serlo un ordenador o una mesa. Al mejorar este sistema la precisión a la hora de realizar las inversiones, disminuye el número de inversiones a realizar por dichas empresas. Se puede considerar, por tanto, un “consumo” responsable de este tipo de activos.