



MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MASTER

DESARROLLO DE UN SISTEMA DE INVERSIÓN EN
ACCIONES BASADO EN LA CREACIÓN DE CARTERAS.
ANÁLISIS TÉCNICO Y DE NOTICIAS PARA SU GESTIÓN
EFICIENTE.

Autor: Diego Gil de Antuñano Durán

Director: Carlos Maté Jiménez

Madrid

Julio de 2021

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
**DESARROLLO DE UN SISTEMA DE INVERSIÓN EN ACCIONES BASADO EN LA
CREACIÓN DE CARTERAS. ANÁLISIS TÉCNICO Y DE NOTICIAS PARA SU
GESTIÓN EFICIENTE.**

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2020/21 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.: Diego Gil de Antuñano Durán

Fecha: 19/07/2021

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Carlos Maté Jiménez

Fecha: 19/07/2021



MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MASTER

DESARROLLO DE UN SISTEMA DE INVERSIÓN EN
ACCIONES BASADO EN LA CREACIÓN DE CARTERAS.
ANÁLISIS TÉCNICO Y DE NOTICIAS PARA SU GESTIÓN
EFICIENTE.

Autor: Diego Gil de Antuñano Durán

Director: Carlos Maté Jiménez

Madrid

Julio de 2021

Agradecimientos

A mis padres por la oportunidad de estudiar en ésta prestigiosa universidad.

DESARROLLO DE UN SISTEMA DE INVERSIÓN EN ACCIONES BASADO EN LA CREACIÓN DE CARTERAS. ANÁLISIS TÉCNICO Y DE NOTICIAS PARA SU GESTIÓN EFICIENTE.

Autor: Gil de Antuñano Durán, Diego.

Director: Maté Jiménez, Carlos.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Este proyecto desarrollará un sistema original, básico y visual para agrupar los valores de varios ITS permitiendo la selección del mejor activo en cada cluster. En consecuencia, se podrían generar alternativas de carteras para un conjunto de acciones. Clasificar estas carteras, utilizar información de noticias u otra fuente de datos no estructurada mediante web scraping, y analizar el comportamiento de las más eficientes es uno de los resultados esperados de este proyecto. Se incluirá la aplicación a los activos del mercado de valores español, incluidos los del IBEX35.

Palabras clave: mercados financieros, clustering, noticias, clasificación, índice fuerza relativa, medias móviles.

1. Introducción

Una de las maravillas de los mercados financieros es que son capaces de reflejar la situación en la que vivimos. Son un indicador numérico, medible y disponible a todo el público en el que se ven reflejadas todas las sucesos que han ocurrido hasta el momento. Por lo tanto, un factor clave para obtener rentabilidades atractivas en mercados financieros es la información que se dispone y el momento en la que se tiene. Por este motivo, y con la cantidad de información disponible hoy en día es imprescindible disponer de un sistema automatizado para la medición del efecto de noticias en los mercados.

2. Definición del proyecto

El objetivo es la transformación de información desestructurada presente en titulares de noticias en variables cuantitativas con las que explicar la evolución de la cotización de empresas en bolsa. Una vez analizada esta información, se plantea la creación de un sistema original de trading basado en la optimización de carteras mediante agrupación de empresas del índice IBEX35, clasificación de noticias en función de su repercusión en determinadas acciones, análisis técnico de cotización de empresas, predicción de dirección de series temporales de intervalo STI's y elaboración de recomendación de carteras.

Para la definición del proyecto es necesario hacer un repaso de los componentes claves necesarios para la consecución de los objetivos planteados. Estas componentes son:

- (i) Análisis técnico básico de mercados financieros. El trabajo se apoya en el índice de fuerza relativa (RSI) y el análisis de tendencias y cruce de medias móviles para la creación de estrategias de trading y selección de carteras.

- (ii) Teoría de creación de carteras. La teoría de Markowitz para gestión de carteras propone la minimización del riesgo y la maximización de la rentabilidad. (Markowitz, 1991) El elevado número de acciones disponible en los mercados financieros hace difícil, ineficiente y, en ocasiones imposible, el uso del modelo de Markowitz. (He-Shan Guan, 2007) Por ello se propone la solución del problema mediante la agrupación de empresas y la posterior selección de la mejor acción de cada grupo para la creación carteras. Esta selección se realiza mediante estrategias basadas en índice de fuerza relativa, medias móviles y análisis de noticias.
- (iii) Estrategias de agrupación de datos. La agrupación de datos es una técnica de análisis estadístico no supervisado meramente descriptivo: a través de la búsqueda de patrones los datos se organizan en grupos con características similares. En este proyecto se proponen 3 métodos de agrupación: agrupación mediante k-medias de dos variables absolutas, agrupación mediante k-medias de dos variables usando valores medios y agrupación jerárquica.
- (iv) Extracción de noticias de internet. Elemento necesario para la obtención de las entradas del sistema. Se accede a información en tiempo real de datos bursátiles y titulares de noticias.
- (v) Sentimiento de mercado y análisis de sentimiento de noticias. Revisión de la Hipótesis de Mercados Eficientes de Eugene Fama y análisis de técnicas supervisadas y no supervisadas para el análisis de textos.
- (vi) Análisis de sistemas existentes de trading. Por último se hace un breve estudio de los sistemas existentes de trading y de las herramientas que estos ponen a disposición de los usuario. Se profundiza en el análisis de los sistema de trading y herramientas de Simply Wall Street y Webull.

3. Descripción del modelo/sistema/herramienta

El sistema es el resultado de una serie de iteraciones en las que se han ido implementando mejoras. El diagrama de la Ilustración 1 es la base que servirá como referencia para la creación de un programa usando la aplicación MATLAB 2020.

El sistema está diseñado para:

1. Extraer información bursátil de los *tickers* seleccionados en tiempo real de forma automática y crear una base de datos local.
2. Extracción de titulares de noticias relacionadas con las compañías objetivo de los principales periódicos económicos en España y creación de una base de datos local.
3. Creación de grupos de las compañías objetivo en base a los datos bursátiles históricos extraídos en el punto 1.
4. Análisis del sentimiento de las noticias de la base de datos local y su clasificación en función de la dirección esperada del retorno que provocan.
5. Creación de recomendaciones de carteras e inversión y uso de estrategias de trading.

Como se puede observar en la Ilustración 1 el sistema dispone de cuatro motores para realizar las principales funciones: (i) motor de agrupación de compañías (clustering), (ii) motor de procesado de noticias y análisis de sentimiento, (iii) motor de clasificación y

(iv) motor de trading. Además, necesita conexión a internet para el acceso a hemerotecas de distintos periódicos y datos de cotización de las empresas objetivo. A partir de estos se crean dos bases de datos locales a partir de las que se nutren los motores del sistema.

Por último, una serie de entradas hacen posible el cumplimiento de las funciones del sistema. Estas entradas son: (i) los tickers de las empresas objetivo, (ii) lexicón de las empresas objetivo, (iii) diccionarios de sentimiento financiero y (iv) análisis de noticias por expertos.

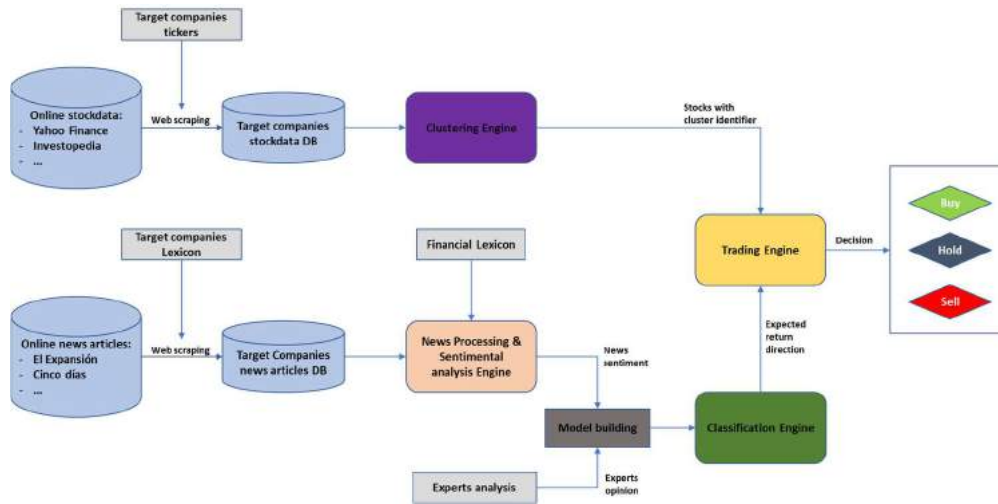


Ilustración 1. Arquitectura sistema creado. Fuente: elaboración propia.

4. Resultados

El resultado es un sistema original para explorar conjuntos de acciones, distintos métodos para el agrupamiento de sus STI, análisis técnico mediante índice de fuerza relativa y medias móviles y análisis de sentimiento de titulares de noticias del El Expansión y Cinco Días. El programa combina toda esta información para dar lugar a recomendaciones de carteras como la mostrada en Ilustración 2.

Hoja	Empresa	Cluster	RSI
29	"Pharma Mar, S.A."	1	37.0370
35	"Viscofan, S.A."	2	53.2110
22	"Industria de Diseño Textil, S.A."	3	28.5068
26	"Meliá Hotels International, S.A."	4	32.9231

Ilustración 2. Recomendación cartera creada por el programa diseñado. Fuente: elaboración propia.

Además, de las recomendaciones de carteras una serie de herramientas permiten al usuarios monitorizar sus inversiones y ayudarles a tomar decisiones. Entre las visualizaciones hay gráficos de medias móviles, análisis tendencia, índice de fuerza relativa, noticias importantes, cruces de medias móviles y gráficos de velas.



Ilustración 3. Visualización del sistema (izq: tendencias con medias móviles, der: análisis RSI) Fuente: elaboración propia.

5. Conclusiones

El programa ofrece multitud de funciones y permite la elección de una serie de parámetros para afinar los análisis en función de las necesidades particulares de cada usuario. Además, el código se plantea como una herramienta para su uso a lo largo del tiempo, permitiendo la extracción de información financiera y de noticias en tiempo real.

Del agrupamiento de compañías se puede concluir que el método k-medias es el más beneficioso ya que disminuye la carga computacional del sistema y da rendimientos superiores para el periodo estudiado. No obstante, es conveniente la ejecución de todos los métodos y experimentar en cada situación los resultados.

Por otro lado, del análisis de sentimiento en titulares de noticias se concluye que (i) existe un sesgo positivo en el sentimiento de los titulares de El Expansión y Cinco días, (ii) la gran diferencia en número de titulares correspondiente a distintas empresas hace que el sistema sea poco robusto, (iii) titulares de periódicos no siempre tienen relación con cotización de empresas y (iv) la existencia de un compromiso entre lexicón especializado en castellano y número de términos.

En cuanto a las visualizaciones el sistema proporciona una serie de gráficos novedosos útiles para el usuario pero sujeto a las limitaciones expresadas anteriormente en cuanto a falta de relación entre noticias y cotización de empresas.

Por último, se observa que el IBEX 35 se trata de un índice más dispar que otros similares como el DAX o el CAC (según las distancias seleccionadas). Además, el IBEX 35 no se comporta adecuadamente según los criterios de índice de fuerza relativa (RSI). Sin embargo, el análisis con medias móviles combinado con RSI es mucho más productivo obteniéndose rentabilidades de la cartera que batían al índice.

6. Referencias

He-Shan Guan, Q.-S. J. (2007). *Cluster Financial Time Series for Portfolio*. Beijing: Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition.

Markowitz, H. M. (1991). *Foundations of Portfolio Theory*. Journal of Finance.

DEVELOPMENT OF A SYSTEM OF INVESTMENT IN SHARES BASED ON THE CREATION OF PORTFOLIOS. TECHNICAL AND NEWS ANALYSIS FOR ITS EFFICIENT MANAGEMENT.

Author: Gil de Antuñano Durán, Diego.

Supervisor: Maté Jiménez, Carlos.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

This project will develop an original, basic and visual system to group the values of several ITS, allowing the selection of the best asset in each cluster. Consequently, portfolio alternatives could be generated for a set of stocks. Classifying these portfolios, using news information or other unstructured data sources through web scraping, and analysing the behaviour of the most efficient ones is one of the expected results of this project. The application to the assets of the Spanish stock market will be included, including those of the IBEX35.

Keywords: financial markets, clustering, news, ranking, relative strength index (RSI), moving averages.

1. Introduction

One of the wonders of financial markets is that they are able to reflect the situation in which we live. They are a numerical indicator, measurable and available to all the public in which all the events that have occurred so far are reflected. Therefore, a key factor in obtaining attractive returns in financial markets is the information that is available and when it is available. For this reason, and with the amount of information available today, it is essential to have an automated system for measuring the effect of news on the markets.

2. Project definition

The objective is the transformation of unstructured information present in news headlines into quantitative variables with which to explain the evolution of the price of companies on the stock market. Once this information has been analyzed, the next step is the creation of an original trading system based on the optimization of portfolios by grouping companies of the IBEX35 index, classification of news based on their impact on certain stocks, technical analysis of company prices, prediction of direction of time series of STI's interval and elaboration of portfolio recommendation.

To define the project, it is necessary to review the key components necessary to achieve the objectives. These components are:

- (i) Basic technical analysis of financial markets. The work is based on the relative strength index (RSI) and the analysis of trends and crossing of moving averages for the creation of trading strategies and portfolio selection.
- (ii) Portfolio creation theory. Markowitz's theory for portfolio management proposes risk minimization and profitability maximization. (Markowitz, 1991) The large number of stocks available in financial markets makes the use of the Markowitz

model difficult, inefficient, and sometimes impossible. (He-Shan Guan, 2007) For this reason, the solution of the problem is proposed by grouping companies and subsequently selecting the best action from each group for the creation of the portfolio. This selection is made through strategies based on relative strength index, moving averages and news analysis.

- (iii) Data grouping strategies. Clustering data is a purely descriptive unsupervised statistical analysis technique: through the search for patterns, the data are organized into clusters with similar characteristics. In this project, 3 clustering methods are proposed: clustering by means of k-means of two absolute variables, grouping by means of k-means of two variables using mean values, and hierarchical clustering.
- (iv) Extraction of news from the internet. Necessary element to obtain system inputs. Real-time information on stock market data and news headlines is accessed.
- (v) Market sentiment and news sentiment analysis. Review of Eugene Fama's Efficient Markets Hypothesis and analysis of supervised and unsupervised techniques for text analysis.
- (vi) Analysis of existing trading systems. Finally, a brief study is made of the existing trading systems and the tools that they make available to users. The project delves into the analysis of the Simply Wall Street and Webull trading systems and tools.

3. Description of the model / system / tool

The system created is the result of a series of iterations in which improvements have been implemented. The diagram in Illustration 1 is the basis that will serve as a reference for creating a program using the MATLAB 2020 application.

The system is designed to:

1. Extract stock information from the selected tickers in real time automatically and create a local database.
2. Extraction of news headlines related to the target companies from the main economic newspapers in Spain and creation of a local database.
3. Creation of groups of target companies based on the historical stock market data extracted in step 1.
4. Analysis of the sentiment of the news from the local database and its classification based on the expected direction of return they cause.
5. Creation of portfolio and investment recommendations and use of trading strategies.

As can be seen in the Illustration 1, the system has four engines to perform the main functions: (i) clustering engine, (ii) news processing engine and sentiment analysis, (iii) classification engine and (iv) trading engine. In addition, you need an internet connection to access the newspaper libraries of different newspapers and price data of the target

companies. From these two local databases are created from which the system engines are fed.

Finally, a series of inputs make it possible to fulfil the functions of the system. These entries are: (i) the tickers of the target companies, (ii) the lexicon of the target companies, (iii) dictionaries of financial sentiment and (iv) news analysis by experts.

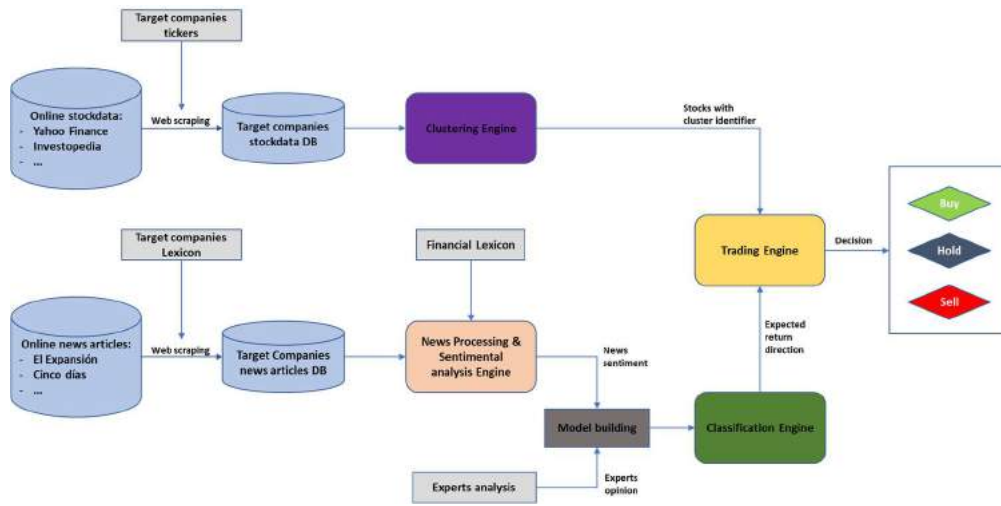


Illustration 4. System architecture created. Source: self-made

4. Results

The result is an original system for exploring sets of stocks, different methods for grouping their STIs, technical analysis using relative strength index and moving averages, and sentiment analysis of El Expansión and Cinco Días news headlines. The created program combines all this information to give rise to portfolio recommendations like the one shown in Illustration 2.

Hoja	Empresa	Cluster	RSI
29	"Pharma Mar, S.A."	1	37.0370
35	"Viscofan, S.A."	2	53.2110
22	"Industria de Diseño Textil, S.A."	3	28.5068
26	"Meliá Hotels International, S.A."	4	32.9231

Illustration 2. Portfolio recommendation created by the designed program. Source: self-made

In addition to portfolio recommendations, a series of tools allow users to monitor their investments and help them make decisions. Visualizations include moving average charts, trend analysis, relative strength index, major news, moving average crossovers, and candlestick charts.



Illustration 3. System visualizations (left: trends with moving averages, right: RSI analysis) Source: self-made)

5. Conclusions

This program offers a multitude of possible functions and allows the user to choose a series of parameters to fine-tune the analyses according to the particular needs of each user. In addition, the code is proposed as a tool for use over time, allowing the extraction of financial information and news in real time.

From the grouping of companies, it can be concluded that the k-means method is the most beneficial since it reduces the computational load of the system and gives superior performance. However, it is convenient to run all the methods and experiment with the results in each situation.

On the other hand, the analysis of sentiment in news headlines concludes that (i) there is a positive bias in the sentiment of the headlines of El Expansión and Cinco Días, (ii) the large difference in the number of headlines corresponding to different companies makes the system is not very robust, (iii) newspaper headlines are not always related to the companies and (iv) the existence of a compromise between specialized lexicon and number of terms.

Regarding the visualizations, the system provides a series of novel graphics useful for the user but subject to the limitations expressed regarding the lack of relationship between news and company prices.

Finally, it is observed that the IBEX 35 is a more disparate index than other similar ones such as the DAX or the CAC (according to the selected distances). Furthermore, the IBEX 35 does not perform adequately according to the relative strength index (RSI) criteria. However, the analysis with moving averages combined with RSI is much more productive, obtaining portfolio returns that beat the index.

6. References

- He-Shan Guan, Q.-S. J. (2007). *Cluster Financial Time Series for Portfolio*. Beijing: Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition.
- Markowitz, H. M. (1991). *Foundations of Portfolio Theory*. Journal of Finance.



MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

DESARROLLO DE UN SISTEMA DE INVERSIÓN EN
ACCIONES BASADO EN LA CREACIÓN DE CARTERAS.
ANÁLISIS TÉCNICO Y DE NOTICIAS PARA SU GESTIÓN
EFICIENTE.

TRABAJO FIN DE MÁSTER
MEMORIA

Autor: Diego Gil de Antuñano
Durán

Director: Carlos Maté Jiménez

Madrid

Julio de 2021

ÍNDICE:

Capítulo 1. Introducción	11
1.1 Motivación	12
1.2 Objetivos del proyecto.....	12
1.3 Alineación con los objetivos de desarrollo sostenible.....	13
1.4 Metodología de trabajo.....	13
1.5 Recursos a emplear.....	14
1.6 Estructura de la memoria.....	14
Capítulo 2. Estado de la cuestión.....	15
2.1 Estrategias básicas de predicción de precios: análisis técnico	15
2.1.1 Índice de fuerza relativa (RSI).....	16
2.1.2 Estrategia con medias móviles:.....	17
2.2 Creación de carteras	25
2.2.1 Fundamentos básicos de gestión de carteras: perfiles de inversor, rentabilidad, riesgo, covarianza y correlación.....	25
2.2.2 Teoría de Markowitz para la gestión de carteras.	27
2.2.3 Problemas del modelo de Markowitz y propuesta de solución.	29
2.3 Fundamentos de agrupación de datos: Clustering.....	31
2.3.1 Agrupación y predicción de series temporales de intervalo en los mercados financieros.	32
2.4 Minería de datos: web scraping.....	34
2.5 Sentimiento de mercado	34
2.5.1 Hipótesis de mercados eficientes y Behavioral Finance	35
2.5.2 Procesado de noticias: Análisis de sentimientos.....	36
2.6 Enfoque combinado y estrategias de trading.....	39
2.7 Sistemas de inversión en acciones y creación de carteras.	40
2.7.1 Simply wall Street: confident investment decisions.....	41
2.7.2 Webull.....	47
Capítulo 3. Sistema desarrollado	53
3.1 Arquitectura del sistema desarrollado.	53
3.2 Entradas al sistema	55
3.2.1 Tickers y lexicón compañías.....	55

3.2.2	<i>Lexicón financiero</i>	56
3.2.3	<i>Análisis supervisado de noticias por expertos</i>	57
3.3	Extracción de información de internet y creación de bases de datos: <i>web scraping</i>	58
3.3.1	<i>Extracción de noticias</i>	58
3.3.2	<i>Datos bursátiles</i>	60
3.4	Motor de agrupación de datos	60
3.4.1	<i>Agrupación natural</i>	61
3.4.2	<i>Agrupación método jerárquico</i>	61
3.4.3	<i>Agrupamiento con k-medias</i>	63
3.5	Motor de análisis de sentimiento de noticias.....	64
3.6	Motor de clasificación.....	65
3.6.1	<i>Aprendizaje no supervisado:</i>	66
3.6.2	<i>Aprendizaje supervisado</i>	67
3.7	Motor de trading.....	68
3.7.1	<i>Selección de carteras basadas en análisis de noticias</i>	68
3.7.2	<i>Selección de carteras basadas en análisis mediante rsi</i>	69
3.7.3	<i>Selección de carteras basado en análisis de rsi y medias móviles</i>	72
3.8	Visualizaciones.....	75
Capítulo 4. Experimentos y pruebas del sistema desarrollado		83
4.1	Agrupación método jerárquico IBEX 35.....	83
4.2	Aplicación clustering jerárquico al DAX 30	86
4.3	Aplicación <i>clustering</i> jerárquico al CAC 40	89
4.4	K- Medias aplicado al IBEX 35:	91
4.4.1	<i>K-medias 2 variables absoluto:</i>	91
4.4.2	<i>K-medias 2 variables media semanal:</i>	93
4.5	Evaluación rendimiento carteras	94
4.5.1	<i>K medias con variables rentabilidad-volatilidad total:</i>	95
4.5.2	<i>K medias con variables rentabilidad-volatilidad media</i>	96
4.5.3	<i>Jerárquico</i>	98
Capítulo 5. Conclusiones y futuros desarrollos.....		101
5.1	Conclusiones	101

5.1.1 Agrupación de compañías	102
5.1.2 Análisis de noticias	103
5.1.3 Visualización y monitorización	104
5.1.4 Conclusiones sobre IBEX 35	105
5.2 Futuros desarrollos	105
Capítulo 6. REFERENCIAS.....	107
Anexo 1: Encuesta análisis de titulares.....	113
Anexo 2: Objetivos de desarrollo sostenible	115

Índice de ilustraciones:

Ilustración 1. Órdenes de compra y venta mediante análisis de RSI. Fuente: elaboración propia.	17
Ilustración 2. Análisis de tendencia mediante media móvil simple BBVA. Fuente: elaboración propia	19
Ilustración 3. Cruce de la muerte, cruce dorado, triple cruce de medias móviles. Fuente: (Crespo, 2019)	20
Ilustración 4. Orden de venta con cruce de tres medias móviles. Fuente: (C., n.d.)	22
Ilustración 5. Análisis orden de compra y venta usando indicador MACD. Fuente: (¿Qué es el Indicador de Análisis Técnico MACD?, n.d.)	23
Ilustración 6. Creación de carteras teoría de Markowitz. Fuente: (Gubu, 2019) modificado.	28
Ilustración 7. Frontera de carteras eficientes. Fuente: (Ferrari, 2016)	29
Ilustración 8. Metodología de creación de carteras a partir de agrupación y análisis de noticias. Fuente: (Gubu, 2019) modificado.	30
Ilustración 9. Metodología agrupamiento STI. Fuente: (Cruz, 2019).	33
Ilustración 10. Funcionamiento de sistema de análisis de texto mediante Lexicón Fuente: (Albertos, 2018).....	37
Ilustración 11. Funcionamiento de sistema de análisis de texto mediante aprendizaje automático. Fuente: (Albertos, 2018).....	38
Ilustración 12. Ejemplo sistema de trading basado en noticias. Fuente: (Stefan Feuerriegel, July 19, 2018)	39
Ilustración 13. Checklist Banco Santander y perfil de empresas según Simply Wall Street. Fuente: Simply Wall Street.....	42
Ilustración 14. Cotización Banco Santander junto con eventos clave. Fuente: Simply Wall Street.....	43
Ilustración 15. Evento importante Banco Santander. Fuente: Simply Wall Street.	43
Ilustración 16. Valoración precio acciones Banco Santander. Fuente: Simply Wall Street.	44

Ilustración 17. Datos financieros Banco Santander. Fuente: Simply Wall Street.	45
Ilustración 18. Actualizaciones monitorización Iberdrola. Fuente: Simply Wall Street.	45
Ilustración 19. Creación de una cartera con Simply Wall Street. Fuente: Simply Wall Street.	46
Ilustración 20. Informe carteras Simply Wall Street. Fuente: Simply Wall Street.	46
Ilustración 21. Gráficos resumen financiero carteras Simply Wall Street. Fuente: Simply Wall Street.	47
Ilustración 22. Cotización Banco Santander junto con medias móviles de 5, 10, 20, 100 y 200 periodos e indicador MACD. Fuente: elaboración propia a partir de (Anquan, n.d.) ..	48
Ilustración 23. Personalización indicadores técnicos aplicación Webull. Fuente: elaboración propia a partir de (Anquan, n.d.) ..	49
Ilustración 24. Ejemplo información financiera disponible en Webull. Fuente: elaboración propia a partir de (Anquan, n.d.) ..	49
Ilustración 25. Ejemplo análisis de expertos Webull. Fuente: elaboración propia a partir de (Anquan, n.d.).....	50
Ilustración 26. Ejemplo pestaña noticias en análisis de Stocks en Webull. Fuente: elaboración propia a partir de (Anquan, n.d.).....	50
Ilustración 27. Búsqueda avanzada de Stock usando el menú Screener de Webull. Fuente: elaboración propia a partir de (Anquan, n.d.).....	51
Ilustración 28. Arquitectura del sistema desarrollado. Fuente: elaboración propia.	54
Ilustración 29. Extracción de noticias de Cinco Días. Fuente: elaboración propia a partir de (Cinco días, 2021) ..	59
Ilustración 30. Visualización cartera seleccionada mediante análisis de RSI. Fuente: elaboración propia.	76
Ilustración 31. Monitorización cotización e Índice de fuerza relativo. Fuente: elaboración propia.....	78
Ilustración 32. Monitorización media móvil simple y tendencia. Fuente: elaboración propia.	79
Ilustración 33. Gráficos cruce de la muerte y cruce dorado. Fuente: elaboración propia. ..	79

Ilustración 34. Análisis convergencia divergencia usando MACD. Fuente: elaboración propia.	80
Ilustración 35. Gráfico cotización junto con medias móviles y datos noticias. Fuente: elaboración propia.	81
Ilustración 36. Matriz de distancias. Fuente: elaboración propia.	84
Ilustración 37. Dendograma clustering jerárquico IBEX 35. Fuente: elaboración propia. .	85
Ilustración 38. Matriz distancias DAX 30. Fuente: elaboración propia.	87
Ilustración 39. Dendograma clustering jerárquico DAX 30. Fuente: Elaboración propia. .	88
Ilustración 40. Matriz distancias CAC 40. Fuente: elaboración propia.....	90
Ilustración 41. Dendograma clustering jerárquico CAC 40. Fuente: elaboración propia. ..	90
Ilustración 42. Agrupación IBEX 35 mediante k medias con dos variables absolutas. Fuente: elaboración propia.	92
Ilustración 43. Agrupación IBEX 35 con k medias con dos variables medias semanales. Fuente: elaboración propia.	93
Ilustración 44. Comparación rendimiento para agrupación con k medias absoluto. Fuente: elaboración propia.	96
Ilustración 45. Comparación rendimiento para agrupación con k medias con variables medias. Fuente: elaboración propia.	98
Ilustración 46. Comparación rendimiento para agrupación con método jerárquico. Fuente: elaboración propia.	100

Índice tablas:

Tabla 1. Planificación. Fuente: elaboración propia.	13
Tabla 2. Formato entradas al sistema Tickers y Lexicón compañías objetivo. Fuente: elaboración propia.	55
Tabla 3. Análisis diccionarios en castellano. Fuente: elaboración propia.	57
Tabla 4. Base de datos titulares de noticias. Fuente: elaboración propia.	59
Tabla 5. Formato extracción datos bursátiles. Fuente: elaboración propia.	60
Tabla 6. Agrupación natural empresas IBEX 35. Fuente: elaboración propia.	61
Tabla 7. Algoritmo para el cálculo de distancia entre activos. Fuente: elaboración propia a partir de (He-Shan Guan, 2007).	62
Tabla 8. Resultado del análisis de sentimiento mediante Lexicón con LIWC. Fuente: elaboración propia.	65
Tabla 9. Agrupación noticias Aena por día. Fuente: elaboración propia.	66
Tabla 10. Algoritmo selección cartera basándose en noticias. Fuente: elaboración propia	69
Tabla 11. Algoritmo selección cartera basándose en RSI. Fuente: elaboración propia.	70
Tabla 12. Asignación RSI y RSI binario a empresas objetivo. Fuente: elaboración propia.	71
Tabla 13. Cartera resultado del análisis basado en RSI. Fuente: elaboración propia.	71
Tabla 14. Algoritmo selección cartera basándose en tendencias mediante media móvil simple. Fuente: elaboración propia.	73
Tabla 15. Asignación tendencias y cruce dorado y cruce de la muerte. Fuente: elaboración propia.	74
Tabla 16. Cartera resultado del análisis basado en RSI y medias móviles. Fuente: elaboración propia.	74
Tabla 17. Petición de elección de la empresa por parte del programa. Fuente: elaboración propia.	77
Tabla 18. Índice IBEX 35. Fuente: elaboración propia.	83
Tabla 19. Agrupación empresas IBEX 35 según clustering jerárquico. Fuente: elaboración propia.	86

Tabla 20. Índice DAX 30. Fuente: elaboración propia.....	86
Tabla 21. Agrupación empresas DAX 30 según clustering jerárquico. Fuente: elaboración propia.....	89
Tabla 22. Índice CAC 30. Fuente: elaboración propia.....	89
Tabla 23. Agrupación empresas CAC 40 según clustering jerárquico. Fuente: elaboración propia.....	91
Tabla 24. Agrupación empresas IBEX 35 según k medias con variables absolutas. Fuente: elaboración propia.	93
Tabla 25. Agrupación empresas IBEX 35 mediante k medias con variables medias. Fuente: elaboración propia.	94
Tabla 26. Cartera generada usando k medias con variables absolutas y análisis mediante índice de fuerza relativa. Fuente: elaboración propia.....	95
Tabla 27. Cartera generada usando k medias con variables absolutas y análisis mediante índice de fuerza relativa y medias móviles. Fuente: elaboración propia.....	95
Tabla 28. Cartera generada usando k medias con variables absolutas y análisis de noticias. Fuente: elaboración propia	96
Tabla 29. Cartera generada usando k medias con variables medias y análisis mediante índice de fuerza relativa. Fuente: elaboración propia.....	97
Tabla 30. Cartera generada usando k medias con variables medias y análisis mediante índice de fuerza relativa y medias móviles. Fuente: elaboración propia.....	97
Tabla 31. Cartera generada usando k medias con variables medias y análisis de noticias. Fuente: elaboración propia	97
Tabla 32. Cartera generada usando método jerárquico y análisis mediante índice de fuerza relativa. Fuente : elaboración propia	98
Tabla 33. Cartera generada usando método jerárquico y análisis mediante índice de fuerza relativa y medias móviles. Fuente: elaboración propia	99
Tabla 34. Cartera generada usando método jerárquico y análisis de noticias. Fuente: elaboración propia	99

Capítulo 1. INTRODUCCIÓN

Una de las maravillas de los mercados financieros es que son capaces de reflejar la situación en la que vivimos. Son un indicador numérico, medible y disponible a todo el público en el que se ven reflejadas todas las noticias que han ocurrido hasta el momento. Cada noticia afecta de una u otra manera a los mercados creando un impacto que queda reflejado en forma de subidas o bajadas de índices, ETFs, stock...

Por lo tanto, un factor clave para obtener rentabilidades atractivas en mercados financieros es la información que se dispone y el momento en la que se tiene. En el siglo XXI existen multitud de fuentes de información y noticias a las que tenemos acceso todos a través de internet. Con la información adecuada a tiempo se facilita la toma de decisiones para obtener mayores retornos o, evitar grandes pérdidas al invertir en mercados financieros. Según autores como (Schumaker, 2009), (Mittermayer, 2004) el efecto de una noticia en bolsa tarda entre 20-60 minutos desde la publicación de la noticia. Por este motivo, y con la cantidad de información disponible hoy en día es imprescindible disponer de un sistema automatizado para la medición del efecto de noticias en los mercados.

La información adecuada implica información fiable y relevante. Estos son los dos aspectos clave a tener en cuenta. Por lo tanto, de cara a este Trabajo de Fin de Máster (en adelante TFM) será imprescindible el uso de webs fiables y robustas de cara a tomar decisiones igual de sólidas.

En segundo lugar, tener la información a tiempo ofrece una ventaja competitiva al usuario brindándole la posibilidad de comprar o vender stock antes que el resto del mercado.

Con esto en mente propongo la creación de un sistema con el que obtener, clasificar y procesar información fiable de la web de forma rápida para obtener una ventaja sobre el resto de inversores. Además, se propone la creación de un sistema de análisis de dicha información con el objetivo de explicar retornos en bolsa y, desarrollar un sistema atractivo de trading y

creación de carteras combinando estrategias de agrupación de datos, teorías tradicionales de trading mediante análisis técnico y estrategias innovadoras de uso de noticias.

1.1 MOTIVACIÓN

A lo largo de la historia del sector financiero se ha nutrido de otras especialidades para optimizar sus procesos. El siglo XXI marcado por la irrupción de las tecnologías y el tratamiento de los datos da nuevas oportunidades en múltiples campos. La aplicación de estas técnicas supone una motivación en sí.

Además, a pesar de disponer de más información que nunca al alcance de todos se hace extremadamente complicado darle uso. Uno de los grandes retos existentes es procesar y unificar toda esta información. La motivación de este TFM es el análisis y procesado de información útil para aplicar al sector financiero, partiendo de datos desestructurados como son noticias en los periódicos, combinado con datos numéricos de trading técnico tradicional. Un reto adicional es la complejidad que requiere este tipo de análisis debido al carácter subjetivo de las noticias y el hecho de que apenas existan este tipo de estudios en lengua española (Albertos, 2018).

1.2 OBJETIVOS DEL PROYECTO

El objetivo es la transformación de información desestructurada presente en titulares de noticias en variables cuantitativas con las que explicar la evolución de la cotización de empresas en bolsa. Una vez analizada esta información, se plantea la creación de un sistema original de trading basado en la optimización de carteras mediante agrupación de empresas del índice IBEX35, clasificación de noticias en función de su repercusión en determinadas acciones, análisis técnico de cotización de empresas, predicción de dirección de series temporales de intervalo STI's y elaboración de recomendación de carteras.

Dado lo novedoso e inexplorado que es el campo del análisis de sentimiento de textos en idioma español existen limitaciones que pueden hacer variar los objetivos.

Independientemente de ello, se pretende crear y sentar las bases del primer sistema de trading (hasta donde se conoce actualmente) basado en análisis de sentimiento de noticias en castellano y agrupación de STI's para aportar a la comunidad científica y fomentar la creación de este tipo de estudios.

1.3 ALINEACIÓN CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE

Este proyecto está ligado con los siguientes objetivos de desarrollo sostenible:

- ODS 8, Promover el crecimiento económico inclusivo y sostenible, el empleo y el trabajo decente para todos. Se busca entender el crecimiento o decrecimiento del mercado. Además, una vez creada la herramienta y en caso de correcto funcionamiento se plantea la creación de una web donde poner a disposición del público la información relevante para el análisis de noticias.

1.4 METODOLOGÍA DE TRABAJO

El método de trabajo del presente proyecto consiste en: (i) la obtención de datos fiables mediante técnicas de extracción de datos de internet (*web scraping*), (ii) preprocesado y agrupación de datos bursátiles, (iii) análisis sentimientos, (iv) clasificación de empresas según sentimiento y (v) creación de carteras.

La cronología del proyecto se muestra en la Tabla 1:

Planning	2020		2021					
Objetivos	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	Junio
Estudio del estado del arte								
Web scraping								
Preprocesado y clustering								
Análisis de sentimientos								
Clasificación y forecasting								
Creación de interfaz								

Tabla 1. Planificación. Fuente: elaboración propia.

1.5 RECURSOS A EMPLEAR

Se emplea el software de Matlab R2020b, con el que se realizarán tareas de extracción de noticias de internet, obtención de datos bursátiles, preprocesado de información, creación de bases de datos en Excel, exportación documentos a .txt, procesado y análisis de información, visualización y automatización de procesos.

Además, se emplea LIWC para análisis de textos mediante léxicos, Excel para el almacenamiento de datos y Onedrive para compartir material online.

1.6 ESTRUCTURA DE LA MEMORIA

La memoria se divide en 6 capítulos y 2 anexos. En el primer capítulo se introduce la motivación para la realización de este trabajo. Se fijan una serie de objetivos y explica la metodología y recursos a emplear. En segundo capítulo se comenta el estado del arte. Se introducen las principales teorías de análisis técnico de trading, teorías financieras de creación de carteras profundizando e, teorías sobre la relación entre mercados financiero e información. Adicionalmente se introducen estrategias minería de datos de la web, análisis de sentimiento y agrupamiento de datos. En este capítulo se revisa también el estado del arte de sistemas de análisis de datos bursátiles y trading existentes. En el tercer capítulo se presenta el sistema desarrollado. Para ello se desglosa cada uno de los elementos claves para sus funcionamiento, desde la creación de bases de datos hasta los motores encargados de llevar a cabo funciones como la clasificación o agrupamiento de datos. También se presenta un apartado dedicado a la visualización de datos. Este sirve como una herramienta de apoyo al sistema creado para monitorizar las acciones de forma visual. En el cuarto capítulo se pone a prueba el sistema simulando la creación de una serie de carteras y observando sus rendimientos. En quinto apartado se exponen las conclusiones del proyecto y se proponen aspectos de mejora para desarrollos futuros. Por último, en el sexto capítulo se incluye la bibliográfica de apoyo para la realización de este TFM.

Capítulo 2. ESTADO DE LA CUESTIÓN

Para la elaboración de este TFM se parte de las hipótesis desarrolladas por (Nuij) por las cuales: (i) existe una relación entre las noticias y los mercados financieros, (ii) el impacto de causado por una noticia puede medirse en términos de cómo va a afectar a los mercados financieros.

Los componentes claves para la consecución de este TFM son distintas técnicas y ramas de conocimiento que usaremos para la obtención de resultados. Los principales componentes son: (i) análisis técnico básico de mercados financieros, (ii) teoría de creación de carteras, (iii) estrategias de agrupación de datos, (iii) extracción de noticias de internet, (iv) sentimiento de mercado y análisis de sentimiento de noticias, (v) estrategias de trading derivadas del resultado del procesado de noticias y la creación de carteras y, (vi) análisis de sistemas existentes de trading. En este capítulo se hará una introducción y estudio bibliográfico de cada uno de esos componentes.

2.1 ESTRATEGIAS BÁSICAS DE PREDICCIÓN DE PRECIOS: ANÁLISIS TÉCNICO

Las técnicas básicas de predicción de precios sirven de apoyo y complemento para el desarrollo un modelo más complejo en el que se incorporara el efecto de las noticias.

Autores como (Nuij) y (Montero, 2020) hacen uso del análisis técnico para definir modelos conjuntamente con el análisis de noticias. Entre los métodos usados destacan el análisis mediante media móvil simple, media móvil exponencial, regresiones lineales, estrategias de *momentum trading* y análisis de índices como el índice de fuerza relativa más conocido por sus siglas en ingles RSI. En este trabajo se profundiza en el análisis técnico mediante RSI y medias móviles los cuales se explican a continuación.

2.1.1 ÍNDICE DE FUERZA RELATIVA (RSI)

En este trabajo se ha optado por el uso del índice de fuerza relativa (RSI) como indicador de momento. Este índice, desarrollado por J.Welles Wilder en 1978, es un indicador de momento de mercado que mide la velocidad en el cambio del precio en acciones para evaluar si las acciones están sobrecompradas o sobrevendidas. Este indicador se muestra como un oscilador que se mueve entre 0 y 100.

El cálculo se realiza usando datos pasados de la cotización de la acción. Típicamente se usan los últimos 14 días siguiendo la fórmula que se muestra a continuación:

$$RSI_{14 \text{ días}} = 100 - \frac{100}{1 - \frac{\text{Ganancias promedio}}{\text{Pérdidas promedio}}}$$

Las pérdidas y ganancias promedio se calculan en forma de porcentaje como la suma de los porcentajes de incremento o decremento de una acción durante los últimos 14 días entre 14.

Este análisis requiere una monitorizar de los valores del RSI para cada compañía. Siguiendo a Wilder una acción esta sobrecomprada cuando el RSI es superior al 70% y un RSI inferior al 30% indica que la acción esta sobrevendida. Por lo tanto se da lugar a las siguientes reglas:

- Si la señal tiene un valor inferior a 30 se manifiesta que el valor de la acción ha caído muy rápidamente y por lo tanto es posible que se produzca un rebote en sentido contrario y suba el precio de la acción. En estos casos se debe lanzar una orden de compra. Alcanzar un valor inferior a 30 no implica el inmediato rebote del valor de las acciones sino que se trata de una referencia. La mejor oportunidad de compra surge ante valores mínimos absolutos de RSI, los cuales pueden llegar a encontrarse en torno a 20 o menos.
- Si la señal supera la barrera de los 70 es un indicador de la acción está siendo sobrecomprada. Esto significa que el precio ha subido muy rápidamente y es posible que rebote en sentido contrario.

A continuación se muestra una imagen ejemplo del índice de fuerza calculado para el Banco Santander durante el periodo diciembre 2020 – abril 2021. Como se puede observar, el mejor momento para la venta se produce con valores máximos de RSI que llegan a estar en torno al 89. Por otro lado, los valores mínimos llegan a encontrarse en torno a los 18 puntos indicando recomendación de compra.



Ilustración 1. Órdenes de compra y venta mediante análisis de RSI. Fuente: elaboración propia.

2.1.2 ESTRATEGIA CON MEDIAS MÓVILES

Una media móvil se puede definir como un promedio de período constante, generalmente de precios, es decir, calculado para cada intervalo de período de gráfico sucesivo. Las medias móviles son un indicador técnico que indica la tendencia seguida gracias al uso de valores promedios.

Existen distintos tipos de medias móviles. Las más comunes usadas en el ámbito del trading son: media móvil simple (SMA) y media móvil exponencial (EMA). No obstante, existen

otros métodos usados como la media móvil geométrica (GMA) o la media móvil ponderada. Las medias móviles se calculan normalmente usando el precio de cierre de las acciones pero no significa que no se pueda calcular usando los precios máximo o mínimos. (Crespo, 2019)

La más común de todas, la media móvil simple (SMA), queda definida como:

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n}$$

Donde:

$$A_n = \text{precio de la acción en el periodo } n$$

$$n = \text{número total de periodos}$$

Por otro lado, la media móvil exponencial (EMA) pondera los valores dando más importancia a aquellos más recientes.

La aparición de las medias móviles da lugar a distintas herramientas que facilitan el trading como el análisis de tendencias o la eliminación de ruido. Además, surgen una serie de estrategias que establecen pautas para la inversión. Las estrategias de medias móviles más comunes son: (i) trading de tendencias con múltiples medias, (ii) cruce de precio con medias móviles, (iii) cruce de dos medias móviles (cruce dorado/cruce de la muerte), (iv) cruce de tres medias móviles, (v) indicador de convergencia y divergencia de medias móviles (MACD).

A continuación se explican las principales estrategias:

- (i) Trading básico o trading de tendencias mediante medias móviles. La regla básica para operar con la media móvil simple es que si el precio de una acción está por encima de su media simple (SMA) entonces la acción se encuentra en una tendencia alcista y por lo tanto se recomienda orden de compra. Por otro lado, si el precio de una acción está por debajo de su media móvil simple está en una tendencia bajista y se recomienda la venta. (Hayes, 2021)

En la siguiente imagen se muestra un ejemplo aplicado a la cotización de BBVA. Durante los meses de noviembre y diciembre se puede observar una mayor diferencia con signo positivo entre cotización y media móvil simple la cual se traduce en una tendencia alcista del precio.



Ilustración 2. Análisis de tendencia mediante media móvil simple BBVA. Fuente: elaboración propia

Distintas medias móviles pueden ser usadas para la detección de la tendencia de una acción. Medias móviles con periodos de estudios más cortos (5-20 periodos) aportan información útil para el trading diario y semanal pero son poco útiles como herramientas para filtrar ruido o estrategias de inversión en largo plazo. Por lo contrario, medias móviles largas (100-200 periodos) sirven para la inversión en el largo plazo.

- (ii) Cruce de precio con medias móviles. Esta estrategia indica posibles cambios en la tendencia en función del cruce del precio de la acción (normalmente se usa el precio de cierre del mercado) con distintas medias móviles generadas. Por lo tanto, esta estrategia sirve como un *trigger* para lanzar órdenes de compra o venta de acciones. Para ello se ha de seguir las siguientes normas:

- a. Regla 1: si el precio de la acción cruza ascendente a la media móvil es un indicador de tendencia creciente e se recomienda lanzar una orden de compra.
 - b. Regla 2: si el precio de la acción cruza descendente a la media móvil es un indicador de tendencia descendente y se recomienda lanzar orden de venta.
- (iii) Cruce dorado/cruce de la muerte. Similar al cruce de medias móviles con precio de acciones. Esta estrategia estudia el cruce entre dos medias móviles de distinta longitud para activar órdenes de compra o venta. Se traduce en las siguientes reglas:
- a. Regla 1 o cruce dorado: si una media móvil rápida/corta (ej. 10 periodos) cruza ascendente a una media móvil lenta/larga (ej. 50 periodos) es un indicador de tendencia creciente y, por lo tanto, se recomienda lanzar una orden de compra. Este cruce es conocido por el nombre de cruz dorada.
 - b. Regla 2 o cruce de la muerte: si la media móvil rápida (ej. 10 periodos) cruza descendente a una media móvil lenta (ej. 50 periodos) es indicador de tendencia decreciente y, por lo tanto, se recomienda lanzar una orden de venta. Este cruce recibe el nombre de cruz de la muerte.

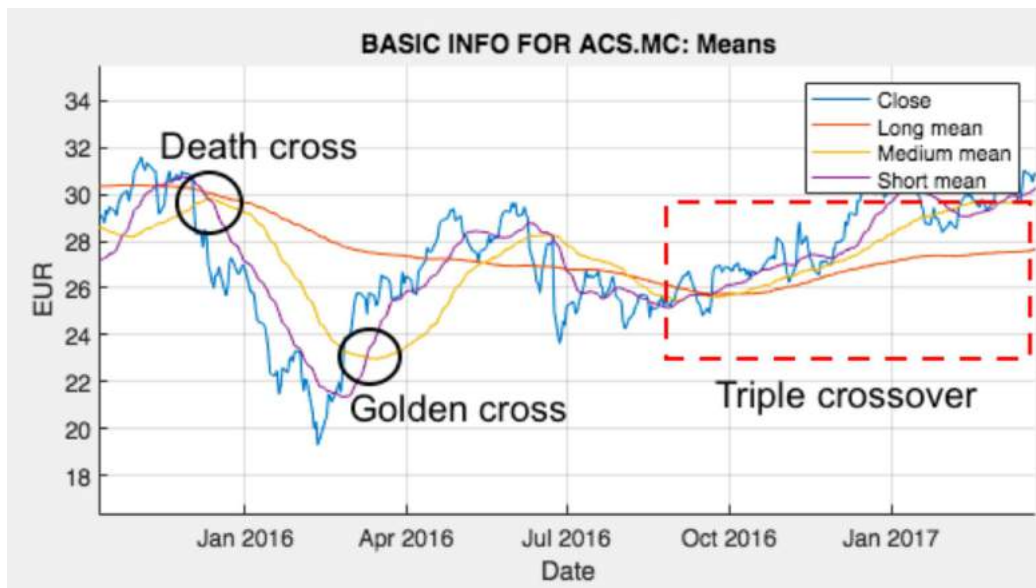


Ilustración 3. Cruce de la muerte, cruce dorado, triple cruce de medias móviles. Fuente: (Crespo, 2019)

- (iv) Cruce de tres medias móviles. Esta estrategia fue definida por Scott Lowry en su libro “*The Magic of Moving Averages*” y, al igual que el cruce dorado y cruce de la muerte sirve para analizar situaciones de compra o venta de distintos instrumentos de trading. El triple cruce de medias móviles es mucho menos frecuente que otros tipos de cruce pero mucho más efectivo. Esto se debe a que al incorporar una tercera media se producen menos señales falsas. (finanzas.com, 2016)

La estrategia original de Lowry se basa en la búsqueda de la convergencia de tres medias móviles exponenciales de 4, 18 y 40 periodos siendo posible adaptar el número de periodos a nuestros gustos en base a nuestra experiencia. De esta forma, autores como (C., n.d.) usan medias móviles exponenciales de 10, 25 y 50 periodos.

Las reglas a seguir para lanzar órdenes de compra o venta con este tipo de cruce son:

- a. Regla 1 u orden compra: la media móvil rápida (4 periodos) cruza ascendente primero a la media móvil media (18 periodos) y acto seguido a la media móvil lenta (40 periodos). En este caso se debe lanzar una orden de compra.
- b. Regla 2 u orden de venta: esta señal se da cuando la media móvil rápida (4 periodos) cruza descendente primero a la media móvil de velocidad media (18 periodos) y acto seguido a la media móvil lenta (40 periodos).

Entre otras funciones esta estrategia se puede usar para la creación de ordenes *stop-loss* que consisten en retirarse del mercado en caso de detectar peligro y bajadas para proteger del inversor. En la Ilustración 4 se puede ver una ejemplo de orden *stop-loss* mediante triple cruce de medias móviles.



Ilustración 4. Orden de venta con cruce de tres medias móviles. Fuente: (C., n.d.)

- (v) Indicador de convergencia y divergencia de medias móviles (MACD). Este indicador se diferencia del resto que usa de medias móviles exponenciales en lugar de simples. Tiene dos componentes, el primero de los componentes es el MACD, que es la diferencia entre las medias móviles exponenciales de 26 periodos y la de 12 periodos. El segundo componente es la media móvil exponencial de 9 días también conocida como señal. El uso de este indicador es muy sencillo siguiendo una serie de normas que indican órdenes de compra o venta:
- a. Regla 1: cada vez que la MACD cruza de abajo a arriba a la señal de la media móvil de 9 días se lanza una orden de compra de esa acción.
 - b. Regla 2: cada vez que la MACD cruza de arriba a abajo la señal de la media móvil de 9 días se lanza una orden de venta de esa acción.



Ilustración 5. Análisis orden de compra y venta usando indicador MACD. Fuente: (¿Qué es el Indicador de Análisis Técnico MACD?, n.d.)

Para este proyecto se analizará la tendencia de las acciones, el indicador de convergencia y divergencia de medias móviles (MACD) y el cruce dorado/cruce de la muerte y se propondrán futuros desarrollos para el cruce de media móviles con precio y triple cruce de medias móviles.

2.1.2.1 Selección de medias móviles

Uno de los principales problemas del uso de medias móviles es la selección del número correcto de periodos a analizar para evitar señales falsas de compra o venta. Existen distintas opiniones al respecto: en función del objetivo de la inversión, el horizonte temporal estudiado, el índice analizado... En este apartado se explicaran algunos de las opiniones más comunes.

- i) Medias móviles simples de 50, 100, 200. Se tratan de los valores más comunes para la sección de periodos. Se usan medias móviles de 50 periodos para el corto plazo, de 100 periodos para el medio plazo y de 200 periodos para el largo plazo. A pesar de ser los valores más aceptados la mayoría de inversores están de

acuerdo en la necesidad de experimentar con distintas medias móviles en cada situación.

- ii) Otras medias móviles. A raíz de modelos iterativos distintos autores han llegado a apoyar el uso de medias móviles de 20, 50, 100 periodos. Además, también existen autores se inclinan por el uso de medias móviles siguiendo patrones de Fibonacci siendo 89 periodos uno de los valores más populares. (Crespo, 2019)

Independientemente de las distintas opiniones la selección de medias móviles es muy dependiente del horizonte temporal de estudio y el nivel de riesgo a asumir. Se puede concluir que inversores que busquen rentabilidades en el largo plazo tienden a usar medias móviles de 100 o 200 periodos. Con ellas se consigue eliminar ruido y hacer frente a fluctuaciones en el valor de la cotización de las empresas. El resultado es una toma de decisiones robusta evitando órdenes de compra o venta falsas, siendo una estrategia propia de inversores conservadores. La principal desventaja de estas medias móviles es el retardo con el que se detectan señales de entrada o salida al mercado pudiendo perder oportunidades trading diario, semanal y mensual.

Por otro lado, para perfiles de inversores más agresivos, que buscan rentabilidades en el corto plazo y hacen trading semanal y mensual son populares las medias móviles de 5, 10, 20 periodos. Estas se ciñen más al valor actual de la cotización de empresas y detectan con más frecuencia señales de compra o venta. En contraposición, esas media móviles tienen más probabilidad de notificar ordenes faltas al usuario.

Para el caso de estudio, se está analizando un horizonte temporal de entre 90 y 120 días. Por lo tanto no es posible hacer uso de medias móviles de 50, 100, 200 periodos. Parece sensato usar un múltiplo de estos valores por lo que se hacen pruebas para medias móviles de 5, 10 y 20 periodos. Como se ha explicado anteriormente estas pueden aumentar el riesgo si se busca inversión en el largo plazo. Sin embargo, al tratarse de un sistema de creación de carteras el riesgo que se asume al usar medias móviles de menor tamaño se reduce por la diversificación en empresas.

2.2 CREACIÓN DE CARTERAS

La creación de carteras es un elemento clave a la hora de invertir. Una cartera es una combinación de activos que cumplen con las características deseadas por el inversor y en los que invierte. Ya sea en mercados capitales, real estate o inversiones alternativas es recomendable la creación de una cartera para diversificar y mitigar riesgos. En ese apartado se introducen los fundamentos de gestión y creación de carteras.

2.2.1 FUNDAMENTOS BÁSICOS DE GESTIÓN DE CARTERAS: PERFILES DE INVERSOR, RENTABILIDAD, RIESGO, COVARIANZA Y CORRELACIÓN

La mayoría de estrategias de gestión de carteras se basan en análisis conjunto de rentabilidad y riesgo. Se busca el equilibrio óptimo entre estos componentes en función del tipo de inversor. Según (MytripleA, 2017) podemos diferenciar tres tipos de inversores en función de su aversión al riesgo:

- (i) Perfil conservador (*risk averse*): evitan cualquier tipo de pérdida, buscan menor riesgo y por consiguiente exigen menor rentabilidad.
- (ii) Perfil moderado (neutral): buscan el equilibrio entre rentabilidades atractivas sin asumir demasiado riesgo.
- (iii) Perfil agresivo (*risk-seeking*): buscan rentabilidades elevadas y asumen gran cantidad de riesgo en sus operaciones.

Una vez conocido el perfil del inversor es necesario definir una serie de variables numéricas medibles y objetivas que sirvan como indicador de rentabilidad y riesgo. A continuación se presentan los conceptos básicos necesarios para el análisis y creación de carteras.

La rentabilidad se mide a través del retorno de la acción (R) y *rate of return* o tasa de retorno (r):

$$R = \frac{x_1}{x_0}$$

$$r = \frac{x_1 - x_0}{x_0}$$

x_1 : precio de la acción en el momento de venta.

x_0 : precio de la acción en el momento de venta.

Por otro lado, el riesgo de una acción se suele medir mediante la varianza y desviación estándar de sus retornos. La varianza de una población viene dada por:

$$\sigma^2 = \frac{\sum_{i=1}^n (R_t - \mu)^2}{n}$$

Sin embargo, en el campo de los mercados financieros y con los recursos disponibles, no se analizará la población entera sino una muestra en un periodo de tiempo. Por este motivo, el cálculo de la varianza muestral será:

$$s^2 = \frac{\sum_{i=1}^n (R_t - \bar{R})^2}{n - 1}$$

La covarianza (Cov) es un indicador de la relación lineal que existe entre dos variables. En este estudio se usará para detectar las similitudes entre pares de acciones a través del cálculo de la covarianza entre retornos de acciones. Su fórmula matemática aplicada al caso de estudio es:

$$Cov_{1,2} = \frac{\sum_{i=1}^n [(R_{t,1} - \bar{R}_1) * (R_t - \bar{R}_2)]}{n - 1}$$

La covarianza es una medida absoluta que depende la magnitud de los retornos de las acciones estudiadas. Para comparar entre multitud de acciones se estandariza la covarianza obteniéndose el coeficiente de correlación (ρ):

$$\rho_{1,2} = \frac{Cov_{1,2}}{\sigma_1 * \sigma_2}$$

La interpretación de este coeficiente es:

- $\rho = +1$: correlación positiva perfecta.
- $\rho = 0$: no existe correlación lineal entre stocks.
- $\rho = -1$: correlación negativa perfecta.

Por último, el ratio de liquidez o *sharpe ratio* es un indicador de la calidad de un activo muy usado en la bibliografía (Pan, 2020), (Gubu, 2019). Este indicador es una medida de riesgo y retorno que sirve para la selección del mejor stock de un conjunto y se define como:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p}$$

2.2.2 TEORÍA DE MARKOWITZ PARA LA GESTIÓN DE CARTERAS

Entre las teorías de gestión de carteras tiene especial importancia la teoría de Markowitz o *Mean-Variance Portfolio Theory* (teoría de la cartera de varianza media) de Harry Markowitz quién ganó el premio Nobel de Economía por su trabajo en el desarrollo de modelos eficientes de portfolios.

Esta teoría parte de una serie de hipótesis. La primera hipótesis es la existencia de incertidumbre. Un inversor que conociese con certeza el retorno futuro de las acciones presentes en el mercado invertiría todo su patrimonio en aquella con mayor rentabilidad. En caso de haber dos acciones con la misma máxima rentabilidad futura el inversor no tendría preferencias entre que acción o combinación de las acciones comprar. Sin embargo, al existir incertidumbre existe riesgo y por lo tanto es sensato y necesario diversificar inversiones dando lugar a carteras. (Markowitz, 1991)

Por lo tanto, el primer paso es identificar que es el riesgo y como se mide. En su trabajo Markowitz propone como medidas de incertidumbre y riesgo la varianza (o desviación típica) o la semi-varianza. Por otro lado, para medir la rentabilidad propone la tasa de retorno. El objetivo del modelo de Markowitz es maximizar el rendimiento a la vez que minimiza el riesgo. Esta teoría es el inicio de la teoría moderna de creación de carteras. Sostiene que

cuanto menor sea la correlación en el precio de los activos, mayor aversión al riesgo tiene la cartera (Pan, 2020). A continuación se explica brevemente el modelo.

Se parte de un valor total de la cartera representado por x_0 , el cual se desea invertir en n activos de valor $x_{0,i}$ cada uno con $i = 1, 2, 3, \dots, n$. La proporción del total de la cartera a invertir en cada activo viene dada por un vector de pesos $w = (w_1, w_2, w_3, \dots, w_n)$. Por lo tanto, para mantener el presupuesto de la cartera se tiene que cumplir $\sum_{i=1}^n w_i = 1$. El valor de los pesos puede ser negativo en caso de posiciones a corto ampliándose el abanico de posibilidades en la creación de portfolios. Sin embargo, en este trabajo para evitar complicaciones matemáticas se impondrá la restricción de comprar a corto. Por lo tanto, $0 < w_i < 1 \forall i$.

El vector con la tasa de retorno histórica de los activos de la cartera se denota con $z^T = (r_1, r_2, r_3, \dots, r_n)$ y a partir de él se calcula su covarianza (Cov(z) o Σ) que es el indicador de volatilidad y riesgo.

Por otro lado, el vector $m^T = (E(r_1), E(r_2), E(r_3), \dots, E(r_n)) = (\mu_1, \mu_2, \mu_3, \dots, \mu_n)$ representa el valor esperado medio de la tasa de retorno o *rate of return* de cada activo siendo el indicador de rentabilidad esperada para el modelo.

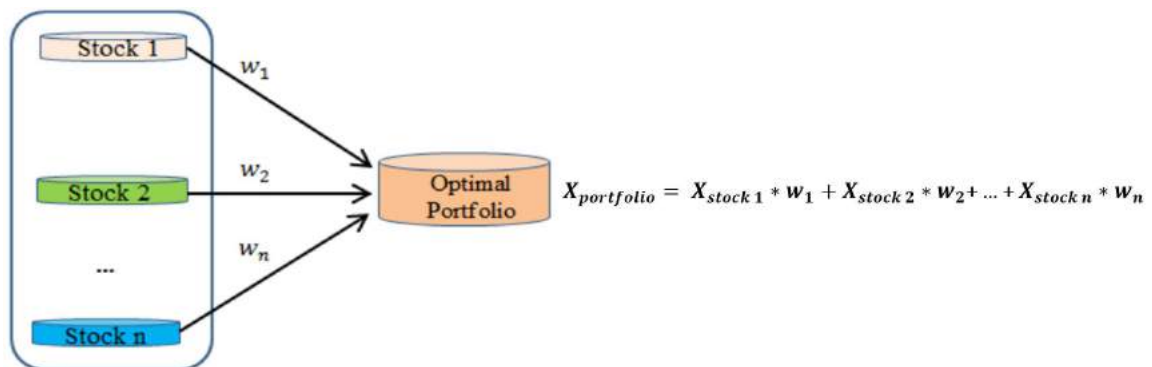


Ilustración 6. Creación de carteras teoría de Markowitz. Fuente: (Gubu, 2019) modificado.

Podemos calcular la tasa de retorno de la cartera (rentabilidad):

$$r = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \dots \\ \mu_n \end{pmatrix} * (w_1, w_2, w_3, \dots, w_n)$$

Por último, para minimizar la volatilidad/riesgo de la cartera y maximizar la rentabilidad se resuelve el siguiente problema de optimización:

$$\begin{aligned} \min & \quad \left\{ \frac{1}{2} w^\tau * (\Sigma) * w - \lambda * m^\tau * w \right\} \\ \text{cond. 1:} & \quad \sum_{i=1}^n w_i = 1 \\ \text{cond. 2:} & \quad 0 < w_i < 1 \forall i. \end{aligned}$$

El valor λ marca la frontera de carteras eficiente y carteras no eficientes como se muestra en la siguiente imagen.

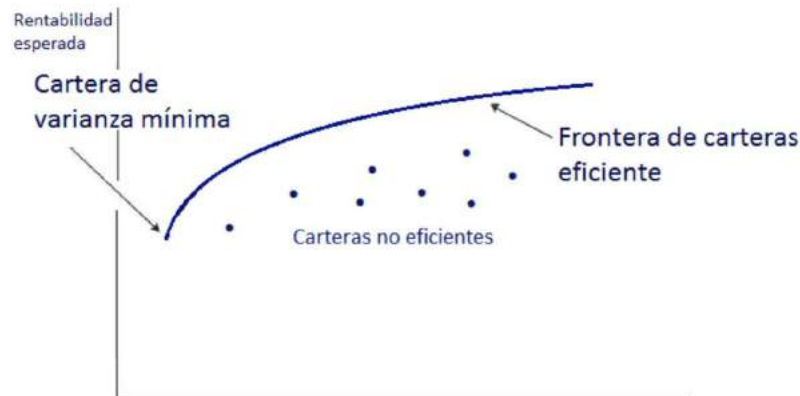


Ilustración 7. Frontera de carteras eficientes. Fuente: (Ferrari, 2016)

2.2.3 PROBLEMAS DEL MODELO DE MARKOWITZ Y PROPUESTA DE SOLUCIÓN.

El elevado número de acciones disponible en los mercados financieros hace difícil, ineficiente y, en ocasiones imposible, el uso del modelo de Markowitz para el cálculo de las proporciones a invertir en cada empresa. (He-Shan Guan, 2007). Por este motivo es útil la agrupación de acciones mediante técnicas de agrupación de datos con *Machine Learning*.

Siguiendo el trabajo de (Gubu, 2019) una posible solución es la creación de grupos de stocks que guardan cierta similitud entre ellos.

Se ilustrará esto con un ejemplo. Si el mercado de valores tuviese 1.000 acciones entre las que invertir siguiendo el modelo de Markowitz para la creación de carteras sería necesario crear un vector de pesos $w = (w_1, w_2, w_3, \dots, w_{1000})$ de dimensión 1.000. En consecuencia el problema de optimización requiere una elevada carga computacional y puede llegar a ser imposible de resolver. Sin embargo, si agrupamos las acciones en 20 grupos de acciones similares y nos quedamos sólo con el mejor valor de cada grupo el problema de optimización se reduce a un vector de pesos $w = (w_1, w_2, w_3, \dots, w_{20})$ de dimensión 20.

Existen distintas propuestas de cómo elegir el mejor valor dentro de cada grupo. Autores como (Gubu, 2019) y (Pan, 2020) usan el ratio de nitidez o *sharpe ratio* para afinar su elección. En este trabajo, se apuesta por el análisis de índice de fuerza relativa (RSI), análisis de medias móviles y análisis de noticias para la selección del mejor valor de cada grupo. Este último tipo de análisis, conocido como new-trading ha sido probado por distintos autores como (Stefan Feuerriegel, July 19, 2018), (Schumaker, 2009), (Montero, 2020), (Mittermayer, 2004) para el estudio de distintos mercados de valores.

Por lo tanto, la metodología a seguir (i) agrupación de datos de empresas, (ii) selección del mejor stock de cada grupo, y (iii) elección de la proporción de la cartera que se le asigna a cada stock seleccionado.

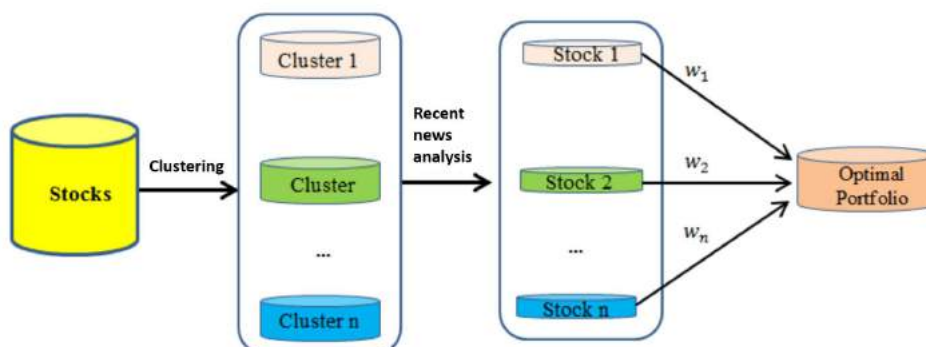


Ilustración 8. Metodología de creación de carteras a partir de agrupación y análisis de noticias. Fuente: (Gubu, 2019) modificado.

2.3 FUNDAMENTOS DE AGRUPACIÓN DE DATOS: CLUSTERING

El *clustering* o agrupación de datos es una técnica de análisis estadístico no supervisado cuyo objetivo es meramente descriptivo: a través de la búsqueda de patrones los datos se organizan en grupos con características similares. Cada instancia puede pertenecer a un solo grupo con el que guarda mayor similitud.

El primer reto a la hora de agrupar individuos por *clusters* es establecer medidas que sirvan para diferenciar observaciones y clasificarlas en grupos. Existen dos tipos de medidas para este propósito:

- (i) Medidas de distancias. La distancia entre dos instancias x_i y x_j se denota como $d(x_i, x_j)$. Una medida de distancia a de ser simétrica $d(x_i, x_j) = d(x_j, x_i)$ y con distancia mínima 0 en caso de misma observación $d(x_i, x_i) = 0$.
- (ii) Medidas de similitud. Alternativa a la distancia que indica la similitud de dos observaciones y se denota $s(x_i, x_j)$. (Rokach, 2014)

Existen distintos métodos para la creación de grupos:

- Métodos jerárquicos:
 - Agrupación divisiva: todas las observaciones se engloban en un solo grupo que se va dividiendo en subgrupos en función de sus distancias o disimilitudes.
 - Agrupación por aglomeración: cada observación representa un *cluster* (grupo). Aquellas observaciones o grupos más próximos o similares entre sí, se van uniendo para formar nuevos grupos en función de la distancia o medida de similitud elegida.

El resultado de estas técnicas es un dendograma, una representación gráfica en forma de árbol donde se representan las agrupaciones/divisiones para todas las observaciones.

- Métodos de partición: estos métodos sirven para clasificar individuos en número de grupos k especificado a priori. El fundamento de este tipo de métodos es realizar particiones sucesivas de los individuos. Los individuos se van reubicando en distintos grupos a medida que se realizan las particiones satisfaciendo dos condiciones:
 - o Cada grupo debe contener al menos un individuo.
 - o Un individuo no puede pertenecer a más de un grupo.

El criterio para establecer buenas particiones consiste en que los individuos de un mismo grupo estén relacionados entre sí y, que objetos de distintos grupos sean diferentes entre sí. El método básico de agrupación por partición es k -medias, el cual se usa en este proyecto. (Romero, 2019)

- Modelos gaussianos: se usan modelos probabilísticos para la agrupación y clasificación de los individuos.

Además, en el caso de acciones de bolsa se pueden agrupar las compañías en función de su actividad natural como hace el CNAE (CNAE Clasificación Nacional de Actividades Económicas, 2009). Este método también ha sido usado por autores como (Cruz, 2019) para realizar grupos de acciones del IBEX 35.

2.3.1 AGRUPACIÓN Y PREDICCIÓN DE SERIES TEMPORALES DE INTERVALO EN LOS MERCADOS FINANCIEROS.

La peculiaridad que tiene la agrupación de acciones de bolsa es que los datos son series temporales de intervalos (STI). En esta sección se explica el concepto de STI, sus características y técnicas para realizar clústeres y predicciones.

“Una serie temporal queda definida como un conjunto de medidas de una misma variables tomadas sobre periodo de tiempo constantes.” (Cruz, 2019) Por otro lado, un intervalo es un conjunto de números reales entre dos valores.

Por lo tanto, una serie temporal de intervalos (STI) es un conjunto de medidas de intervalos tomados a lo largo de un periodo constante de tiempo. Los datos bursátiles comúnmente se

expresan en forma de STI siendo cada intervalo un conjunto de valores entre el mínimo y el máximo valor de la acción del día.

Para la agrupación se calcula la distancia entre series temporales de intervalos (STI) con separación de tiempo uniforme. Para ello se calcula la distancia de cada observación de la serie y luego se engloban en una distancia como se muestra a continuación: (Cruz, 2019)

$$d_t(x, y) = \frac{1}{T} \sum_{i=1}^Y d(x_t, y_t)$$

Distintos autores como (He-Shan Guan, 2007) o (Maté C. , 2012) discuten medidas de distancias aplicadas en series temporales de intervalo.

Una vez ejecutada la distancia para todos los pares de STI's se obtiene la matriz de distancias para los datos estudiados. Con esta matriz y el método deseado se obtienen los grupos como se muestra a continuación.

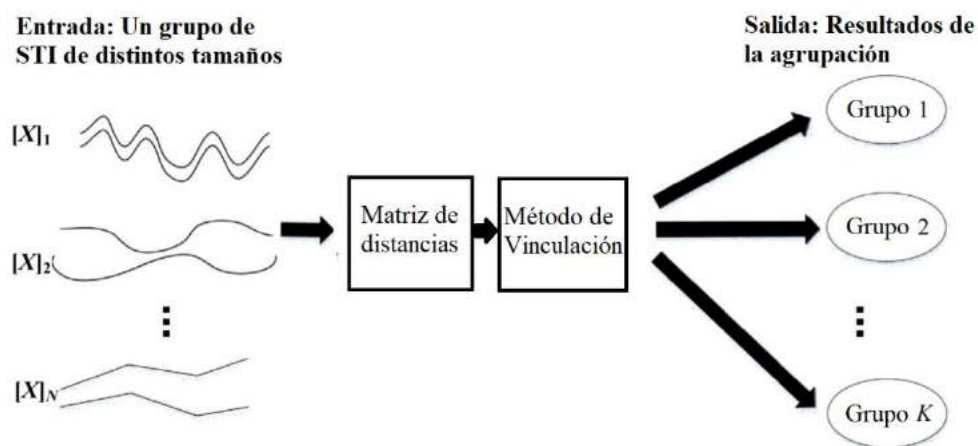


Ilustración 9. Metodología agrupamiento STI. Fuente: (Cruz, 2019).

Por otro lado, para la predicción de STI se sigue a (Maté C. , 2021). En sus estudios se presentan dos métodos de modelado y predicción:

- (i) Métodos puros, que usan la aritmética para el cálculo de nuevas STI's a partir de otras STI's.

- (ii) Pseudo métodos para previsión de intervalos. Este último método se basan en el modelado de elementos de STI como series temporales clásicas y su posterior combinación (Ej. modelado de una serie temporal de centros de los intervalos y una serie temporal de radios y su posterior combinación) (Maté C. , 2021).

2.4 MINERÍA DE DATOS: *WEB SCRAPING*

El siguiente componente clave para la elaboración del trabajo es la extracción de datos de internet o *web scrapping*. Se trata de una serie de técnicas usadas para la extracción de información de internet. El método comúnmente usado para llevar a cabo esta labor es mediante *Hyper-text Transfer Protocol* (HTTP) (Zhao, 2017). Se usa este método para acceder a distintas fuentes de información como (Cinco días, 2021) o (Expansión, 2021).

2.5 SENTIMIENTO DE MERCADO

Otro aspecto clave y de lo más innovador para la elaboración de este trabajo es el análisis de sentimiento de mercado. Este se puede definir como el consenso de la actitud del conjunto de participantes en los mercados financieros que puede, o no, repercutir en la cotización de las empresas. El sentimiento de mercado se detecta generalmente a raíz de la información disponible como pueden ser artículos periodísticos, entrevistas a expertos o *tweets* (mensaje digital a través de la red social Twitter a través del cual un usuario expone su opinión, da información o comparte contenido).

En este apartado se introduce la teoría de mercados eficientes de Eugene Fama, Premio Nobel que establece la relación entre el precio de las acciones y la información disponible. Por otro lado se explican las principales técnicas de análisis de sentimientos.

2.5.1 HIPÓTESIS DE MERCADOS EFICIENTES Y *BEHAVIORAL FINANCE*

Eugene Fama, economista estadounidense, ganó el Premio Nobel de Economía en 2013 por su análisis y relación entre precios de activos e información. Además, es considerado el padre de la Hipótesis de Mercados Eficientes.

En sus estudios Fama define una relación entre información y precio de los activos. Sostiene que un mercado es “informacionalmente eficiente” cuando el precio de los activos recoge la información existente en cada momento: “*Un mercado en el que los precios reflejan la totalidad de la información existente es denominado eficiente*” (Fama, 1970).

Además, un mercado eficiente no puede ser predecible. Fama establece que el precio futuro de una acción es reflejo de la información futura. Por lo tanto, como no se puede predecir la información futura tampoco se puede predecir el mercado de valores. A medida que surge información nueva los actores del mercado actúan racionalmente produciéndose alteraciones en los valores que reflejen la información disponible. Por consiguiente, se sigue la teoría del *Random Walk* o recorrido aleatorio (Malkiel, 1973).

Fama establece tres niveles de eficiencia dentro de un mercado:

- (i) Eficiencia débil. El precio de una acción refleja la información bursátil pasada contenida en las series históricas de los precios de esa acción. Por lo tanto, para obtener rentabilidades extraordinarias, un inversor tiene que analizar información pública (noticias, anuncios...) e información confidencial (datos financieros de la empresas, decisiones del consejo...). El análisis técnico no tiene importancia ya que te solo te da el precio de la acción en ese momento.
- (ii) Eficiencia semi-fuerte. El precio de una acción refleja la información bursátil pasada contenida en las series históricas de los precios de esa acción y la información pública disponible. Por lo tanto, para obtener rentabilidades extraordinarias es necesario el análisis de información confidencial.
- (iii) Eficiencia fuerte: El precio de una acción refleja toda la información existente, tanto las series históricas pasadas, como la información pública como la

confidencial. En este escenario ningún inversor dispone de medio para obtener rentabilidad extraordinaria ya que no tiene ninguna ventaja competitiva respecto al mercado. (Quiroga, 2017)

En contraposición a la Hipótesis de Mercados Eficientes de Fama surgen las finanzas de comportamiento o *behavioral finance*. Esta rama propone una teoría alternativa capaz de explicar sucesos como las burbujas financieras, que no se explican con la teoría de mercados eficientes. Esta teoría admite conductas irracionales de los inversores y las incorpora en los modelos. Es decir, se admiten sesgos cognitivos y emocionales de los inversores que causan anomalías e ineficiencias en los mercados (Ferreruela, 2017).

2.5.2 PROCESADO DE NOTICIAS: ANÁLISIS DE SENTIMIENTOS

(Stefan Feuerriegel, July 19, 2018) En uno de sus artículos dijo: “*Las reacciones del mercado de valores a las divulgaciones financieras están impulsadas por información cualitativa, ya que las noticias suelen ir encarnadas en mensajes de texto.*” Para analizar mensajes de texto y obtener información valiosa para la toma de decisiones de inversión se hace uso de técnica de Procesamiento Natural de Lenguaje (PLN). Dentro de este campo es especialmente útil el análisis de sentimientos (AS), con el cual se busca obtener un valor numérico que indique el tono de un texto dentro de una escala previamente definida.

Por lo general, el AS se puede usar para (i) evaluar un texto (negativo/neutro/positivo), (ii) para establecer la intensidad del sentimiento expresado (mediante una escala que refleje el grado de negatividad o positividad) y (iii) para clasificar el texto en grupos según una lista de sentimiento expresados (ej. felicidad, enfado, angustia...) o temáticas (ej. textos de carácter religioso, financiero, médico...).

Las principales técnicas usadas para la extracción de sentimientos son: (i) lexicón, (ii) aprendizaje supervisado y (iii) métodos híbridos.

- (i) Lexicón o enfoque basado en léxicos. Se trata de un análisis no supervisado¹ basado en el uso de diccionarios (lexicón) para el cálculo del valor sentimental del texto. Estos diccionarios están formadas por palabras a las cuales se les asigna un valor sentimental. El método consiste en la búsqueda de estas palabras a lo largo del texto a analizar y la contabilización del sentimiento que tienen pre asignado cada una de las mismas.

A continuación se muestra un ejemplo en el que el lexicón está formado por 5 palabras (deliberado, divertido, encantar, odio y reflexivo), a cada una de estas se le asigna un valor (0, +1 ó -1). En el texto a analizar se localizan las palabras del lexicón y se realiza un sumatorio de su polaridad. El resultado es el sentimiento del texto. (Albertos, 2018)

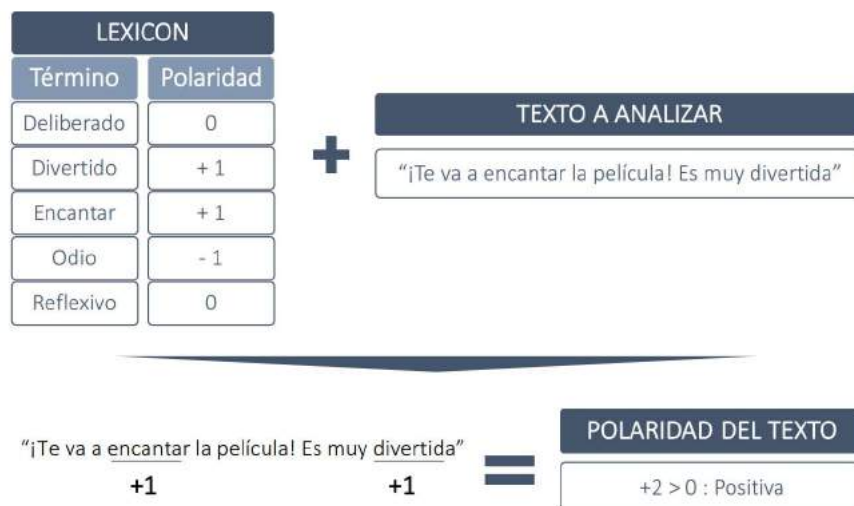


Ilustración 10. Funcionamiento de sistema de análisis de texto mediante Lexicón

Fuente: (Albertos, 2018)

El principal problema de esta técnica es la escasez de diccionarios para clasificación de palabras por sentimientos. Durante la realización de este trabajo solo se ha tenido constancia de la existencia de dos diccionarios de sentimiento

¹ Modelos en los que no disponemos de datos clasificados previamente. Este tipo de análisis se centra en la búsqueda de patrones dentro de nuestros datos y en entender la estructura de la información que se analiza. Los métodos más comunes son el *clustering* y la reducción de dimensiones o PCA.

en lengua castellana. En primero de todos es un diccionario genérico disponible en LIWC (LIWC, 2015). En segundo lugar, (Bernal, 2020) dispone del “*primer Lexicón de sentimiento financiero en Castellano*” según el mismo comenta en su trabajo. No obstante, otros autores como (Montero, 2020) han realizado uso de análisis de sentimiento mediante lexicón haciendo uso de un reducido número de palabras.

- (ii) Enfoque basado en aprendizaje automático (*machine learning*). Se trata de un análisis supervisado² en el que se dispone de un conjunto de noticias previamente etiquetado usado para el entrenamiento de un algoritmo (Fase 1: Entrenamiento). Este algoritmo es usado para clasificar un nuevo conjunto de noticias asignándoles una etiqueta (Fase 2: Predicción). (Cunningham) (Albertos, 2018)

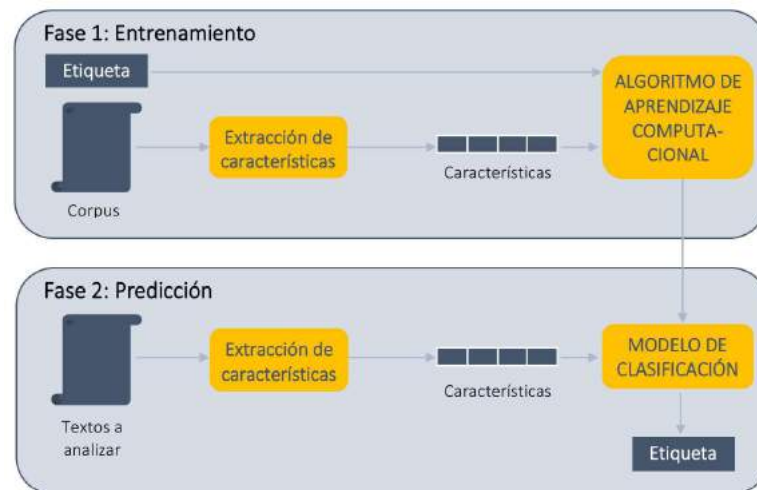


Ilustración 11. Funcionamiento de sistema de análisis de texto mediante aprendizaje automático. Fuente: (Albertos, 2018)

Este tipo de técnicas ha sido usada por autores como (Schumaker, 2009) para el análisis de noticias o (Baviera, 2016) para el análisis de sentimientos en mensajes de Twitter. El principal problema de esta técnica es la necesidad de un conjunto

² Modelos en los que disponemos de datos clasificados previamente. Por lo tanto, en este tipo de análisis se busca la relación entre variables de entrada y variables de salida. Los métodos más comunes son la clasificación, regresión y predicción.

de entrenamiento amplio previamente etiquetado. En el caso del análisis de noticias el etiquetado se puede hacer por dos métodos. Siguiendo a (Nuij) se puede optar por el etiquetado de noticias mediante el análisis de las mismas por expertos. En sus estudios un grupo de analistas categorizó en base a sus conocimientos financieros un conjunto de noticias con el objetivo de comprobar la correlación que existía con los movimientos del mercado bursátil. Por otro lado, (Mittermayer, 2004) clasifica las noticias como positivas o negativas en función de la variación de la cotización en bolsa a los 60 minutos de su publicación.

- (iii) Por último, existen sistemas híbridos que hacen uso del Lexicón y de algoritmos de aprendizaje automático para el análisis de noticias.

2.6 ENFOQUE COMBINADO Y ESTRATEGIAS DE TRADING

Según (Stefan Feuerriegel, July 19, 2018) un sistema de trading basado en noticias consta de dos pasos: (i) extracción del valor sentimental de una noticia que es un indicador del retorno esperado y (ii) ejecución de estrategias de trading para toma de decisiones.

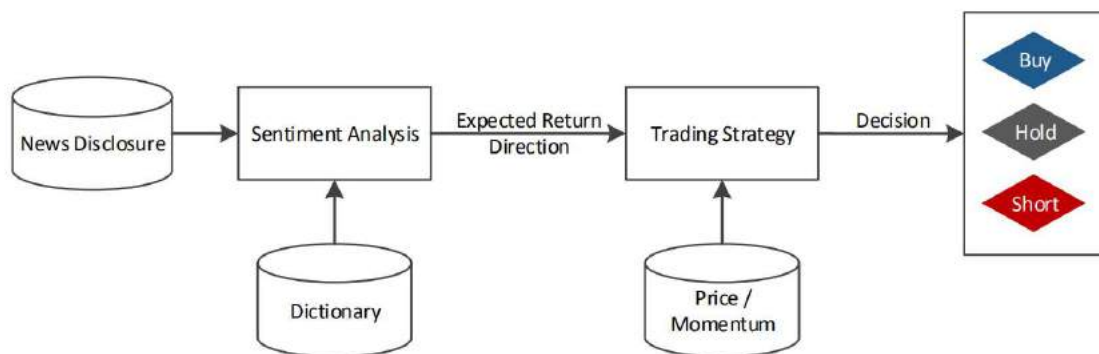


Ilustración 12. Ejemplo sistema de trading basado en noticias. Fuente: (Stefan Feuerriegel, July 19, 2018)

Las principales estrategias son (i) trading de momento (*momentum trading*), (ii) trading basado en reglas y, (iii) aprendizaje automático (*machine learning*).

Como explica (Takeuchi, 2013) el *momentum trading* establece que stocks que han tenido retornos elevados durante los últimos 3 a 12 meses son susceptibles de seguir teniendo esos

retornos y viceversa. La teoría es que si hay suficiente fuerza impulsando el precio en una cierta dirección entonces ese movimiento es probable que continúe por algún tiempo (Markets, 2020).

Autores como (Nuij) han estudiado la relación entre el impacto medido por expertos que puede causar una noticia y el *rate-of-change* que se produce en los *stocks* correspondientes concluyendo que existe una relación lineal entre ambas variables en la mayoría de los casos.

Otros autores como (Stefan Feuerriegel, July 19, 2018) proponen estrategias basadas en reglas (*rule-based trading*) que sirven de gatillo para lanzar órdenes de compra o venta cuando se detecta un nivel de sentimiento determinado. Además, (Stefan Feuerriegel, July 19, 2018) y (Nuij) combinan esta estrategia con estrategias de *momentum trading* en sus estudios mediante árboles de decisión y operadores lógicos.

Estrategias de machine learning junto con análisis de texto también han sido utilizadas por autores como (Schumaker, 2009) para predecir las tendencias en los precios de las acciones y clasificarlas. A su vez, (Schumaker, 2009) indica que los principales algoritmos usados por otros autores para el análisis predictivo de stock basándose en análisis de textos son (i) algoritmos genéticos, (ii) Naïve Bayes y (iii) SVM (Support Vector Machines).

2.7 SISTEMAS DE INVERSIÓN EN ACCIONES Y CREACIÓN DE CARTERAS

A partir de los años noventa una serie de avances tecnológicos relacionados con sistemas de la información y comunicación propiciaron el crecimiento de las operaciones en los mercados financieros. El aumento de capacidad de los ordenadores y la conexión a internet permitió que cientos de usuarios pudieran operar cómodamente. Además, esto permitió que los *traders* eligiesen su bróker.

A lo largo de los últimos años, nuevos avances tecnológicos han producido un incremento aún mayor de la popularidad de la inversión online y, en concreto, de la inversión en

acciones. De esta forma se ha acuñado el término *trading online* que se puede definir como el método de compra o venta de acciones en el mercado financiero a través de aplicaciones conectadas que facilitan el aumento del valor de las transacciones reduciendo los costes de las mismas. Esta nueva modalidad proporciona útiles ventajas a sus usuarios como:

- El acceso a herramientas e indicadores para el análisis bursátil ayudando a los usuarios en la toma de decisiones.
- Automatización de procesos posibilitando acciones instantáneas.
- Elementos semi automatizado para su uso por expertos.
- Reducción del coste medio por transacción.

A continuación explicaremos dos sistemas de trading y algunas de sus funcionalidades.

2.7.1 SIMPLY WALL STREET: CONFIDENT INVESTMENT DECISIONS

Simply Wall Street es una startup australiana fundada en junio de 2014 que ayuda a las personas a invertir gracias a herramientas visuales que convierten datos financieros complejos en gráficos comprensibles. El origen de Simple Wall Street surgió a causa del aumento de inversores de los últimos años, siendo la característica de estos que toman sus decisiones basándose únicamente en noticias. Debido a estos los rendimientos esperados son menores.

Por lo tanto, el objetivo de la Simply Wall Street es ayudar a inversores tanto experimentados como noveles a desarrollar estrategias de inversión rentables en el largo plazo a través de la ayuda de gráficos para la toma de decisiones. Este sistema no hace recomendaciones sobre posibles inversiones ni toma decisiones sino que mediante sus herramientas aporta consistencia a nuestras propias decisiones y busca mantenernos emocionalmente desligados de ellas.

Los análisis disponibles en este sistema se nutren de información financiera de más de más de 150.000 compañías y diversos análisis externos para completar información referente a valoraciones y otros aspectos. Las herramientas que ofrece son:

- Informes sobre compañías. El equipo de Simply Wall Street analiza empresas cada 24 horas creando reportes detallados para sus usuarios con información útil para la toma de decisiones. Estos informes incluyen:
 - o Resumen ejecutivo. Se trata de un *checklist* de recompensas y riesgos que presenta la compañía. Este *checklist* incluye elementos como cantidad de deuda de la compañía o sostenibilidad de los dividendos. Mediante este análisis se evalúa el perfil del tipo de compañía y se obtiene un gráfico con cinco ejes: dividendos, valor, futuro, pasado y salud que se muestra a continuación.

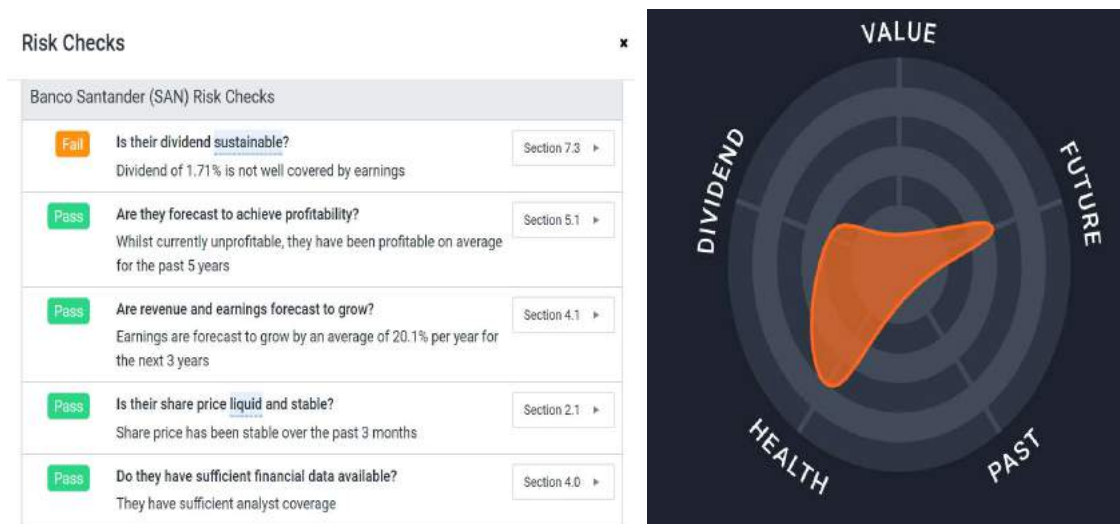


Ilustración 13. Checklist Banco Santander y perfil de empresas según Simply Wall Street. Fuente: Simply Wall Street.

- o Por otro lado, se muestra la cotización de las acciones de la empresa y se relaciona con los distintos eventos ocurridos en el pasado. Los eventos se dividen en 5 categorías: (i) financieros (morado), (ii) gestión y dirección (verde claro), (iii) dividendos (verde oscuro), (iv) estratégicos (naranja) y (v) otros (azul).



Ilustración 14. Cotización Banco Santander junto con eventos clave. Fuente: Simply Wall Street.

Pinchando en los puntos de colores se puede acceder a la información detallada de cada evento abriéndose el desplegable que se muestra a continuación.

1 Event - Tue, 17 Nov 2020

● Banco Santander Appoints Paul Horlock as Chief Payment Officer

Banco Santander hired Paul Horlock from Pay.UK as its chief payments officer. He is to assume his duties by the beginning of next year and will be responsible for creating a single UK Payments team as part of One Santander. Horlock is currently chief executive of Pay.UK, which he joined in October 2017.

Ilustración 15. Evento importante Banco Santander. Fuente: Simply Wall Street.

- Valoraciones. Simple Wall Street proporciona valoraciones de las compañías usando distintos métodos. Hace uso de descuento de flujo de caja para obtener el valor real de las acciones de cada empresa y compararlo con su valor actual. Para este cálculo hace uso de información de analistas externos o extrapola el flujo de caja como predicciones basándose en datos pasados. También se indican valores de ratios como PE ratio o PB ratio para la valoración de la compañía.

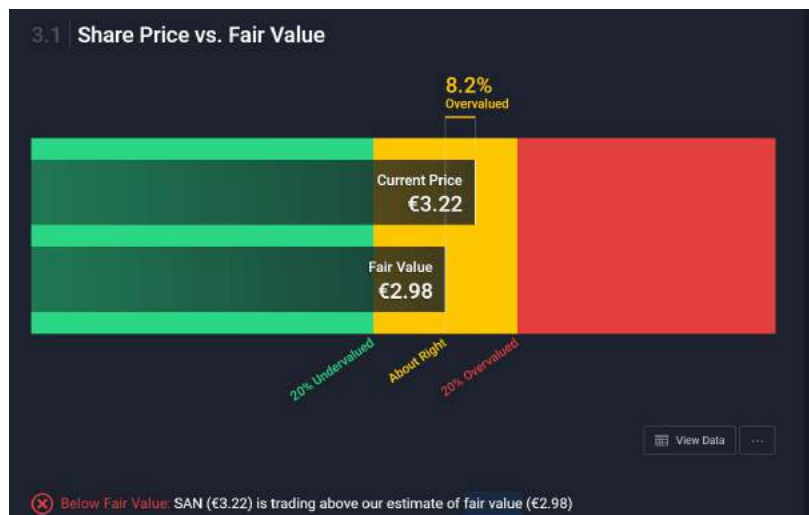


Ilustración 16. Valoración precio acciones Banco Santander. Fuente: Simply Wall Street.

- Predicciones futuras de crecimiento y datos del rendimiento pasado. Mediante distintos gráficos Simply Wall Street proporciona predicciones sobre ganancias, flujos de caja, ganancias por acción y otros indicadores. Para ello hace uso de la información de distintos analistas externos. Además, para periodos en el pasado muestra la comparativa entre las predicciones de estos analistas y los resultados reales.
- Datos financieros y dividendos. Muestra un análisis de la posición financiera de las empresas, su balance y distintas comparativas sobre los dividendos repartidos por la empresa.

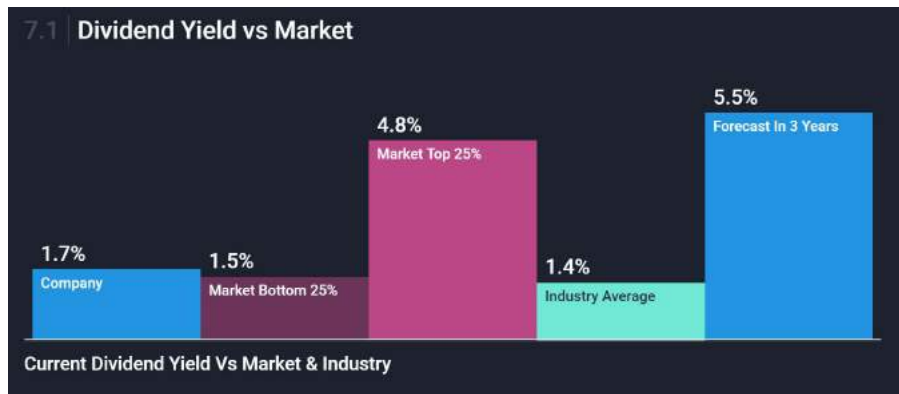


Ilustración 17. Datos financieros Banco Santander. Fuente: Simply Wall Street.

- Datos sobre gestión, localización.
- Monitorización. Por otro lado, Simply Wall Street permite la creación de listas de seguimiento para la monitorización de una serie de activos. Con ellas el usuario se mantiene informado de las principales actualizaciones que impactan a las acciones deseadas. La información se recibe siguiendo el siguiente formato.

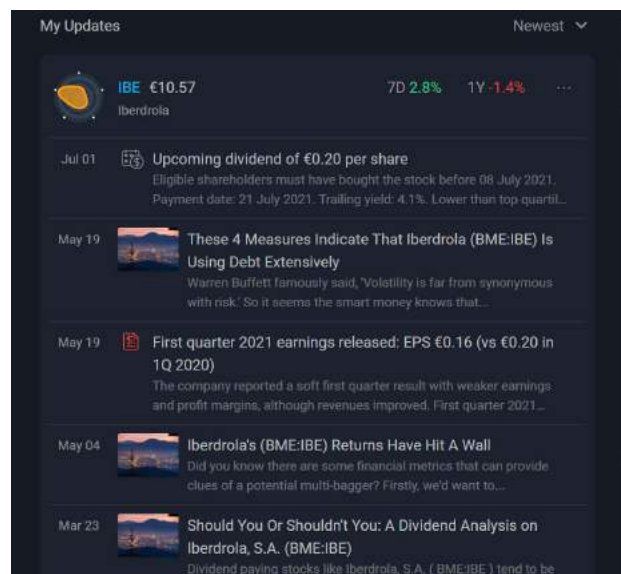


Ilustración 18. Actualizaciones monitorización Iberdrola. Fuente: Simply Wall Street.

- Creación de carteras. Por último, es posible crear carteras indicando el nombre de la empresa y la cantidad o precio de las acciones a comprar. A continuación se muestra un ejemplo de la creación de una cartera para cuatro acciones pertenecientes al IBEX 35:

Total	Total Bought	Total shares	Current Price	Current Value	Return			Total Gain (inc Currency)	Avg Years	Total Return
					Capital Gains	Dividends				
BME:SAN		52.00	€3.16	€164.42	€51.05	€1.43	€52.48	1.00	46.29%	
BME:IBE		50.00	€10.57	€528.50	€-13.25	€16.20	€2.95	1.00	0.55%	
BME:ITX		50.00	€29.70	€1,485.00	€292.50	€28.35	€320.85	1.00	26.91%	
BME:VIS		50.00	€59.25	€2,962.50	€47.50	€68.44	€115.94	1.00	3.98%	

Ilustración 19. Creación de una cartera con Simply Wall Street. Fuente: Simply Wall Street.

Una vez creada la cartera se generan informes similares al reporte de compañías explicado anteriormente. Alguno de los gráficos más innovadores que proporciona Simply Wall Street son el perfil de la cartera y la diversificación en industrias. El perfil de la cartera se crea como una combinación de los perfiles de las acciones seleccionadas (explicado anteriormente) ponderando sus pesos. Por otro lado, también se informa al usuario sobre la diversificación en sectores en función de sus empresas.

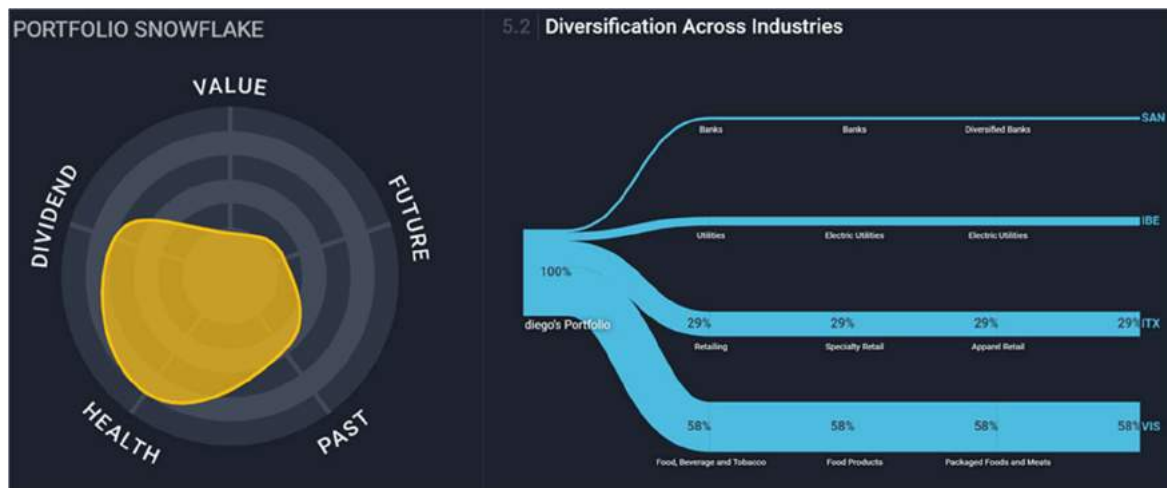


Ilustración 20. Informe carteras Simply Wall Street. Fuente: Simply Wall Street.

Se incluye también un resumen financiero con datos como los dividendos esperados de la cartera correspondiente, la deuda y el *equity* conjunto de tu cartera, valoración de la cartera usando los métodos explicados anteriormente o un análisis de la volatilidad comparándola con la del mercado.



Ilustración 21. Gráficos resumen financiero carteras Simply Wall Street. Fuente: Simply Wall Street.

2.7.2 WEBULL

Webull Financial LLC es una app de trading vía aplicación o página web con cero comisiones y sin depósito mínimo para empezar a invertir. Fue creada en 2017 por Wang Anquan, ex empleado del grupo Alibaba. Esta plataforma ofrece una serie de herramientas de análisis de compañías para mejorar los retornos de sus usuarios. Además, la aplicación pone a disposición variedad de productos para la creación de carteras diversificadas. Entre estos productos se pueden encontrar acciones, opciones, ETFs y ADRs. Actualmente Webull no opera en mercados europeos pero si en Estados Unidos, Canada, Hong Kong y recientemente en el mercado de Criptomonedas.

Una de las herramientas más novedosa que disponen Webull es la compra de fracciones de acciones de empresas (con un mínimo de cinco euros por fracción) aumentando así las posibilidades para la diversificación de carteras de menor capital.

La interfaz de Webull se divide en distintas ventanas cada una con distintas funciones que se explicarán a continuación.

- Stocks. Es la ventana principal que permite analizar la cotización de empresas. Además de información básica bursátil pone a disposición de los usuarios una serie de herramientas de análisis. A continuación se explican las opciones más populares:
 - o Indicadores y otras herramientas para análisis técnico. Es uno de los puntos fuertes de la aplicación. Dispone de más de 50 indicadores permitiendo su personalización. Entre los indicadores hay bandas de Boiller, bandas de Donchian, medias móviles simples, medias móviles exponenciales, niveles de Fibonacci, nubes de Ichimoku y MACD.



Ilustración 22. Cotización Banco Santander junto con medias móviles de 5, 10, 20, 100 y 200 periodos e indicador MACD. Fuente: elaboración propia a partir de (Anquan, n.d.)

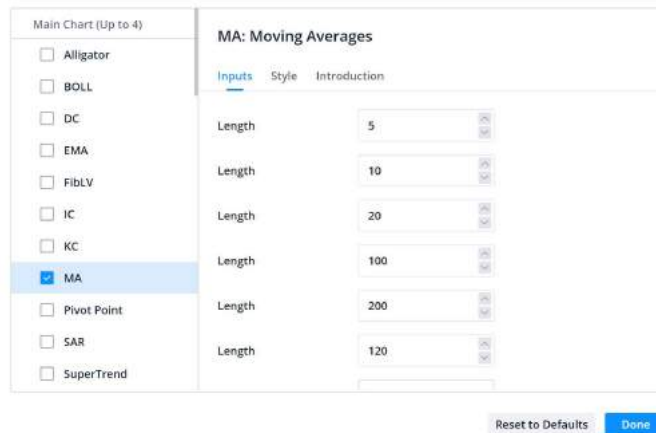


Ilustración 23. Personalización indicadores técnicos aplicación Webull. Fuente: elaboración propia a partir de (Anquan, n.d.)

- Información financiera. Se incluye información financiera referente al balance de situación, cuenta de resultados y flujos de caja. Además, se proporciona una serie de predicciones de los resultados financieros para los próximos cuatrimestres de la empresa. Esta información permite la toma de decisiones para inversores que buscan rentabilidades en el largo plazo.

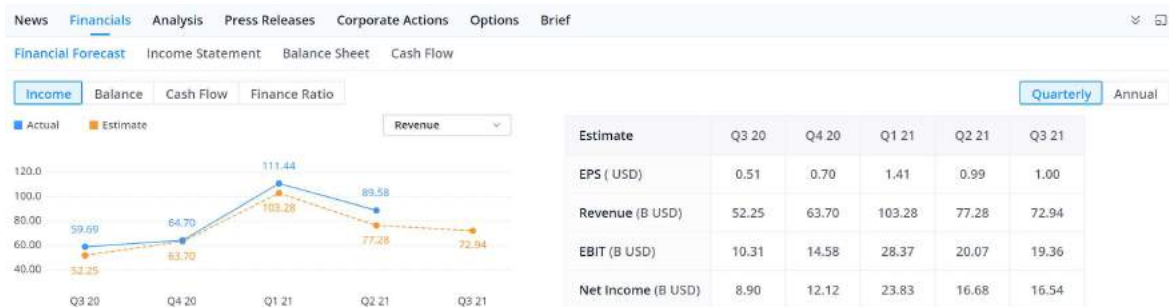


Ilustración 24. Ejemplo información financiera disponible en Webull. Fuente: elaboración propia a partir de (Anquan, n.d.)

- Análisis de expertos. Webull tiene una serie de analistas externos que incluyen sus informes a la aplicación. En estos informes se da la opinión sobre si aguantar una posición, comprar, vender...



Ilustración 25. Ejemplo análisis de expertos Webull. Fuente: elaboración propia a partir de (Anquan, n.d.)

- Noticias. Se incluyen los titulares de las últimas noticias referentes a la compañía analizada junto con un link que permite acceder a la web origen de la noticia.



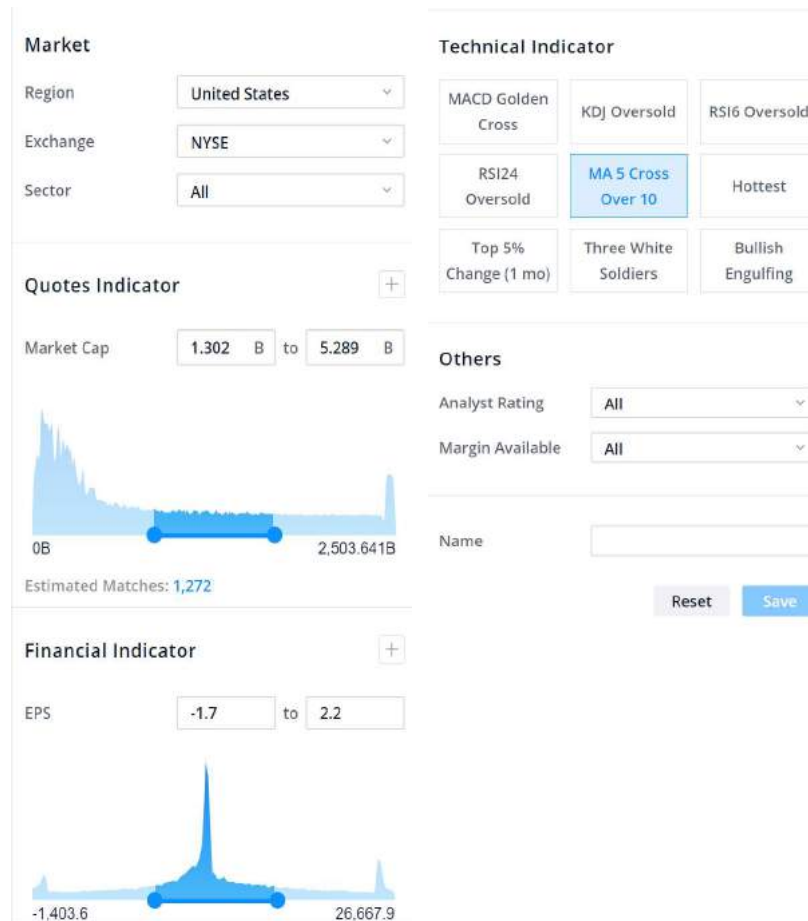
The screenshot shows a news feed with the following items:

News	Financials	Analysis	Press Releases	Corporate Actions	Options	Brief
Banco Santander to buy primary dealer of USTs for \$600M, expanding CIB business						Seekingalpha - 2d ago
UPDATE 2-Santander to buy U.S. fixed-income broker Amherst Pierpont for \$600 mln						reuters.com - 2d ago
Santander U.S. holding to buy broker Amherst Pierpont for \$600 mln						reuters.com - 2d ago
BRIEF-Santander To Buy U.S. Fixed-Income Broker Dealer Amherst Pierpont						reuters.com - 2d ago
Sector Update: Financial						MT Newswire - 2d ago
Deals of the day-Mergers and acquisitions						reuters.com - 2d ago
Banco Santander-Chile (BSAC) Is Oversold: Can It Recover?						Zacks - 2d ago
Retired detective lost life savings in a 'clone firm' investment scam						The Guardian - 2d ago
Top Ranked Income Stocks to Buy for July 13th						Zacks - 2d ago
Banco Santander Sells Nonperforming Loan Portfolio to Marathon -Source						marketwatch.com - 2d ago

Ilustración 26. Ejemplo pestaña noticias en análisis de Stocks en Webull. Fuente: elaboración propia a partir de (Anquan, n.d.)

- Listas de seguimiento. En esta sección se permite al usuario la creación de listas de seguimiento de acciones. El objetivo es la monitorización de las empresas deseadas. En este apartado se incluye la misma información que en el apartado *Stocks* pero con la posibilidad de crear alarmas y notificaciones. Estas notificaciones se pueden personalizar e incluyen entre otros la posibilidad de programar alarmas en caso de cruce medias móviles con el precio, porcentaje de subidas, alcanzar máximos o mínimos locales...
- Ventana de Mercados. Dispone de mercados en Estados Unidos, Canadá, Hong Kong y recientemente ha incluido el mercado de Criptomonedas.
- Venta de simulación. Webull pone a disposición un simulador con el que permite invertir de forma virtual y poner a prueba las decisiones de inversión.

- Menu Screener. Esta herramienta sirve para la búsqueda de empresas que cumplen con una serie de características. Webull pone a la disposición una serie de filtros avanzados como búsqueda por región, mercado, sector, ganancias por acción, índice de fuerza relativa o cruce de medias móviles entre otros.



The screenshot shows the Webull Screener interface with the following sections:

- Market:** Region (United States), Exchange (NYSE), Sector (All).
- Technical Indicator:** A grid of indicators including MACD Golden Cross, KDJ Oversold, RSI6 Oversold, RSI24 Oversold, MA 5 Cross Over 10 (highlighted), Hottest, Top 5% Change (1 mo), Three White Soldiers, and Bullish Engulfing.
- Quotes Indicator:** Market Cap filter set to 1.302 B to 5.289 B, with a corresponding area chart and 'Estimated Matches: 1,272'.
- Financial Indicator:** EPS filter set to -1.7 to 2.2, with a corresponding area chart.
- Others:** Analyst Rating (All) and Margin Available (All).
- Name:** A search input field.
- Buttons:** Reset and Save.

↓ **Resultado búsqueda**

No.	Symbol	Name	Price	After Hours	Change	% Change	Open	Prev Close
1	FLO	Flowers Foods	24.34	0.00% (24.34)	+0.15	+0.62%	24.19	24.19
2	WOW	WideOpenWest	21.56	0.00% (21.56)	+0.07	+0.33%	21.64	21.49

Ilustración 27. Búsqueda avanzada de Stock usando el menú Screener de Webull. Fuente: elaboración propia a partir de (Anquan, n.d.)

Capítulo 3. SISTEMA DESARROLLADO

(Schumaker, 2009), (Mittermayer, 2004) afirman que el efecto de una noticia en bolsa se equilibra entre 20-60 minutos después de la publicación de la misma. Por este motivo, con la cantidad de información disponible para tener cierta ventaja competitiva respecto al resto de inversores es imprescindible un sistema automatizado, rápido y eficiente para la medición del efecto de noticias en los mercados financieros. Además, con el desarrollo de nuevas tecnologías y el aumento de la capacidad computacional es probable que la brecha de oportunidad de entre 20-60 minutos continúe reduciéndose.

En este capítulo se propone un sistema semiautomático que suponga una herramienta útil para la creación de carteras y análisis de estrategias de inversión basándose en el análisis de noticias y análisis técnico. En el primer apartado 3.1 se explicará la arquitectura del sistema. En el apartado 3.2 se introducen las entradas del sistema. En el apartado 3.3 se exponen los métodos y bases de datos usadas en los distintos motores del sistema. Los motores del sistema se explican en los apartados 3.4 Agrupación de datos, 3.5 Análisis de sentimiento, 3.6 Clasificación y 3.7 Trading. Por último, en el apartado 3.8 se presenta la salida del sistema junto con una serie de visualizaciones.

3.1 ARQUITECTURA DEL SISTEMA DESARROLLADO.

En esta sección se describe la arquitectura del sistema de trading. Este sistema está diseñado para:

1. Extraer información bursátil de los *tickers* seleccionados en tiempo real y de forma automática y creación de una base de datos local.
2. Extracción de titulares de noticias relacionadas con las compañías objetivo de los principales periódicos económicos en España y creación de una base de datos local.

3. Creación de grupos de las compañías objetivo en base a los datos bursátiles históricos extraídos en el punto 1.
4. Análisis del sentimiento de las noticias de la base de datos local y su clasificación en función de la dirección esperada de retorno que provocan.
5. Creación de recomendaciones de carteras e inversión.

El sistema dispone de cuatro motores para realizar las principales funciones: (i) motor de agrupación de compañías (*clustering*), (ii) motor de procesado de noticias y análisis de sentimiento, (iii) motor de clasificación y (iv) motor de *trading*. Además, el sistema necesita conexión a internet para el acceso a hemerotecas de distintos periódicos y datos de cotización de las empresas objetivo. A partir de estos se crean dos bases de datos locales a partir de las que se nutren los motores del sistema.

Una serie de entradas hacen posible el cumplimiento de las funciones del sistema. Estas entradas son: (i) los *tickers* de las empresas objetivo, (ii) lexicón de las empresas objetivo, (iii) diccionarios de sentimiento financiero y (iv) análisis de noticias por expertos.

El resultado del sistema será una recomendación de compra de una serie de compañías para la creación de una cartera.

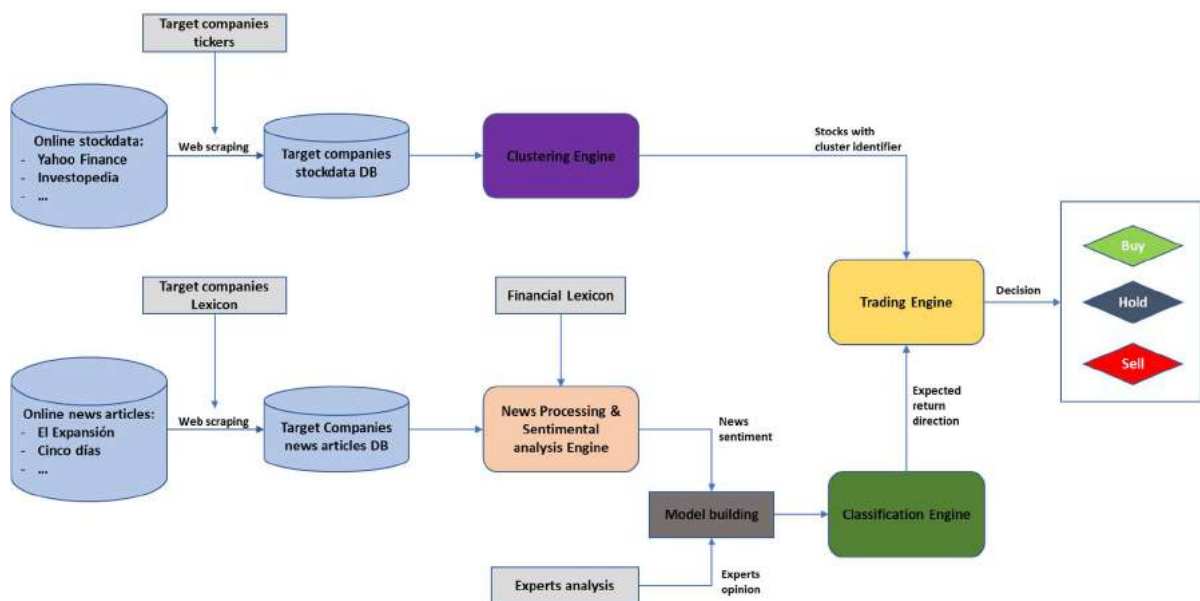


Ilustración 28. Arquitectura del sistema desarrollado. Fuente: elaboración propia.

3.2 ENTRADAS AL SISTEMA

A continuación se detallan las entradas necesarias para el sistema y su formato.

3.2.1 TICKERS Y LEXICÓN COMPAÑÍAS

La primera entrada necesaria para el sistema es la lista de compañías target a analizar. Esta lista se proporciona al sistema en formato Excel con el nombre *0_Target_companies.xlsx*. El formato que ha de seguir se muestra en la Tabla 2. Formato entradas al sistema Tickers y Lexicón compañías objetivo. Fuente: elaboración propia.

- Columna A. Nombre completo de la compañía.
- Columna B. El *ticker* que usa *Yahoo Finance* para identificar la compañía. El programa usa esta información para extraer los datos bursátiles de la compañía por lo que es muy importante poner el *ticker* correcto.
- Columna C. Lexicón de la compañía. Este campo se usa para buscar titulares de noticias en los periódicos que tengan este término. Por lo tanto, es importante acertar con el lexicón para una buena extracción de noticias y análisis de sentimiento.

	A	B	C
1	Company	Stock Ticker	Lexicon
2	Mapfre, S.A.	MAP.MC	Mapfre
3	Banco Santander, S.A.	SAN.MC	Santander
4	Banco Bilbao Vizcaya Argentaria, S.A.	BBVA.MC	BBVA
5	Amadeus IT Group, S.A.	AMS.MC	Amadeus
6	Naturgy Energy Group, S.A.	NTGY.MC	Naturgy
7	Siemens Gamesa Renewable Energy, S.A.	SGRE.MC	Gamesa
8	Inmobiliaria Colonial, SOCIMI, S.A.	COL.MC	Colonial

Tabla 2. Formato entradas al sistema Tickers y Lexicón compañías objetivo. Fuente: elaboración propia.

En caso de querer analizar otras compañías habría que actualizar este Excel con los datos correspondientes. Además, el número de empresas a incluir puede ser mayor, el inconveniente es que esto eleva la carga computacional del sistema.

3.2.2 LEXICÓN FINANCIERO

Esta entrada es el diccionario a usar para el análisis de sentimiento. El formato del diccionario es un .txt con un listado de palabras clasificadas según distintas categorías. En este trabajo se lleva a cabo una clasificación binaria de palabras que aportan una connotación negativa y palabras que aportan connotación positiva (ej. La palabra “bajadas” tendrá asignado el valor 1 para el campo de sentimiento negativo y el valor 0 para el campo de sentimiento positivo).

El lexicón financiero es uno de los grandes retos de este trabajo, supone uno de los cuellos de botella para el análisis de sentimiento y componente a tener en cuenta en futuros desarrollos. Dado lo novedoso que es el análisis de sentimientos solo hay constancia (hasta donde se conoce) de un diccionario de sentimiento financieros en castellano publicado en (Bernal, 2020). El propio Ángel Bernal dice: *“se ha creado, hasta donde conocemos, el primer diccionario en español de palabras con connotación positiva, negativa o neutra dentro del contexto de la estabilidad financiera”*. Otro autores, como (Albertos, 2018) en sus estudios se encontraron con el mismo cuello de botella dada la falta de investigación hasta el momento.

Se lleva a cabo un análisis de sentimiento con dos lexicón distintos:

- Lexicón genérico en castellano proporcionado por LIWC. (LIWC, 2015)
- Primer lexicón financiero en Español creado por (Bernal, 2020).

Además, se han probado otros lexicón como el usado por autores como (Montero, 2020). No obstante, el reducido número de términos hizo imposible su uso en este trabajo.

El primer paso para evaluar la calidad de cada diccionario es comparar el número de palabras que detecta cada uno de un conjunto de titulares de noticias. Para este experimento se analiza un conjunto de 1531 titulares de noticias. Se calcula el porcentaje medio de palabras que el diccionario reconoce y se incluye en la Tabla 3. Tabla 3. Análisis diccionarios en castellano. Fuente: elaboración propia.

Diccionario	% de palabras detectadas	Nº palabras registradas
<i>LIWC en castellano</i>	64,5%	12.625 términos más sus posibles derivaciones
<i>(Bernal, 2020)</i>	0,6%	564 palabras

Tabla 3. Análisis diccionarios en castellano. Fuente: elaboración propia.

A raíz de este experimento se concluye que el diccionario más completo es el de LIWC en castellano. No obstante, aún no se puede confirmar que sea el más apto para este trabajo ya que se trata de un diccionario de uso genérico que incluye pronombres y sustantivos mientras que el diccionario de Bernal está compuesto por términos positivos y negativos en el ámbito financiero.

3.2.3 ANÁLISIS SUPERVISADO DE NOTICIAS POR EXPERTOS

En este trabajo se ha optado por incluir estrategias híbridas para el análisis de sentimientos. Para el lexicón se hace uso de los diccionarios explicados en el apartado anterior y para el desarrollo de aprendizaje automático supervisado es necesario etiquetar las noticias entre aquellas con sentimiento positivo y negativo.

Con las hipótesis de (Mittermayer, 2004) y (Schumaker, 2009) el efecto de una noticia en el mercado bursátil tarda entre 20 y 60 minutos diluirse. Por lo tanto, si queremos etiquetar noticias en función de su efecto bursátil necesitaríamos recoger datos al minuto. Debido las limitaciones de carga computacional se ha descartado esta opción y se ha optado por el análisis manual de un conjunto de noticias para el entrenamiento del algoritmo de clasificación. Por lo tanto, y siguiendo a (Nuij) se propone analizar una muestra de noticias por un grupo de personas cualificadas. Para llevar a cabo esta tarea se va a seguir la metodología Delphi. Se profundizará en esta metodología y su implementación en el

apartado 3.6.2. Además, en el Anexo 1 se incluye un formulario para su análisis por un grupo de expertos.

3.3 EXTRACCIÓN DE INFORMACIÓN DE INTERNET Y CREACIÓN DE BASES DE DATOS: *WEB SCRAPING*

Se ha realizado labores de extracción de información de internet para la obtención de noticias de distintos periódicos financieros y datos bursátiles de Yahoo Finance.

3.3.1 EXTRACCIÓN DE NOTICIAS

La obtención de noticias para realizar el análisis de sentimiento se ha realizado extrayendo información en tiempo real de los periódicos Cinco Días y Expansión.

Para la extracción de información se ha usado la función *webread* de Matlab 2021 cuya función es la lectura de contenido de una página web especificada mediante su URL. Para acceder a información de días pasados se han detectado patrones en los URL para la automatización del proceso de búsqueda.

El siguiente paso es la búsqueda de patrones en el documento para la extracción de los titulares de noticias. Para ello se ha usado la función *extractBetween* como se muestra en la imagen siguiente.

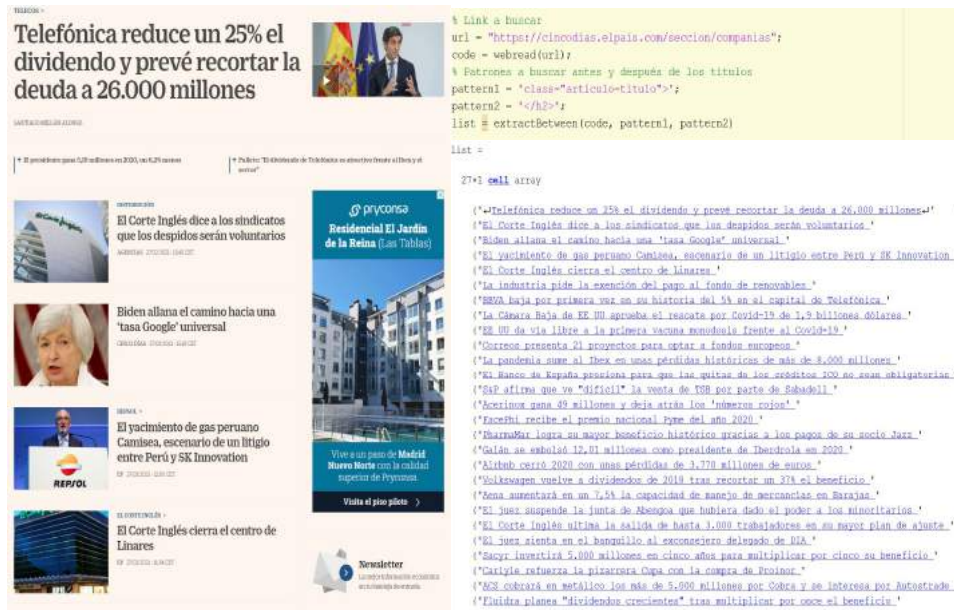


Ilustración 29. Extracción de noticias de Cinco Días. Fuente: elaboración propia a partir de (Cinco días, 2021)

Una vez obtenidos los titulares estos se procesan buscando nombres o referencias a las empresas objetivo a analizar. Los titulares seleccionados se almacenan en un Excel que es la base de datos de la que se nutre el motor de análisis de sentimiento. El formato final de la base de datos proporciona información de: titulares, empresa objetivo mencionada, fecha titular, fuente del titular, número de palabras.

	A	B	C	D	E
1	Noticias	Empresa	Fecha	Fuente	NumPalabras
2	Repsol cumple 2 años en el negocio eléctrico con 1,1 millones de clientes y 2 GW de proyectos nuevos 'verdes'	Repsol	2020/11/01	Expansión	20
3	Amancio Ortega cobra mañana 646,8 millones de euros del dividendo de Inditex	Inditex	2020/11/01	Expansión	12
4	Los analistas mejoran las estimaciones para Fluidra tras presentar sus resultados hasta septiembre	Fluidra	2020/11/01	Expansión	13
5	ACS cancela un contrato de derivados y lo sustituye por uno 'forward' de 12 millones de acciones	ACS	2020/11/02	Expansión	17
6	Telefónica y Vértice 360 se alían para la digitalización y gestión de catálogos audiovisuales	Telefónica	2020/11/02	Expansión	14
7	PNM Resources (Iberdrola) acelera su salida del carbón en 2024, siete años antes de lo previsto	Iberdrola	2020/11/02	Expansión	16
8	Alantra crea una titulizadora europea y ficha en Santander a su responsable	Santander	2020/11/03	Expansión	12
9	Juan Abelló triplica su inversión en Merlin en el tercer trimestre	Merlin	2020/11/03	Expansión	11
10	Santander arranca las conversaciones con los sindicatos previas al ERE	Santander	2020/11/03	Expansión	10
11	Santander y Norges entran en Soltec, estrella bursátil de las renovables en España	Santander	2020/11/03	Expansión	13
12	Alantra lanza una titulizadora liderada por un ex-Santander	Santander	2020/11/04	Expansión	8
13	Cellnex pierde 84 millones hasta septiembre por los costes de crecimiento	Cellnex	2020/11/04	Expansión	11
14	Cellnex confirma "conversaciones avanzadas" con Hutchison para comprar sus torres en Europa valoradas en 10.000 millones	Cellnex	2020/11/04	Expansión	16
15	La compra de las torres de Hutchison aprovecharía al máximo la capacidad financiera actual de Cellnex	Cellnex	2020/11/04	Expansión	16
16	Endesa gana 1.511 millones tras estabilizar el impacto del Covid	Endesa	2020/11/04	Expansión	10
17	Santander Totta ganó 255 millones hasta septiembre, un 35 % menos	Santander	2020/11/04	Expansión	11
18	Cómo gestionan Mapfre y Catalana Occidente el Covid fuera de España	Mapfre	2020/11/04	Expansión	11
19	BBVA lanza con Amazon una plataforma en la nube para operar en Bolsa	BBVA	2020/11/05	Expansión	13
20	Cómo le va a Sabadell tras vender su gestora?	Sabadell	2020/11/05	Expansión	9
21	Santander: nuevos recortes de plantilla en Europa	Santander	2020/11/05	Expansión	7
22	Santander remodela la cúpula de Consumer para la fusión con Openbank	Santander	2020/11/05	Expansión	11

Tabla 4. Base de datos titulares de noticias. Fuente: elaboración propia.

3.3.2 DATOS BURSÁTILES

La obtención de datos bursátiles se realiza a través de *Yahoo Finance*. El programa escanea los *tickers* objetivo que son una entrada del sistema y los busca en la web para a continuación guardar los datos. El código se ha realizado con ayuda del código de Josiah Renfree disponible en *File Exchange* de *MathWorks* (Renfree, 2021). Se ha modificado el código para adaptar el formato de los datos a tablas y poder exportar a Excel para crear una base de datos local. Además, se ha incorporado el cálculo de la tasa de retorno diaria tanto aritmética como logarítmica (columnas *DailyReturn* y *Log_DailyReturn*) y una variable *SignodailyReturn* que indica el signo de la tasa de retorno diaria (1: tasa de retorno positiva en el día, -1: tasa de retorno negativa en el día). Se profundiza en la utilidad de esta variable en el apartado de agrupación de datos.

La información extraída se guarda en una base datos local en formato Excel, como el mostrado en la Tabla 5. Formato extracción datos bursátiles. Fuente: elaboración propia.

	A	B	C	D	E	F	G	H	I	J
1	Date	Open	High	Low	Close	AdjClose	Volume	DailyReturn	Log_DailyReturn	SignodailyReturn
2	02/11/2020 00:00	1,309	1,343	1,292	1,322	1,284024	7795320	0,009833585	0,009882255	1
3	03/11/2020 00:00	1,321	1,382	1,318	1,375	1,335502	8486550	0,039272727	0,040064706	1
4	04/11/2020 00:00	1,345	1,375	1,294	1,351	1,312191	8395882	0,004441155	0,004451046	1
5	05/11/2020 00:00	1,359	1,39	1,34	1,381	1,34133	13498832	0,015930485	0,016058739	1
6	06/11/2020 00:00	1,375	1,399	1,36	1,364	1,324818	8954403	-0,008064516	-0,008032172	-1
7	09/11/2020 00:00	1,349	1,533	1,342	1,528	1,484107	13588634	0,117146597	0,124596114	1
8	10/11/2020 00:00	1,511	1,619	1,509	1,606	1,559866	15521627	0,059153176	0,060974932	1
9	11/11/2020 00:00	1,61	1,626	1,552	1,564	1,519073	10331645	-0,029411765	-0,028987537	-1
10	12/11/2020 00:00	1,556	1,588	1,53	1,578	1,532671	7972756	0,013941698	0,014039797	1
11	13/11/2020 00:00	1,57	1,6	1,56	1,578	1,532671	6840073	0,005069708	0,005082603	1
12	16/11/2020 00:00	1,604	1,639	1,579	1,614	1,567636	9679997	0,006195787	0,00621506	1

Tabla 5. Formato extracción datos bursátiles. Fuente: elaboración propia.

3.4 MOTOR DE AGRUPACIÓN DE DATOS

Para facilitar la creación de carteras se agrupan compañías siguiendo tres métodos distintos dando lugar a distintas carteras: (i) agrupación natural, (ii) agrupación con método jerárquico y (iii) agrupación mediante k-medias con variables rentabilidad-volatilidad. A su vez, el

agrupamiento mediante k-medias se puede desarrollar de dos formas distintas en función de las preferencias del usuario que se explicaran más adelante.

3.4.1 AGRUPACIÓN NATURAL

Una de las opciones a la hora de diversificar una cartera es diversificar a través de distintos sectores y actividades. De esta forma se intenta replicar la economía global o del país y no un nicho de empresas centradas en un tipo de actividad.

Esta agrupación se realiza siguiendo las pautas de la Bolsa de Madrid donde se clasifican por sectores las empresas públicas. (Madrid, n.d.)

A continuación se muestra la clasificación sectorial de las empresas que forman parte del IBEX 35 durante el periodo en el que se realizó el trabajo.

1. Petróleo y Energía	2. Mat. Básicos, Industria y construcción	3. Bienes de Consumo	4. Servicios de Consumo
ENG.MC	ANA.MC	ALM.MC	AENA.MC
ELE.MC	ACX.MC	GRF.MC	IAG.MC
IBE.MC	ACS.MC	ITX.MC	MEL.MC
NTGY.MC	MTS.MC	PHM.MC	
REE.MC	CIE.MC	VIS.MC	
REP.MC	FER.MC		
SLR.MC	FDR.MC		
	SGRE.MC		
5. Servicios Financieros	6. Tecnología y Telecomunicaciones	7. Servicios Inmobiliarios	
BBVA.MC	AMS.MC	COL.MC	
SAB.MC	CLNX.MC	MRL.MC	
SAN.MC	IDR.MC		
BKT.MC	TEF.MC		
CABK.MC			
MAP.MC			

Tabla 6. Agrupación natural empresas IBEX 35. Fuente: elaboración propia.

3.4.2 AGRUPACIÓN MÉTODO JERÁRQUICO

El segundo método de agrupación del sistema se basa en los estudios de (He-Shan Guan, 2007) con alguna variante. Se trata de agrupación jerárquica a partir de una matriz de distancias entre empresas que se explica a continuación.

Los pasos para el cálculo de la distancia entre dos empresas se calcula como:

Paso 1: dada dos acciones X e Y se calcula su tasa de retorno diaria a lo largo del periodo a estudiar $i = 1, 2, 3, \dots, n$.

Paso 2: evaluación del signo de las tasas de retorno diarias de las empresas X e Y.

$$C_{X,i} = \begin{cases} 1 & \text{si } r_{X,i} > 0 \\ 0 & \text{si } r_{X,i} = 0 \\ -1 & \text{si } r_{X,i} < 0 \end{cases} \quad C_{Y,i} = \begin{cases} 1 & \text{si } r_{Y,i} > 0 \\ 0 & \text{si } r_{Y,i} = 0 \\ -1 & \text{si } r_{Y,i} < 0 \end{cases}$$

Paso 3: evaluación de similitudes:

$$\beta = \begin{cases} 0 & \text{si } \text{signo}(r_{X,i}) = \text{signo}(r_{Y,i}) \\ 1 & \text{si } \text{signo}(r_{X,i}) \neq \text{signo}(r_{Y,i}) \end{cases}$$

Paso 4: cálculo de la distancia entre la cotización de X e Y.

$$\text{dist}(X, Y) = \sum_{i=1}^n \beta_i$$

Tabla 7. Algoritmo para el cálculo de distancia entre activos. Fuente: elaboración propia a partir de (He-Shan Guan, 2007).

Esta medida se centra en evaluar similitudes en el signo de las rentabilidades diarias de dos acciones. Dos empresas serán muy similares (la distancia será pequeña) si sus rentabilidades coinciden en signo los mismos días.

Esto resulta extremadamente útil ya que se tiene en cuenta posibles comportamientos cíclicos en rentabilidades de grupos de empresas. Es común que una serie de empresas compartan signo de rentabilidad durante periodos de tiempo.

El problema que tiene este método es que no tiene en cuenta el riesgo o volatilidad de las empresas estudiadas.

3.4.3 AGRUPAMIENTO CON K-MEDIAS

Por último, con el objetivo de aportar de variedad al usuario, se aplican técnicas de agrupamiento mediante K-medias. La función K-medias no admite como entrada datos en forma de series temporales de intervalo. Por lo tanto, es necesario manipular los datos para obtener una entrada al sistema en la que se le pueda asignar a cada acción un valor de cada variable para el periodo completo.

La primera aproximación se llevara a cabo mediante K-medias de dos variables absolutas en el periodo de estudio elegido por el usuario. Para ser coherente con la teoría de creación de cartera de Markowitz explicada en el apartado 2.2.2 se ha elegido rentabilidad y riesgo como variables de entrada. Este tipo de agrupamiento ha sido probado con anterioridad por autores como (Cruz, 2019).

La rentabilidad absoluta se define como la rentabilidad del periodo de estudio usando los datos de cierre según la siguiente formula:

$$\text{Rentabilidad absoluta} = \frac{x_{\text{último cierre periodo}} - x_{\text{primer cierre periodo}}}{x_{\text{primer cierre periodo}}}$$

Por otro lado, se modela el riesgo como la volatilidad absoluta del periodo de estudio. A su vez, la volatilidad se define como la diferencia entre el máximo de los máximos valores de cierre menos el mínimo de los mínimos valores de cierre. A continuación se muestra la formula usada.

$$\text{Volatilidad absoluta} = \frac{\max(x_{\text{maximo diario}}) - \min(x_{\text{minimo diario}})}{\min(x_{\text{minimo diario}})}$$

El problema que presenta esta metodología de agrupamiento de datos es la dependencia respecto al periodo de tiempo elegido. Este problema es causado fundamentalmente por la variable rentabilidad. En caso de que una empresa suba o baje su valor muy rápidamente es

posible que el periodo de tiempo escogido corte en mitad de dicha subida y por lo tanto el resultado no represente la realidad del rendimiento de la empresa.

Por este motivo se desarrolla la idea de (Cruz, 2019) usando como variables la rentabilidad media semanal y volatilidad media semanal. De esta forma se tiene información más precisa sobre las acciones obteniéndose un resultado más robusto erradicando el problema mencionado anteriormente.

3.5 MOTOR DE ANÁLISIS DE SENTIMIENTO DE NOTICIAS

El sentimiento de las noticias se analiza con ayuda del software de LIWC (LIWC, 2015) y los distintos diccionarios explicados anteriormente. LIWC es un sistema diseñado por James W. Pennebaker para el análisis del lenguaje capaz de categorizar textos en más 70 categorías a través del uso del lexicón (LIWC, 2015). A pesar de la variedad de categorías y ser el lexicón más extenso disponible hasta la fecha en castellano (hasta dónde llega el conocimiento del autor de este trabajo) se trata de un lexicón genérico por lo que el encaje puede no ser el mejor para este trabajo. Por este motivo se evalúa también el uso del diccionario de (Bernal, 2020).

Se realiza el análisis de titulares de noticias extraídas entre el 01/11/2020 y el 05/03/2021 de los periódicos Expansión (Expansión, 2021) y Cinco días (Cinco días, 2021).

Los indicadores usados para el análisis de sentimiento son:

$$EmoPos = \frac{Num\ palabras\ positivas}{Num.\ total\ palabras} \times 100$$

$$EmoNeg = \frac{Num\ palabras\ negativas}{Num.\ total\ palabras} \times 100$$

A continuación se muestra el resultado de este bloque del sistema usando el diccionario genérico de LIWC 2008:

Source (A)	Empresa	Fecha	Source (D)	Num. Palabras	EmoPos	EmoNeg
Repsol cumple 2 años en el negocio eléctrico con 1,1 millones de clientes y 2 GW de proyectos nuevos 'verdes'	Repsol	2020/11/01	Expansión	20	0,00	0,00
Amancio Ortega cobra mañana 646,8 millones de euros del dividendo de Inditex	Inditex	2020/11/01	Expansión	12	0,00	0,00
Los analistas mejoran las estimaciones para Fluidra tras presentar sus resultados hasta septiembre	Fluidra	2020/11/01	Expansión	13	7,69	0,00
PNM Resources (Iberdrola) acelera su salida del carbón en 2024, siete años antes de lo previsto	Iberdrola	2020/11/02	Expansión	16	0,00	0,00
Alantra crea una tituladora europea y ficha en Santander a su responsable	Santander	2020/11/03	Expansión	12	8,33	0,00
Juan Abelló triplica su inversión en Merlin en el tercer trimestre	Merlin	2020/11/03	Expansión	11	0,00	0,00
Santander arranca las conversaciones con los sindicatos previas al ERE	Santander	2020/11/03	Expansión	10	10,00	0,00
Santander y Norges entran en Soltec, estrella bursátil de las renovables en España	Santander	2020/11/03	Expansión	13	7,69	0,00
Cellnex confirma "conversaciones avanzadas" con Hutchison para comprar sus torres en Europa valoradas en 10.000 millones	Cellnex	2020/11/04	Expansión	16	6,25	0,00
La compra de las torres de Hutchison aprovecharía al máximo la capacidad financiera actual de Cellnex	Cellnex	2020/11/04	Expansión	16	6,25	0,00
Endesa gana 1.511 millones tras estabilizar el impacto del Covid	Endesa	2020/11/04	Expansión	10	0,00	0,00
Santander Totta ganó 255 millones hasta septiembre, un 35 % menos	Santander	2020/11/04	Expansión	11	20,00	0,00
Cómo gestionan Mapfre y Catalana Occidente el Covid fuera de España	Mapfre	2020/11/04	Expansión	11	0,00	0,00
BBVA lanza con Amazon una plataforma en la nube para operar en Bolsa	BBVA	2020/11/05	Expansión	13	0,00	0,00
Cómo le va a Sabadell tras vender su gestora?	Sabadell	2020/11/05	Expansión	9	0,00	0,00
Santander: nuevos recortes de plantilla en Europa	Santander	2020/11/05	Expansión	7	14,29	0,00
Santander remodela la cúpula de Consumer para la fusión con Openbank	Santander	2020/11/05	Expansión	11	9,09	0,00
Siemens Gamesa tiene pérdidas récord de 918 millones en su año más "complejo"	Gamesa	2020/11/05	Expansión	13	0,00	7,69
ArcelorMittal triplica pérdidas	rcelorMitt	2020/11/05	Expansión	3	0,00	33,33
Megaplan de Iberdrola con 75.000 millones de inversión y 19.000 millones de dividendos hasta 2025	Iberdrola	2020/11/05	Expansión	15	0,00	0,00

Tabla 8. Resultado del análisis de sentimiento mediante Lexicón con LIWC. Fuente: elaboración propia.

3.6 MOTOR DE CLASIFICACIÓN

Una vez disponemos de un nivel de sentimiento (salida boque Sentimiento explicado en 3.5) se busca conocer la dirección esperada del movimiento en bolsa. Como indica (Schumaker, 2009) el análisis de noticias solo puede ayudar a predecir la dirección de movimiento. Por lo tanto, el objetivo de este bloque es la clasificación de las noticias en tres categorías (C_i):

- Noticia positiva: implica el aumento del valor de la acción, $C_i = 1$.
- Noticia neutra: indiferencia del precio de la acción, se traduce como un $C_i = 0$.
- Noticia negativa: disminución del valor de la acción, se traduce como un $C_i = -1$.

Con este propósito se propone un análisis no supervisado mediante reglas y, adicionalmente, se crea una estructura y metodología a desarrollar en futuros trabajos para análisis supervisado basado en la opinión de expertos. Debido a la falta de expertos para el análisis de noticias y restricción de tiempo, este análisis y esta metodología no llego a implementarse. No obstante, se desarrolla la idea de cara a desarrollos futuros.

3.6.1 APRENDIZAJE NO SUPERVISADO:

Para la clasificación supervisada de noticias se crea la positividad neta que se muestra a continuación. Es un indicar que contabiliza en términos absolutos (tiene en cuenta todas las palabras y resta la negatividad) la positividad o negatividad del texto. Por lo tanto, un texto con positividad neta superior a 0 es, teóricamente, una noticia positiva y viceversa.

$$Positividad\ neta = \frac{Num\ palabras\ positivas - Num\ palabras\ negativas}{Num.\ total\ palabras} \times 100$$

Que se puede traducir en:

$$Positividad\ neta = EmoPos - EmoNeg$$

El cálculo de positividad neta se realiza por días, es decir, en caso de que exista más de una noticia para una misma empresa en un mismo día, se realizara el cálculo de la positividad neta de cada noticia y a continuación se calculará la media de la positividad neta de las noticias obteniéndose así la positividad neta de cada empresa para cada día.

A continuación se muestra un ejemplo de las noticias de Aena. A la izquierda de la imagen está el resultado de análisis de sentimiento de cada noticia y a la derecha el análisis de sentimiento del día para la empresa como resultado de la combinación de noticias.

empresa							Noticias_agrupadas_dia				
41x5 table							34x5 table				
	1	2	3	4	5	6	1	2	3	4	5
	Var1	Var2	Var3	Var4	Var5		Var1	GroupCount	mean_Var3	mean_Var4	mean_Var5
1	11-11-2020'Aena'		0	0	0		04-11-2020	1	0	7.1400	-7.1400
2	13-11-2020'Aena'		0	0	0		11-11-2020	1	0	0	0
3	18-11-2020'Aena'		0	0	0		13-11-2020	1	0	0	0
4	21-11-2020'Aena'		0	0	0		18-11-2020	1	0	0	0
5	25-11-2020'Aena'		0	0	0		21-11-2020	1	0	0	0
6	25-11-2020'Aena'		6.6700	0	6.6700		25-11-2020	4	6.8025	0	6.8025
7	25-11-2020'Aena'		14.2900	0	14.2900		03-12-2020	1	7.6900	0	7.6900
8	03-12-2020'Aena'		7.6900	0	7.6900		10-12-2020	1	0	0	0
9	18-12-2020'Aena'		0	0	0		14-12-2020	1	0	7.1400	-7.1400
10	21-12-2020'Aena'		8.3300	8.3300	0		18-12-2020	1	0	0	0
11	22-12-2020'Aena'		0	0	0		21-12-2020	1	8.3300	8.3300	0
12	15-01-2021'Aena'		0	0	0		22-12-2020	1	0	0	0
13	18-01-2021'Aena'		0	0	0		24-12-2020	1	0	0	0
14	22-01-2021'Aena'		6.6700	0	6.6700		13-01-2021	1	0	5.8800	-5.8800
15	25-01-2021'Aena'		0	0	0		15-01-2021	2	0	0	0
16	29-01-2021'Aena'		0	0	0		17-01-2021	1	0	0	0

Tabla 9. Agrupación noticias Aena por día. Fuente: elaboración propia.

Por último, se aplican las siguientes reglas (*rule-based trading*) para la obtención de la tendencia de cada día en cada empresa:

- Si *Positividad neta de* $N_i > 0,1$ entonces $C_i = 1$.
- Si *Positividad neta de* $0 < N_i < 0,1$ entonces $C_i = 0$.
- Si *Positividad neta de* $N_i < 0$ entonces $C_i = -1$.

Los intervalos creados no son simétricos ya que se realizaron una serie de pruebas y se observó un sesgo de positividad en los resultados de sentimiento de titulares de noticias. Se permite un margen de error para la positividad neta de entre 0 y 0,1 entre el cual las tendencia del día y empresa se considerada neutra. Con ello se pretende eliminar ruido generado por (i) defectos en las técnicas de análisis de sentimiento y ausencia de diccionarios extensos y completos en español y, (ii) posibles noticias con leve connotación positiva o negativa pero no suficiente como para causar impacto significativo en el precio de la acción y (iii) eliminar posibles sesgos en los titulares de periódicos cuando tienden a ser alarmistas.

3.6.2 APRENDIZAJE SUPERVISADO

Para la creación de un sistema de clasificación supervisado es necesario tener un conjunto de datos previamente etiquetados. Para ello se propone la clasificación manual de titulares de un conjunto de noticias publicadas entre el 01/11/2020 y el 01/04/2021.

Para asegurar máxima fiabilidad y validez en los resultados del estudio se ha realizado una adaptación de la metodología Delphi. Se trata de una metodología en la que se hacen consultas a un grupo de personas (considerados expertos) con el objetivo de encontrar la opinión más consensuada sobre el tema a tratar (Liaño, 2012). Por lo tanto, para llevar a cabo esta metodología es importante hacer un seguimiento cercano del grupo de participantes en el experimento para asegurarse de la calidad de los datos.

Los pasos a seguir para la obtención de resultados adaptando la metodología Delphi son:

1. Extracción de una muestra de titulares noticias 01/11/2020 y el 01/04/2021.

2. Creación de un formulario en línea (se usó un documento Excel compartido en OneDrive con contraseña solo disponible para él envió al grupo de expertos (formulario disponible en el Anexo 1).
3. Contacto con expertos y confirmación de su participación en el estudio.
4. Asignación aleatoria de titulares a expertos.
5. Recogida de resultados.
6. Verificación de resultados.

Una vez disponemos de un conjunto de datos se puede llevar a cabo el aprendizaje supervisado. Con un conjunto de entrenamiento se procede a entrenar distintos algoritmos de clasificación y a continuación se llevaran a cabos test para analizar el rendimiento de los algoritmos. Para algoritmos de clasificación revisar la bibliografía expuesta en el apartado Procesado de noticias: Análisis de sentimientos.

3.7 MOTOR DE TRADING

El motor de trading se encarga de elaborar las decisiones de inversión usando como entradas los datos procesados en el resto de bloques. Con el objetivo de crear un sistema más robusto y poder comparar resultados el bloque tiene incorporada tres estrategias de para creación de carteras:

- Selección de carteras basadas en análisis técnico mediante índice de fuerza relativa (RSI).
- Selección de carteras basadas en análisis combinado de índice de fuerza relativa (RSI) y medias móviles.
- Selección de carteras basadas en análisis de sentimiento de noticias.

3.7.1 SELECCIÓN DE CARTERAS BASADAS EN ANÁLISIS DE NOTICIAS

En este caso se trata de predecir la dirección esperada las noticias analizando los catorce días previos a la realización de la inversión (resultado bloque Clasificación explicado en 3.7).

A continuación se resumen los pasos para la selección de la cartera en función del sentimiento de las noticias:

Paso 1: cálculo de la clasificación diaria (C) para cada empresa en función de las noticias en función de la positividad neta (apartado 3.7.1):

- Si Positividad neta de $N_i > 0,1$ entonces $C_i = 1$.
- Si Positividad neta de $0 < N_i < 0,1$ entonces $C_i = 0$.
- Si Positividad neta de $N_i < 0$ entonces $C_i = -1$.

Paso 2: cálculo de la inercia de los último 14 días.

$$\text{Inercia noticias} = \frac{\sum C_i * 100}{14}$$

Paso 3: selección acciones con mayor inercia de cada grupo.

Tabla 10. Algoritmo selección cartera basándose en noticias. Fuente: elaboración propia

3.7.2 SELECCIÓN DE CARTERAS BASADAS EN ANÁLISIS MEDIANTE RSI

En segundo lugar se desarrolla un algoritmo aplicando teoría de análisis técnico mediante RSI explicado en el apartado 2.1.1. Para ello se calcula el RSI durante todo el periodo de estudio para cada una de las compañías objetivo. Estos datos servirán para analizar el comportamiento de las compañías y mostrar distintos gráficos explicados en el apartado 3.8. Por otro lado, para la selección de carteras se usa el RSI el día deseado (día de creación de la cartera). Se eligen las compañías con menor RSI de cada grupo y estas serán las que conformen la cartera.

A continuación se resume el algoritmo que sigue el programa:

Paso 1: cálculo del RSI a 14 días en el periodo de estudio.

$$RSI = 100 - \frac{100}{1 + RS}$$

siendo RS:

$$RS = \frac{\text{promedio subidas}}{\text{promedio bajadas}}$$

Paso 2: idealmente selección de inversiones siguiendo las siguientes reglas:

- Si $RSI > 70$ entonces colocar orden de venta.
- Si $RSI < 30$ entonces colocar orden de compra.

Paso 3: en caso de no haber acciones que cumplan estas condiciones se selecciona las acciones con menor inercia de cada grupo.

Tabla 11. Algoritmo selección cartera basándose en RSI. Fuente: elaboración propia.

Una vez ejecutado el algoritmo el resultado es una tabla con la información de la compañía, su RSI y su RSI binario (decisión de inversión resultado de la ejecución del paso 2 de la Tabla 11. Algoritmo selección cartera basándose en RSI. Fuente: elaboración propia. Este resultado se exporta a un archivo tipo .txt para guardar y mantener un seguimiento de las inversiones. El formato del resultado se muestra en la siguiente imagen:

1 Hoja	2 Empresa	3 RSI	4 RSI_binario
1	"Acciona, S.A."	79.8851	-1'
2	"Acerinox, S.A."	72.4537	-1'
3	"ACS, Actividades de Construcción y Servicios, S.A."	62.6316	0'
4	"Aena S.M.E., S.A."	36.7521	0'
5	"Almirall, S.A."	65.2174	0'
6	"Amadeus IT Group, S.A."	40.8517	0'
7	"ArcelorMittal"	69.4818	0'
8	"Banco Bilbao Vizcaya Argentaria, S.A."	41.1681	0'
9	"Banco de Sabadell, S.A."	51.1712	0'
10	"Banco Santander, S.A."	48.0587	0'
11	"Bankinter, S.A."	60.5144	0'
12	"CaixaBank, S.A."	63.1653	0'
13	"Cellnex Telecom, S.A."	54.8982	0'
14	"CIE Automotive, S.A."	43.3962	0'
15	"Enagás, S.A."	75.2577	-1'
16	"Endesa, Sociedad Anonima"	69.5067	0'
17	"Ferrovial, S.A."	52.5714	0'
18	"Fluidra, S.A."	53.0973	0'
19	"Grifols, S.A."	76.0218	-1'
20	"Iberdrola, S.A."	53.7671	0'
21	"Indra Sistemas, S.A."	46.6165	0'
22	"Industria de Diseño Textil, S.A."	28.5068	1'
23	"Inmobiliaria Colonial, SOCIMI, S.A."	52.1236	0'
24	"International Consolidated Airlines Group, S.A."	40.9769	0'
25	"Mapfre, S.A."	69.4524	0'
26	"Meliá Hotels International, S.A."	32.9231	0'
27	"MERLIN Properties SOCIMI, S.A."	42.8571	0'
28	"Naturgy Energy Group, S.A."	40.7767	0'
29	"Pharma Mar, S.A."	37.0370	0'
30	"Red Eléctrica Corporación, S.A."	68.9944	0'
31	"Repsol, S.A."	44.9735	0'
32	"Siemens Gamesa Renewable Energy, S.A."	64.1182	0'
33	"Solaria Energía y Medio Ambiente, S.A."	41.1960	0'
34	"Telefonica, S.A."	38.2263	0'
35	"Viscofan, S.A."	53.2110	0'

Tabla 12. Asignación RSI y RSI binario a empresas objetivo. Fuente: elaboración propia.

Por último, el resultado de la cartera de inversión se guarda en una variable adicional en forma de tabla como se muestra a continuación. Este resultado también se exporta a .txt para facilitar la visualización y él envió de información entre usuarios.

1 Hoja	2 Empresa	3 Cluster	4 RSI	5 RSI_binario
29	"Pharma M...	1	37.0370	0'
8	"Banco Bilb...	2	41.1681	0'
26	"Meliá Hot...	3	32.9231	0'
22	"Industria d...	4	28.5068	1'

Tabla 13. Cartera resultado del análisis basado en RSI. Fuente: elaboración propia.

3.7.3 SELECCIÓN DE CARTERAS BASADO EN ANÁLISIS DE RSI Y MEDIAS MÓVILES

Por último, se desarrolla un sistema de selección de carteras combinando estrategias de análisis de índice de fuerza relativa y medias móviles. Para ello se usa el RSI del apartado anterior, se analiza la tendencia con la diferencia entre periodo de cierre y media móvil y se comprueba no estar en un cruce de la muerte.

A continuación se resume el algoritmo desarrollado:

Paso 1: selección de la media móvil simple con la que analizar la tendencia.

Paso 2: cálculo de la diferencia entre precio cierre y media móvil simple:

$$diff = \text{precio cierre periodo} - \text{media movil simple periodo}$$

Paso 3: asignación de tendencia.

- Si $diff > 0$ entonces tendencia = 1 indicando tendencia ascendente.
- Si $diff < 0$ entonces tendencia = 0 indicando tendencia descendente.

Paso 4: análisis de la tendencia de la última semana de cada compañía (5 días laborables) y cálculo de la tendencia media de la semana:

$$\text{tendencia semanal} = \frac{\sum_i^5 \text{tendencia}_i}{5}$$

Paso 5: comprobación cruce de la muerte.

$$\begin{cases} \text{Si hay cruce de la muerte bandera} = 1 \\ \text{No cruce de la muerte bandera} = 0 \end{cases}$$

Paso 6: selección compañías con menor RSI con una tendencia alcista durante la última semana y que no se encuentren en cruce de la muerte. Se busca satisfacer:

$$\begin{cases} \min(RSI_{\text{compañía}}) \\ \text{tendencia semanal} > 0,6 \\ \text{bandera cruce de la muerte} = 0 \end{cases}$$

Tabla 14. Algoritmo selección cartera basándose en tendencias mediante media móvil simple. Fuente: elaboración propia.

En caso de no existir acciones con tendencia semanal $> 0,6$ se procede a coger aquellas que maximicen las tendencias minimizando el RSI.

Una vez más el resultado se muestra en formato de tabla y se exporta a un archivo .txt para guardar los resultados y poder enviarlos fácilmente.

1 Hoja	2 Empresa	3 Tendencia último día	4 Tendencia último semana	5 Cruce dorado	6 Cruce de la muerte
1	"Acciona, S.A."	1		1'0'	'0'
2	"Acerinox, S.A."	1		1'0'	'0'
3	"ACS, Actividades de Construcción y Servicios, S.A."	1	0.6000'1'		'0'
4	"Aena S.M.E., S.A."	1	0.2000'0'		'0'
5	"Almirall, S.A."	1	0.6000'0'		'0'
6	"Amadeus IT Group, S.A."	1	0.4000'0'		'0'
7	"ArcelorMittal"	1	0.8000'0'		'0'
8	"Banco Bilbao Vizcaya Argentaria, S.A."	0		0'0'	'0'
9	"Banco de Sabadell, S.A."	0	0.2000'0'		'0'
10	"Banco Santander, S.A."	1	0.4000'0'		'0'
11	"Bankinter, S.A."	1		1'0'	'0'
12	"CaixaBank, S.A."	1	0.8000'0'		'0'
13	"Cellnex Telecom, S.A."	1	0.4000'0'		'0'
14	"CIE Automotive, S.A."	1	0.4000'0'		'0'
15	"Enagás, S.A."	1		1'0'	'0'
16	"Endesa, Sociedad Anonima"	1		1'0'	'0'
17	"Ferrovial, S.A."	1	0.8000'0'		'0'
18	"Fluidra, S.A."	1	0.6000'0'		'0'
19	"Grifols, S.A."	1		1'0'	'0'
20	"Iberdrola, S.A."	1		1'0'	'0'
21	"Indra Sistemas, S.A."	0		0'0'	'0'
22	"Industria de Diseño Textil, S.A."	1	0.4000'0'		'0'
23	"Inmobiliaria Colonial, SOCIMI, S.A."	1		1'0'	'0'
24	"International Consolidated Airlines Group, S.A."	1	0.4000'0'		'0'
25	"Mapfre, S.A."	0	0.8000'0'		'0'
26	"Meliá Hotels International, S.A."	0		0'0'	'0'
27	"MERLIN Properties SOCIMI, S.A."	0		0'0'	'0'
28	"Naturgy Energy Group, S.A."	1		1'0'	'0'
29	"Pharma Mar, S.A."	0		0'0'	'0'
30	"Red Eléctrica Corporación, S.A."	1		1'0'	'0'
31	"Repsol, S.A."	0	0.4000'0'		'0'
32	"Siemens Gamesa Renewable Energy, S.A."	1	0.8000'0'		'0'
33	"Solaria Energía y Medio Ambiente, S.A."	1	0.4000'0'		'0'
34	"Telefónica, S.A."	0		0'0'	'0'
35	"Viscofan, S.A."	1	0.8000'0'		'0'

Tabla 15. Asignación tendencias y cruce dorado y cruce de la muerte. Fuente: elaboración propia

Por último, el resultado de la cartera de inversión se guarda en una variable adicional en forma de tabla como se muestra a continuación. Este resultado también se exporta a .txt para facilitar la visualización y el envío de información entre usuarios.

1 Hoja	2 Empresa	3 Cluster	4 RSI
35	"Viscofan, S.A."	1	53.2110
18	"Fluidra, S.A."	2	53.0973
7	"ArcelorMittal"	3	69.4818
28	"Naturgy Energy Group, S.A."	4	40.7767

Tabla 16. Cartera resultado del análisis basado en RSI y medias móviles. Fuente: elaboración propia

3.8 VISUALIZACIONES

Adicionalmente, se han creado funciones para visualizar rendimientos, cotizaciones, cartera, medias móviles, tendencias, RSI y otros gráficos. Esta herramienta se incorpora al motor de trading y sirve para que el analista valore las decisiones de inversión y tenga los datos de forma visual. Además, al estar usando estrategias comprar y mantener (*buy&hold*) es importante el uso de estas funciones que sirven como elemento para la monitorización de la cartera una vez seleccionada. De esta forma el usuario puede hacer un seguimiento preciso de sus inversiones y en caso de verlas peligrar cambiar la cartera.

Una vez seleccionada la cartera, se usan gráficos como los de subida y bajada que muestran el máximo y mínimo valor de las acciones seleccionadas para la cartera. También se representan una comparación del rendimiento diario de la cartera, frente al índice de referencia, en este caso el IBEX 35. Mediante estos gráficos se presenta una visualización rápida de la cartera elegida sin entrar en detalles sobre cada una de las compañías adquiridas.

Estos gráficos muestran carteras elegidas mediante RSI, carteras elegidas mediante estrategias combinadas de RSI junto con medias móviles y carteras elegidas mediante análisis de noticias.

A continuación se muestra dicha visualización para la cartera seleccionada mediante análisis de RSI.

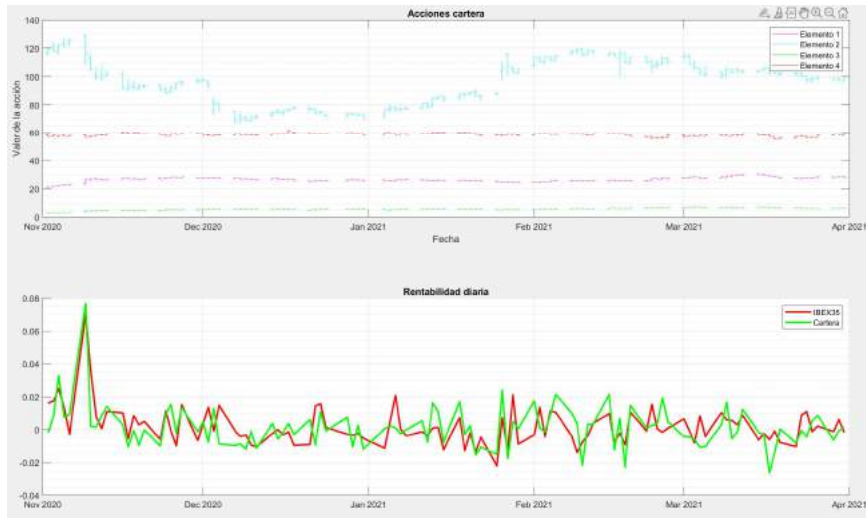


Ilustración 30. Visualización cartera seleccionada mediante análisis de RSI. Fuente: elaboración propia.

Para un análisis detenido de cada compañía se incorporan una serie de visualizaciones con opciones avanzadas para el análisis de los distintos indicadores comentados a lo largo de este proyecto.

El primer paso es la selección de la compañía que se desea analizar. Para ello se muestra una lista para que el usuario elija:

Lista de empresas:

- 1 - Acciona, S.A. ANA.MC,
- 2 - Acerinox, S.A. ACX.MC
- 3 - ACS, Actividades de Construcción y Servicios, S.A. ACS.MC
- 4 - Aena S.M.E., S.A. AENA.MC
- 5 - Almirall, S.A. ALM.MC
- 6 - Amadeus IT Group, S.A. AMS.MC
- 7 - ArcelorMittal MTS.MC
- 8 - Banco Bilbao Vizcaya Argentaria, S.A. BBVA.MC
- 9 - Banco de Sabadell, S.A. SAB.MC
- 10 - Banco Santander, S.A. SAN.MC
- 11 - Bankinter, S.A. BKT.MC
- 12 - CaixaBank, S.A. CABK.MC
- 13 - Cellnex Telecom, S.A. CLNX.MC
- 14 - CIE Automotive, S.A. CIE.MC
- 15 - Enagás, S.A. ENG.MC
- 16 - Endesa, Sociedad Anonima ELE.MC
- 17 - Ferrovial, S.A. FER.MC
- 18 - Fluidra, S.A. FDR.MC
- 19 - Grifols, S.A. GRF.MC
- 20 - Iberdrola, S.A. IBE.MC
- 21 - Indra Sistemas, S.A. IDR.MC
- 22 - Industria de Diseño Textil, S.A. ITX.MC
- 23 - Inmobiliaria Colonial, SOCIMI, S.A. COL.MC
- 24 - International Consolidated Airlines Group, S.A. IAG.MC
- 25 - Mapfre, S.A. MAP.MC
- 26 - Meliá Hotels International, S.A. MEL.MC
- 27 - MERLIN Properties SOCIMI, S.A. MRL.MC
- 28 - Naturgy Energy Group, S.A. NTGY.MC
- 29 - Pharma Mar, S.A. PHM.MC
- 30 - Red Eléctrica Corporación, S.A. REE.MC
- 31 - Repsol, S.A. REP.MC
- 32 - Siemens Gamesa Renewable Energy, S.A. SGRE.MC
- 33 - Solaria Energía y Medio Ambiente, S.A. SLR.MC
- 34 - Telefónica, S.A. TEF.MC
- 35 - Viscofan, S.A. VIS.MC

¿Que empresas desea visualizar?

Tabla 17. Petición de elección de la empresa por parte del programa. Fuente: elaboración propia.

Una vez seleccionado el número de la empresa a visualizar se exponen los siguientes gráficos para la monitorización:

1. Índice de fuerza relativa (RSI). Se visualiza la cotización de la empresa seleccionada junto con el oscilador de fuerza de inercia. Adicionalmente, se incorporan los límites superiores (en 70) e inferiores (en 30) para facilitar la monitorización por parte del usuario.

El objetivo de esta gráfica es informar sobre el estado de sobrecompra o sobreventa de las acciones seleccionadas.

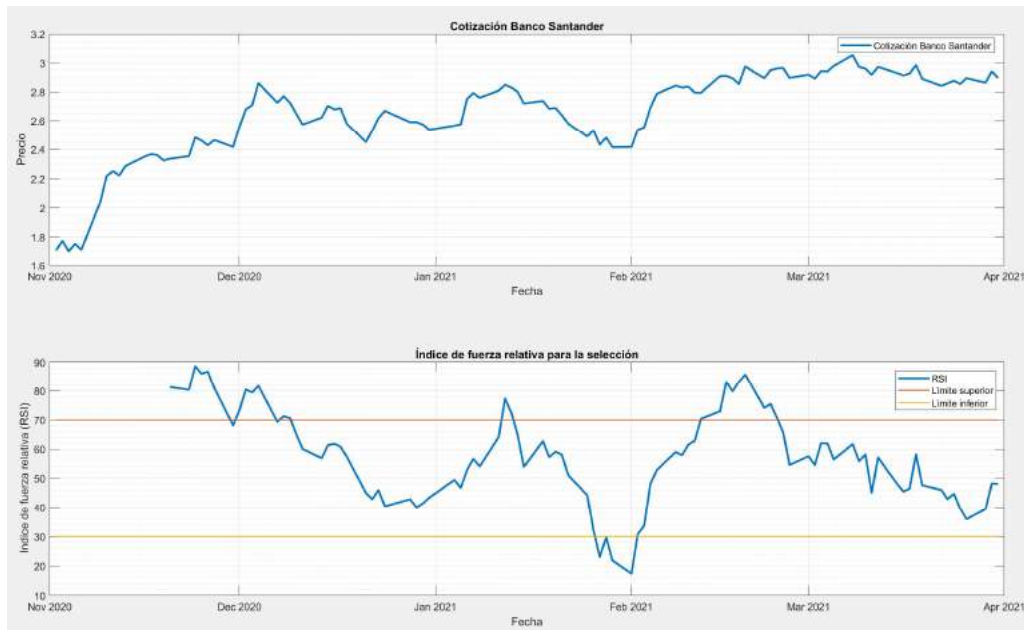


Ilustración 31. Monitorización cotización e Índice de fuerza relativo. Fuente: elaboración propia.

2. Media móvil y tendencia. Se muestra la cotización de la empresa junto una media móvil simple de 10 periodos (el valor de la media móvil puede modificarse en función de las preferencias del usuario). Además, se representa la diferencias entre la cotización la media móvil simple en forma gráfico de barras, usando colores verdes para tendencias alcistas y colores rojos para tendencias a la baja.
El objetivo es informar sobre posibles cambios en las tendencias y la fuerza que tienen. Para mayor tamaño de las barras más fuerte es la tendencia que están representando.

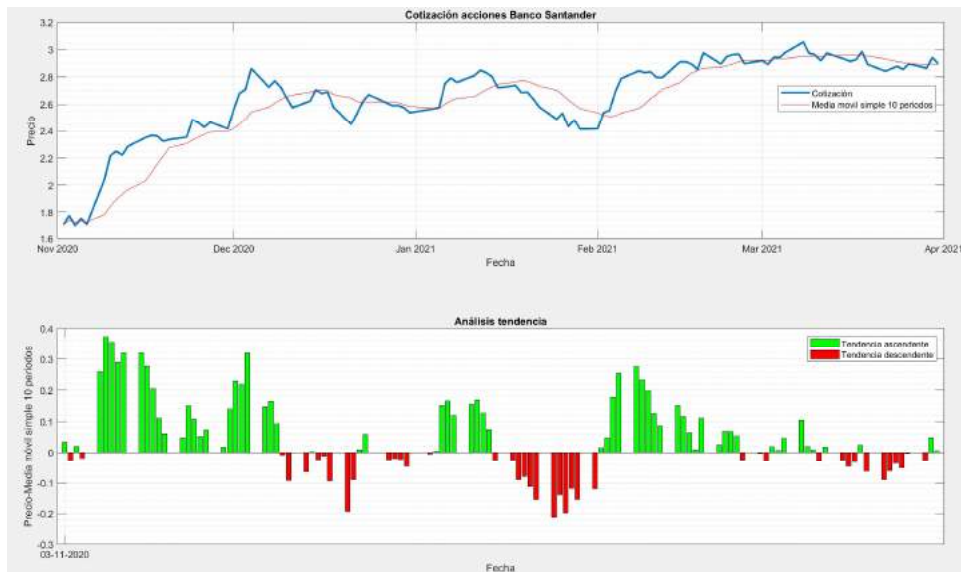


Ilustración 32. Monitorización media móvil simple y tendencia. Fuente: elaboración propia.

3. Cruce de medias móviles: cruce dorado y cruce de la muerte. Esta herramienta sirve para monitorizar posibles puntos de inflexión en la cotización de la empresa. Mediante cruces negros se representan los cruces de la muerte indicando recomendación de venta y con cruces dorados, indicando recomendación de compra.

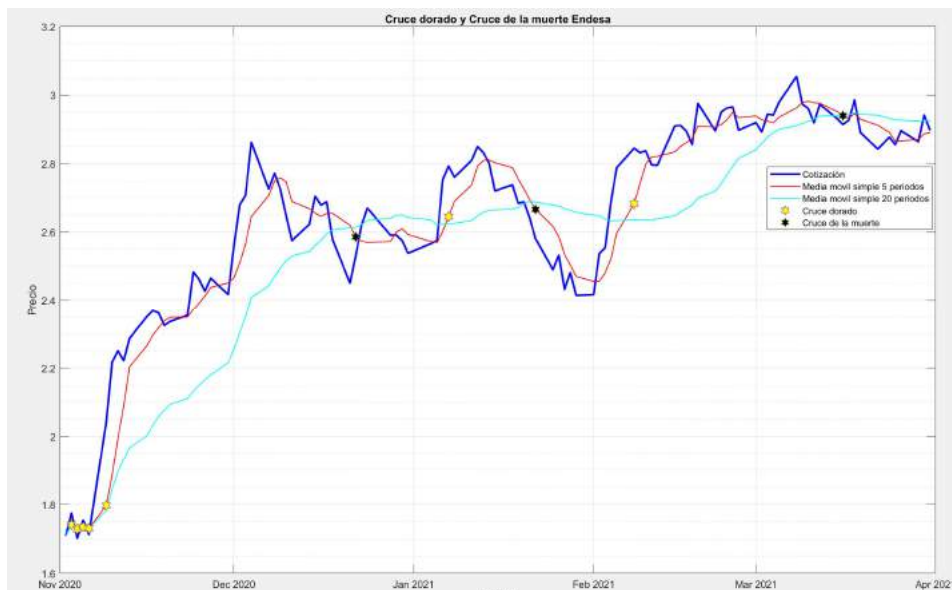


Ilustración 33. Gráficos cruce de la muerte y cruce dorado. Fuente: elaboración propia.

4. Análisis convergencia/divergencia de media móviles (MACD). Este gráfico es una herramienta similar al cruce de la muerte y cruce de dorado. Se representa la cotización de la empresa estudiada, la señal MACD, la señal de una media móvil exponencial de 9 periodos y unas barras con la diferencia entre MACD y media móvil exponencial de 9 periodos.

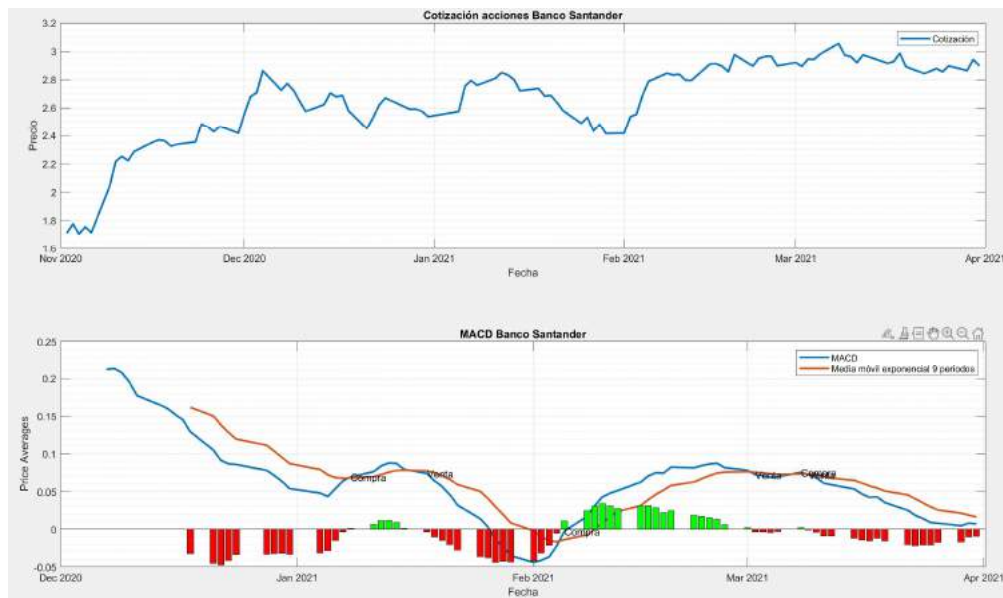


Ilustración 34. Análisis convergencia divergencia usando MACD. Fuente: elaboración propia.

5. Análisis de noticias. Por último se incluye un análisis de la cotización de la empresa seleccionada, en la que se indican con triángulos verdes momentos en los que se han detectado noticias con una connotación positiva, indicando un ascenso en la cotización de las acciones. Por otro lado, con triángulos rojos se representan las noticias negativas. Adicionalmente se representan dos medias móviles de 15 y 25 periodos.



Ilustración 35. Gráfico cotización junto con medias móviles y datos noticias. Fuente: elaboración propia.

Capítulo 4. EXPERIMENTOS Y PRUEBAS DEL SISTEMA DESARROLLADO

A lo largo de este capítulo se realizan pruebas y experimentos para testar el sistema desarrollado. Para ello se profundizará en el análisis de las compañías pertenecientes al índice IBEX 35 mostrado en Tabla 18. Índice IBEX 35. Fuente: elaboración propia. Adicionalmente, se recurrirá al análisis de otros índices como el DAX 30 y CAC 40 para realizar comparaciones.

Nº Company	Stock Ticker	Nº Company	Stock Ticker
1 Acciona, S.A.	ANA.MC	19 Grifols, S.A.	GRF.MC
2 Acerinox, S.A.	ACX.MC	20 Iberdrola, S.A.	IBE.MC
3 ACS, Actividades de Construcción y Servicios, S.A.	ACS.MC	21 Indra Sistemas, S.A.	IDR.MC
4 Aena S.M.E., S.A.	AENA.MC	22 Industria de Diseño Textil, S.A.	ITX.MC
5 Almirall, S.A.	ALM.MC	23 Inmobiliaria Colonial, SOCIMI, S.A.	COL.MC
6 Amadeus IT Group, S.A.	AMS.MC	24 International Consolidated Airlines Group, S.A.	IAG.MC
7 ArcelorMittal	MTS.MC	25 Mapfre, S.A.	MAP.MC
8 Banco Bilbao Vizcaya Argentaria, S.A.	BBVA.MC	26 Meliá Hotels International, S.A.	MEL.MC
9 Banco de Sabadell, S.A.	SAB.MC	27 MERLIN Properties SOCIMI, S.A.	MRL.MC
10 Banco Santander, S.A.	SAN.MC	28 Naturgy Energy Group, S.A.	NTGY.MC
11 Bankinter, S.A.	BKT.MC	29 Pharma Mar, S.A.	PHM.MC
12 CaixaBank, S.A.	CABK.MC	30 Red Eléctrica Corporación, S.A.	REE.MC
13 Cellnex Telecom, S.A.	CLNX.MC	31 Repsol, S.A.	REP.MC
14 CIE Automotive, S.A.	CIE.MC	32 Siemens Gamesa Renewable Energy, S.A.	SGRE.MC
15 Enagás, S.A.	ENG.MC	33 Solaria Energía y Medio Ambiente, S.A.	SLR.MC
16 Endesa, Sociedad Anonima	ELE.MC	34 Telefónica, S.A.	TEF.MC
17 Ferrovial, S.A.	FER.MC	35 Viscofan, S.A.	VIS.MC
18 Fluidra, S.A.	FDR.MC		

Tabla 18. Índice IBEX 35. Fuente: elaboración propia.

4.1 AGRUPACIÓN MÉTODO JERÁRQUICO IBEX 35

En este apartado se aplica el método jerárquico descrito en la sección 3.4.2 aplicado a las 35 acciones del IBEX 35. Esta prueba ha sido realizada extrayendo datos bursátiles en directo de Yahoo Finance para el período de tiempo entre el 01/11/2020 y el 01/04/2021. Se trata por lo tanto de la agrupación de una serie de 106 días de cotización en el IBEX 35.

En primer lugar, a partir de los datos individuales (Apertura, Cierre, Máximo, Mínimo) se ha calculado la tasa de retorno diaria de cada acción durante todo el periodo de estudio. Con la serie de tasas de retorno se ha ejecutado el algoritmo de la Tabla 13 para calcular la

distancia (medida de similitud) entre acciones obteniéndose la matriz de distancias de 35x35 que es la entrada para la agrupación jerárquica de empresas.

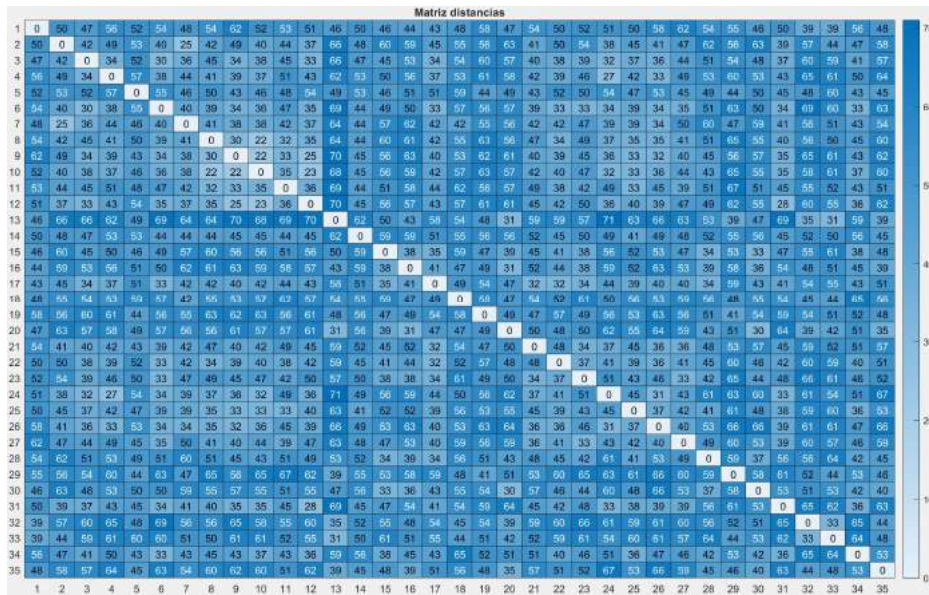


Ilustración 36. Matriz de distancias. Fuente: elaboración propia.

Estamos usando un conjunto de 106 de días por lo tanto la máxima distancia posible entre dos acciones es 106. Sin embargo, la distancia de la gran mayoría de valores es menor a 65 y no existen valores que disten entre ellos más de una puntuación de 71. Por lo tanto, como primera conclusión observamos que al tratarse de empresas del mismo índice mantienen todas cierta similitud.

En segundo lugar, se usa la matriz de distancias para agrupar las 35 empresas usando *clustering* jerárquico y distancias *Ward*. El dendograma correspondiente se muestra a continuación. Se ha realizado un corte a 65 unidades de distancia obteniéndose un total de 5 clústeres.³

³ Se puede observar que el dendograma de la Ilustración 37 tiene 30 nodos de salida en lugar de 35. Además 4 de los nodos tiene número en lugar de nombre. Esto se debe a una limitación de Matlab que permite un máximo de 30 nodos iniciales. Al tener 5 en nuestro ejemplo los agrupa antes del árbol. Por lo tanto, el nodo 9 por ejemplo es la combinación de Caixa Bank, Santander y Sabadell.

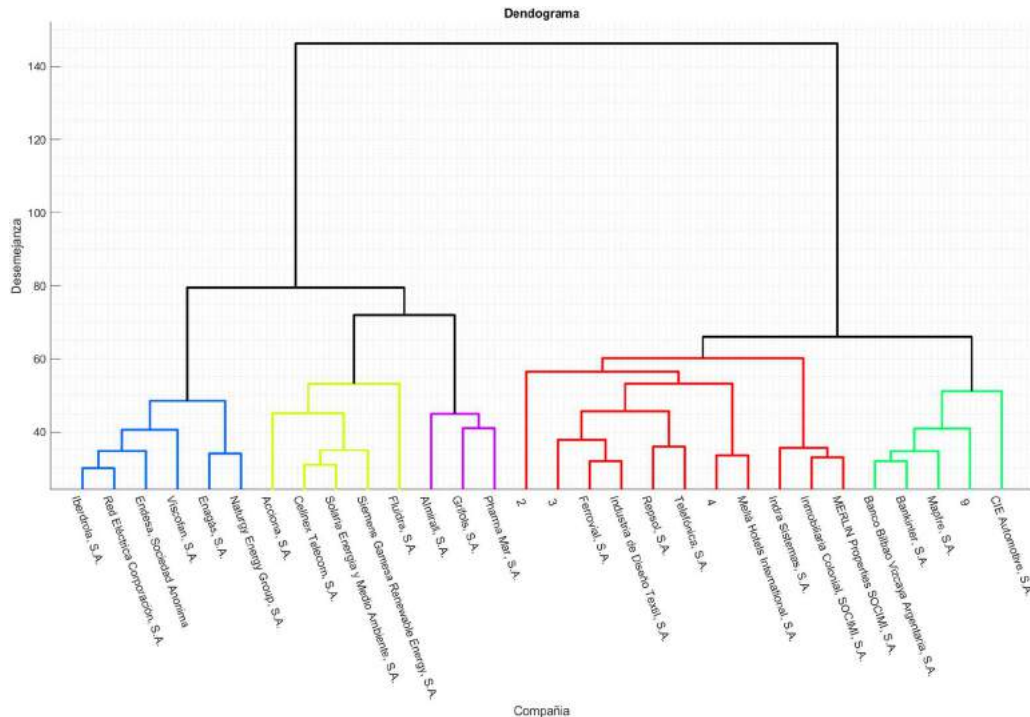


Ilustración 37. Dendrograma clustering jerárquico IBEX 35. Fuente: elaboración propia.

Finalmente, en la Tabla 19. Agrupación empresas IBEX 35 según clustering jerárquico. Fuente: elaboración propia. Se muestran las empresas agrupadas. Como conclusión final se puede observar que los grupos tienden a crearse por empresas de sectores similares. En el *Cluster_1* se agrupan 6 empresas del sector financiero y CIE Automotriz, del sector industrial. En el *Cluster_2* se agrupan 3 empresas farmacéuticas (sector bienes de consumo). En el *Cluster_3* se encuentran agrupadas 5 empresas del sector eléctrico y energético y Viscofan, una farmacéutica. El *Cluster_4* y *Cluster_5* son una mezcla de sectores con sector materiales básicos e industria, tecnología y telecomunicaciones, servicios inmobiliarios y servicios de consumo.

A raíz de esto podemos concluir que los sectores energético, farmacéutico y financiero son cíclicos o nicho y sus empresas tienden a evolucionan de forma similar.

Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5
BBVA.MC	ALM.MC	ENG.MC	ACX.MC	ANA.MC
SAB.MC	GRF.MC	ELE.MC	ACS.MC	CLNX.MC
SAN.MC	PHM.MC	IBE.MC	AENA.MC	FDR.MC
BKT.MC		NTGY.MC	AMS.MC	SGRE.MC
CABK.MC		REE.MC	MTS.MC	SLR.MC
CIE.MC		VIS.MC	FER.MC	
MAP.MC			IDR.MC	
			ITX.MC	
			COL.MC	
			IAG.MC	
			MEL.MC	
			MRL.MC	
			REP.MC	
			TEF.MC	

Tabla 19. Agrupación empresas IBEX 35 según clustering jerárquico. Fuente: elaboración propia.

4.2 APLICACIÓN CLUSTERING JERÁRQUICO AL DAX 30

En este apartado se repite el proceso de agrupación para la acciones del índice alemán DAX. El objetivo es probar la versatilidad del programa creado, capaz de extraer y analizar datos de cualquier índice en tiempo real. Además, se analiza si existe diferencia en la dispersión de las acciones del IBEX 35 y DAX 30. Para el mismo periodo (01/11/2020 y el 01/04/2021) se comparan las matrices de distancias y número de grupos para el mismo corte del árbol jerárquico.

Los títulos del índice DAX 30 analizados se muestran en siguiente tabla:

Nº Company	Stock Ticker	Nº Company	Stock Ticker
1 Henkel AG & Co. KGaA	HEN3.DE	16 Fresenius SE & Co. KGaA	FRE.DE
2 SAP SE	SAP.DE	17 Deutsche Bank Aktiengesellschaft	DBK.DE
3 Siemens Aktiengesellschaft	SIE.DE	18 Fresenius Medical Care AG & Co. KGaA	FME.DE
4 Bayer Aktiengesellschaft	BAYN.DE	19 Siemens Energy AG	ENR.DE
5 Münchener Rückversicherungs-Gesellschaft Aktiengesellschaft in München	MUV2.DE	20 E.ON SE	EOAN.DE
6 Vonovia SE	VNA.DE	21 Daimler AG	DAI.DE
7 Deutsche Börse AG	DB1.DE	22 BASF SE	BAS.DE
8 Allianz SE	ALV.DE	23 Deutsche Telekom AG	DTE.DE
9 Linde plc	LIN.DE	24 RWE Aktiengesellschaft	RWE.DE
10 Bayerische Motoren Werke Aktiengesellschaft	BMW.DE	25 Deutsche Post AG	DPW.DE
11 adidas AG	ADS.DE	26 Continental Aktiengesellschaft	CON.DE
12 Volkswagen AG	VOW3.DE	27 Delivery Hero SE	DHER.DE
13 MERCK Kommanditgesellschaft auf Aktien	MRK.DE	28 HeidelbergCement AG	HEI.DE
14 MULTI-UNITS LUXEMBOURG - Lyxor Euro Government Bond (DR) UCITS ETF - Acc	MTX.DE	29 Infineon Technologies AG	IFX.DE
15 Deutsche Wohnen SE	DWNI.DE	30 Covestro AG	1COV.DE

Tabla 20. Índice DAX 30. Fuente: elaboración propia.

Siguiendo el mismo proceso que en el apartado anterior se obtiene la siguiente matriz de distancias:

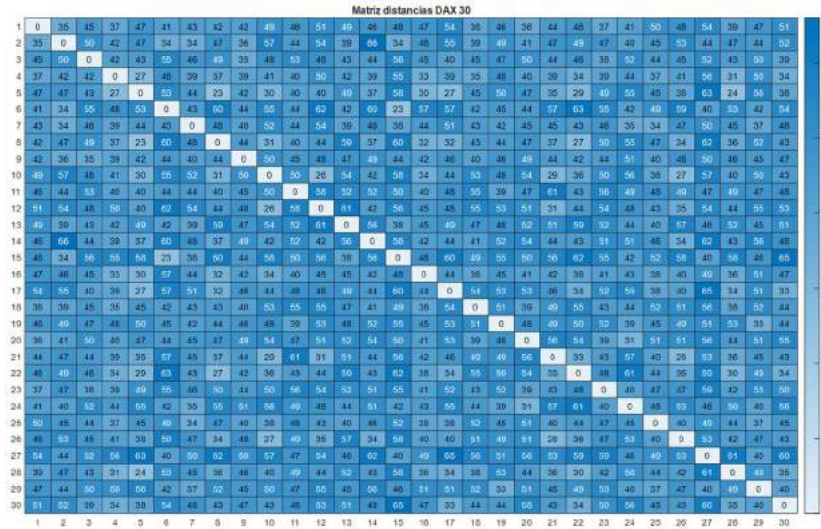


Ilustración 38. Matriz distancias DAX 30. Fuente: elaboración propia.

Como primera conclusión se observan menores distancias entre compañías. Por lo tanto, se puede afirmar que el DAX ha sido un índice más compacto cuyas acciones se mueven más uniformemente durante el periodo estudiado. Dentro del DAX la mayor distancia entre empresas es de 66 para un periodo de 106 días.

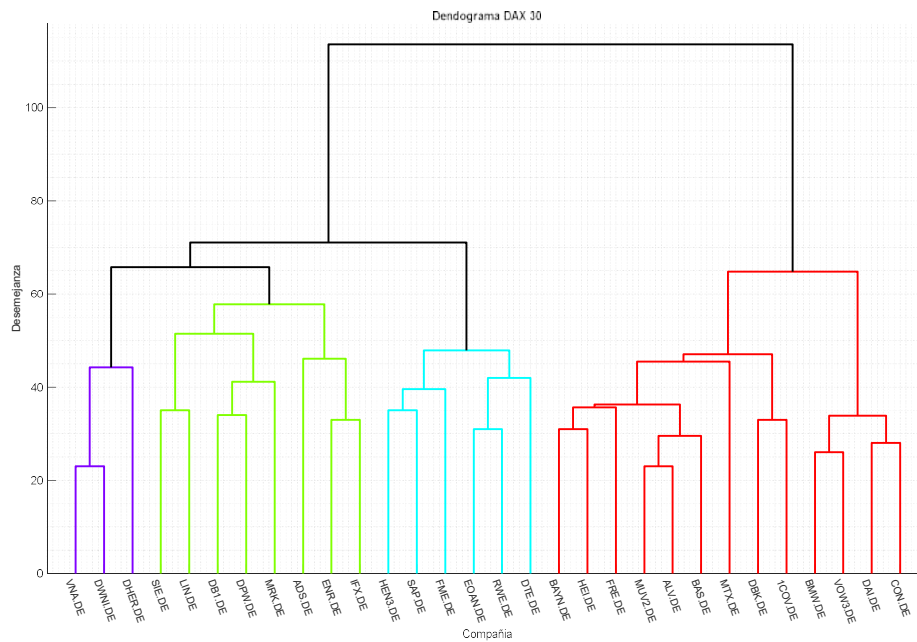


Ilustración 39. Dendrograma clustering jerárquico DAX 30. Fuente: Elaboración propia.

Finalmente los grupos obtenidos se muestran en la Tabla 21. Agrupación empresas DAX 30 según clustering jerárquico. Fuente: elaboración propia. El primer grupo muestra dos empresas inmobiliarias y una de reparto de comida a domicilio. El *Cluster_2* muestra empresas de distintos sectores: 1 trading Marketplace, 2 de industria, 4 de tecnología, 1 de distribución y logística. El *Cluster_3* tiene 2 empresas tecnológica y telecomunicaciones, 1 de servicios sanitarios, 2 energética, 1 de industria.

Cluster_1	Cluster_2	Cluster_3	Cluster_4
VNA.DE	SIE.DE	HEN3.DE	BAYN.DE
DWNI.DE	DB1.DE	SAP.DE	MUV2.DE
DHER.DE	LIN.DE	FME.DE	ALV.DE
	ADS.DE	EOAN.DE	BMW.DE
	MRK.DE	DTE.DE	VOW3.DE
	ENR.DE	RWE.DE	MTX.DE
	DPW.DE		FRE.DE
	IFX.DE		DBK.DE
			DAI.DE
			BAS.DE
			CON.DE
			HEI.DE
			1COV.DE

Tabla 21. Agrupación empresas DAX 30 según clustering jerárquico. Fuente: elaboración propia.

4.3 APLICACIÓN CLUSTERING JERÁRQUICO AL CAC 40

Se repite el proceso para las empresas pertenecientes al CAC 40 siguiendo el listado de (Euronext Paris, 2021) que se muestra a continuación:

Nº	Company	Stock Ticker	Nº	Company	Stock Ticker
1	Acelormital SA	MT.AS	21	L'Oréal S.A.	OR.PA
2	Airbus SE	AIR.PA	22	LVMH Moët Hennessy - Louis Vuitton, Société Européenne	MC.PA
3	Altom SA	ALO.PA	23	Orange S.A.	ORA.PA
4	Atos SE	ATO.PA	24	Pernod Ricard SA	RI.PA
5	AXA SA	CS.PA	25	Publicis Groupe S.A.	PUB.PA
6	BNP Paribas SA	BNP.PA	26	Renault SA	RNO.PA
7	Bouygues SA	EN.PA	27	Safran	SAF.PA
8	Capgemini SE	CAP.PA	28	Sanofi	SAN.PA
9	Carrefour SA	CA.PA	29	Schneider Electric S.E.	SU.PA
10	Compagnie de Saint-Gobain S.A.	SGO.PA	30	Société Générale Société anonyme	GLE.PA
11	Compagnie Générale des Établissements Michelin Société en commandite par actions	ML.PA	31	Stellantis N.V.	STLA.PA
12	Crédit Agricole S.A.	ACA.PA	32	STMicroelectronics N.V.	STM.PA
13	Danone S.A.	BN.PA	33	Teleperformance SE	TEP.PA
14	Dassault Systèmes SE	DSY.PA	34	Thales S.A.	HQ.PA
15	ENGIE SA	ENGI.PA	35	TOTAL SE	FP.PA
16	EssilorLuxottica Société anonyme	EL.PA	36	Veolia Environnement S.A.	VIE.PA
17	Hermès International Société en commandite par actions	RMS.PA	37	VINCI SA	DG.PA
18	Kering SA	KER.PA	38	Vivendi SE	VIV.PA
19	L'Air Liquide S.A.	AI.PA	39	Worldline SA	WLN.PA
20	Legrand SA	LR.PA	40	Unibail-Rodamco-Westfield SE	URW.AS

Tabla 22. Índice CAC 30. Fuente: elaboración propia

La matriz de distancias y dendograma resultantes de la agrupación se muestran a continuación:

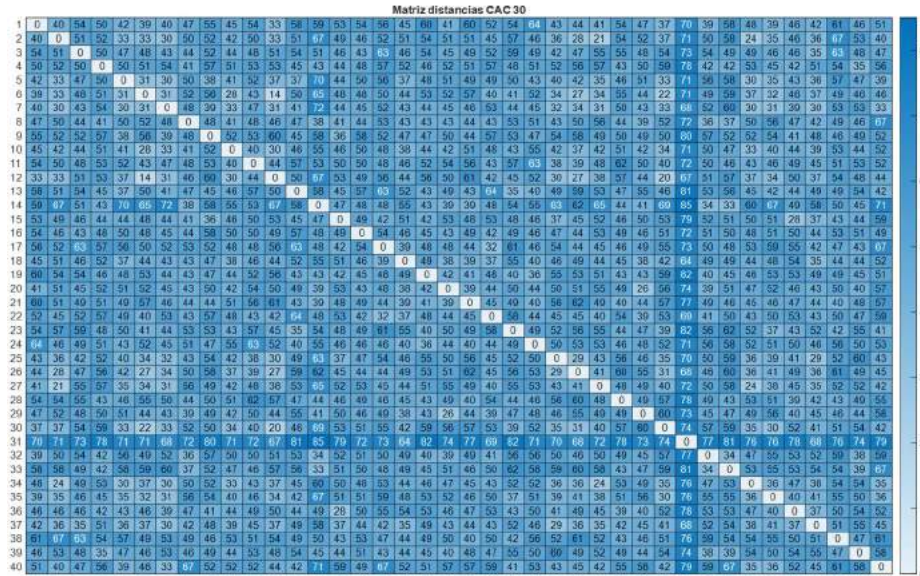


Ilustración 40. Matriz distancias CAC 40. Fuente: elaboración propia.

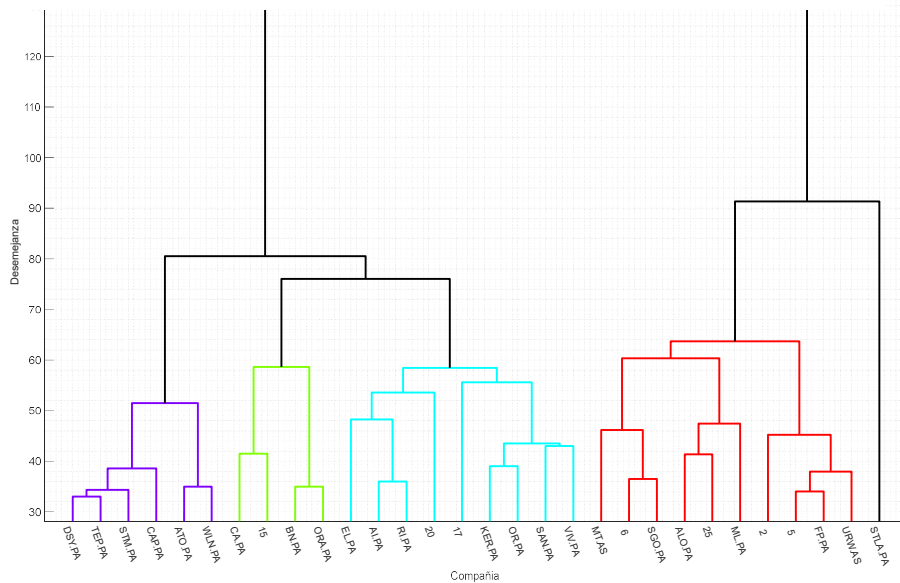


Ilustración 41. Dendograma clustering jerárquico CAC 40. Fuente: elaboración propia.

El título número 31 Stellantis N.V muestra un comportamiento alejado del resto de acciones considerándose un outlier. Se trata de un grupo automovilístico internacional fruto de la fusión de Fiat Chrysler Automobiles y Groupe PSA. El grupo se consolidó empezando a

cotizar en bolsa de Paris y Milán el 18 de enero de 2021 y en Nueva York 19 de enero de 2021. Esto tiene lugar en mitad de nuestro periodo de estudio explicándose así su comportamiento anormal respecto al resto de valores. Los grupos obtenidos del índice CAC 40 se muestran a continuación:

Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5
EL.PA	CA.PA	ATO.PA	MT.AS	STLA.PA
RMS.PA	BN.PA	CAP.PA	AIR.PA	
KER.PA	ENGI.PA	DSY.PA	ALO.PA	
AI.PA	ORA.PA	STM.PA	CS.PA	
LR.PA	VIE.PA	TEP.PA	BNP.PA	
OR.PA		WLN.PA	EN.PA	
MC.PA			SGO.PA	
RI.PA			ML.PA	
SAN.PA			ACA.PA	
SU.PA			PUB.PA	
VIV.PA			RNO.PA	
			SAF.PA	
			GLE.PA	
			HO.PA	
			FP.PA	
			DG.PA	
			URW.AS	

Tabla 23. Agrupación empresas CAC 40 según clustering jerárquico. Fuente: elaboración propia.

4.4 K-MEDIAS APLICADO AL IBEX 35:

Por otro lado, se aplica el método K-medias para el intervalo de estudio (01/11/2020 y el 01/04/2021) para la acciones del IBEX 35. El método K-medias no acepta series temporales ni series temporales de intervalo como input y, por lo tanto, es necesario manipular los datos.

4.4.1 K-MEDIAS 2 VARIABLES ABSOLUTO:

Siguiendo a (Cruz, 2019) se ha realizado una agrupación mediante el uso de las variables rentabilidad y volatilidad absolutas de cada acción durante el periodo de estudio. La rentabilidad se calcula según la siguiente fórmula para todo el periodo. Es decir, para el caso de estudio:

$$rentabilidad = \frac{Precio\ cierre_{01/04/2021} - Precio\ apertura_{01/11/2020}}{Precio\ apertura_{01/11/2020}}$$

Por otro lado, la volatilidad se define como:

$$\text{volatilidad} = \frac{\text{Max. Nivel Periodo} - \text{Min. Nivel Periodo}}{\text{Min. Nivel Periodo}}$$

Se ejecutan ambas operaciones para los 35 títulos del IBEX 35 y se aplica el método K-medias usando la distancia euclídea y con 100 iteraciones de inicio. Como se puede ver en la Ilustración 42. Agrupación IBEX 35 mediante k medias con dos variables absolutas. Fuente: elaboración propia. El número óptimo de grupos pueden ser 3 o 4 siguiendo la regla del codo.

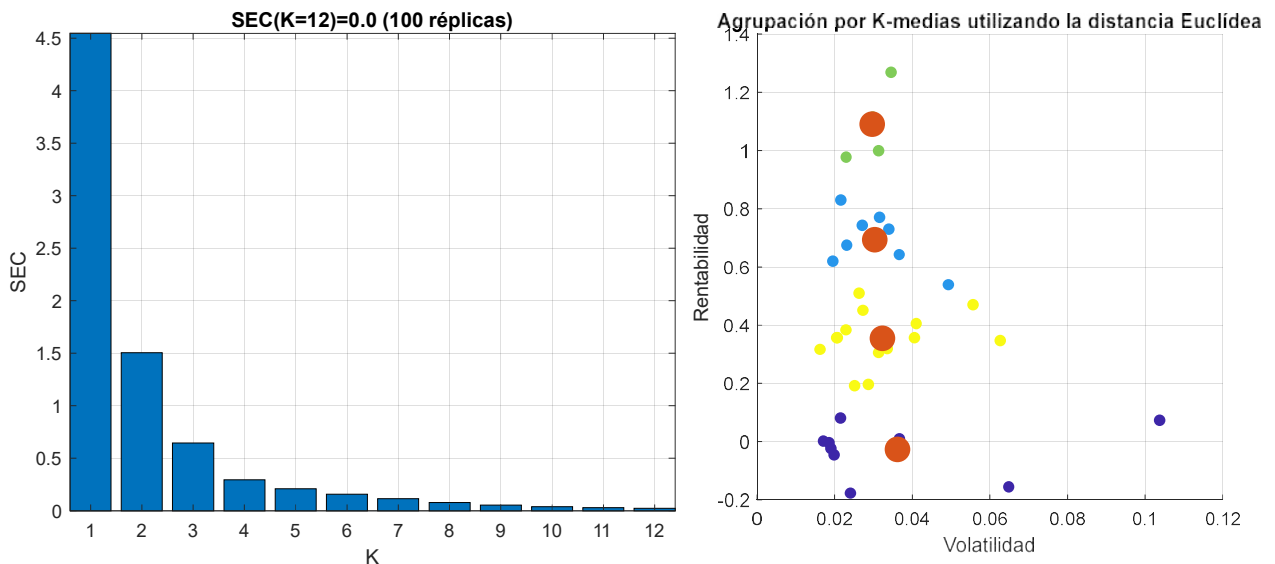


Ilustración 42. Agrupación IBEX 35 mediante k medias con dos variables absolutas. Fuente: elaboración propia.

Finalmente se escogen 4 grupos para representar las acciones del IBEX 35. La solución gráfica se observa en la Tabla 24. El *Cluster_1* (morado), el *Cluster_2* (azul), el *Cluster_3* (en verde), el *Cluster_4* (amarillo).

Cluster_1	Cluster_2	Cluster_3	Cluster_4
CLNX.MC	ANA.MC	MTS.MC	ACS.MC
ENG.MC	ACX.MC	IAG.MC	AENA.MC
ELE.MC	BBVA.MC	MEL.MC	ALM.MC
GRF.MC	SAB.MC	REP.MC	AMS.MC
IBE.MC	SAN.MC		CIE.MC
PHM.MC	BKT.MC		FER.MC
REE.MC	CABK.MC		IDR.MC
SLR.MC	FDR.MC		ITX.MC
VIS.MC			COL.MC
			MAP.MC
			MRL.MC
			NTGY.MC
			SGRE.MC
			TEF.MC

Tabla 24. Agrupación empresas IBEX 35 según k medias con variables absolutas. Fuente: elaboración propia.

4.4.2 K-MEDIAS 2 VARIABLES MEDIA SEMANAL:

Al trabajar usando sólo los valores de apertura y cierre del periodo es posible que se comentan errores por la aparición de valores anormales en los extremos. Para mitigar esto se mejora el método de *clustering* del apartado anterior ejecutando las rentabilidades y volatilidades semanales de cada acción y calculando la media de las mismas para todo el periodo de estudio.

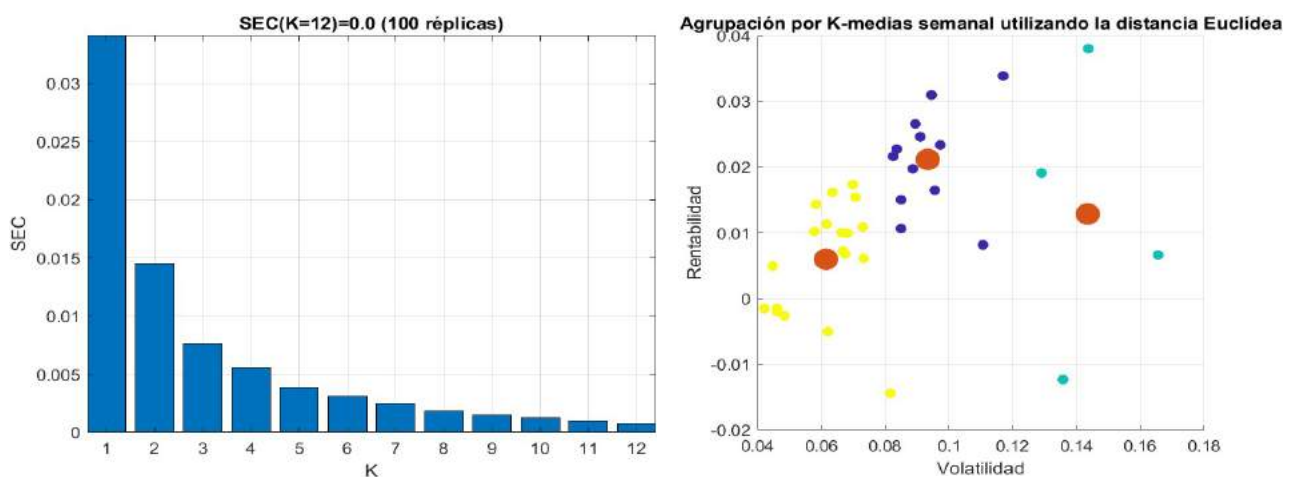


Ilustración 43. Agrupación IBEX 35 con k medias con dos variables medias semanales. Fuente: elaboración propia.

El cluster 1 (morado) representa aquellas acciones con volatilidad media y rentabilidad positiva, el cluster 2 (amarillo) representa acciones con rentabilidad media y volatilidad baja, el cluster 3 (cian) representa un grupo de acciones anormales con alta volatilidad y rentabilidades dispares.

Cluster_1	Cluster_2	Cluster_3
ANA.MC	SGRE.MC	SAB.MC
AMS.MC	TEF.MC	IAG.MC
MTS.MC		PHM.MC
BBVA.MC	SLR.MC	ALM.MC
SAN.MC		CLNX.MC
BKT.MC		CIE.MC
CABK.MC		ENG.MC
		NTGY.MC

Tabla 25. Agrupación empresas IBEX 35 mediante k medias con variables medias. Fuente: elaboración propia.

4.5 EVALUACIÓN RENDIMIENTO CARTERAS

Además del sistema explicado en el Capítulo 3. se ha creado un programa para testear los rendimientos de los distintos tipos de carteras que se pueden generar. El objetivo es evaluar el rendimiento de la cartera en los días siguientes a su creación. El código permite la selección de número de días que se desea evaluar.

A continuación se realizan una serie de pruebas.

En la pestaña a0_Inicialización se selecciona el periodo de estudio para analizar compañías. En este caso seleccionamos el periodo entre 01 de noviembre de 2020 y el 1 de abril de 2021. Además se selecciona el método de agrupación deseado: jerárquico, k medias con variables rentabilidad-volatilidad total, k medias con variables rentabilidad-volatilidad media semanal.

Una vez cargada esta información se cargan los *tickers* de las empresas objetivo. Para este ejercicio se cogen las empresas pertenecientes al IBEX35 durante el periodo de estudio.

Con esta información se ejecuta el programa creándose las carteras correspondientes en cada caso y se evalúa a catorce días vista el rendimiento de cada una de ellas. A continuación se muestran los resultados obtenidos.

4.5.1 K MEDIAS CON VARIABLES RENTABILIDAD-VOLATILIDAD TOTAL:

Las tres carteras generadas mediante k medias usando variables globales de rendimiento y volatilidad son:

Usando el indicador de fuerza relativa o RSI:

Hoja	Empresa	Cluster	RSI
26	"Meliá Hotels International, S.A."	1	32.9231
29	"Pharma Mar, S.A."	2	37.0370
8	"Banco Bilbao Vizcaya Argentaria, S.A."	3	41.1681
22	"Industria de Diseño Textil, S.A."	4	28.5068

Tabla 26. Cartera generada usando k medias con variables absolutas y análisis mediante índice de fuerza relativa. Fuente: elaboración propia

Usando estrategias combinando indicador de fuerza relativa (RSI) junto con análisis de medias móviles:

Hoja	Empresa	Cluster	RSI
7	"ArcelorMittal"	1	69.4818
35	"Viscofan, S.A."	2	53.2110
18	"Fluidra, S.A."	3	53.0973
28	"Naturgy Energy Group, S.A."	4	40.7767

Tabla 27. Cartera generada usando k medias con variables absolutas y análisis mediante índice de fuerza relativa y medias móviles. Fuente: elaboración propia

Por último, usando análisis del sentimiento en las noticias publicadas durante los últimos días en Expansión y Cinco Días:

Hoja	Empresa	Cluster	RSI
31	"Repsol, S.A."	1	44.9735
13	"Cellnex Telecom, S.A."	2	70.7234
10	"Banco Santander, S.A."	3	48.0587
34	"Telefónica, S.A."	4	38.2263

Tabla 28. Cartera generada usando k medias con variables absolutas y análisis de noticias. Fuente: elaboración propia

Finalmente, el resultado de los distintos rendimientos comparado con el IBEX 35 es:

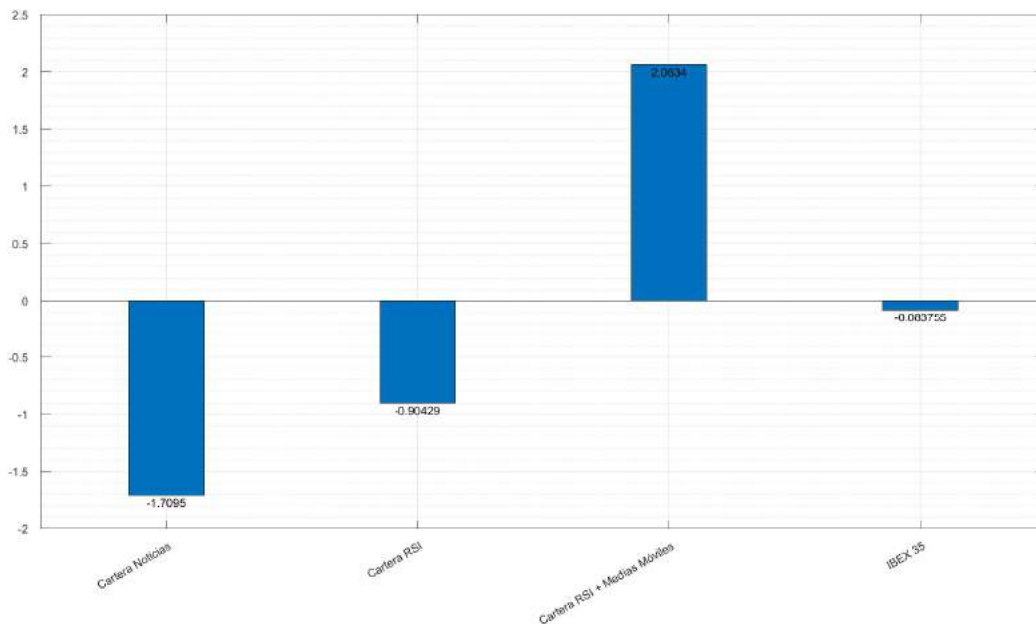


Ilustración 44. Comparación rendimiento para agrupación con k medias absoluto. Fuente: elaboración propia.

4.5.2 K MEDIAS CON VARIABLES RENTABILIDAD-VOLATILIDAD MEDIA

Las tres carteras generadas mediante K medias usando variables globales de rendimiento y volatilidad son:

Usando el indicador de fuerza relativa o RSI:

Hoja	Empresa	Cluster	RSI
29	"Pharma Mar, S.A."	1	37.0370
35	"Viscofan, S.A."	2	53.2110
22	"Industria de Diseño Textil, S.A."	3	28.5068
26	"Meliá Hotels International, S.A."	4	32.9231

Tabla 29. Cartera generada usando k medias con variables medias y análisis mediante índice de fuerza relativa. Fuente: elaboración propia

Usando estrategias combinando indicador de fuerza relativa (RSI) junto con análisis de medias móviles:

Hoja	Empresa	Cluster	RSI
9	"Banco de Sabadell, S.A."	1	51.1712
35	"Viscofan, S.A."	2	53.2110
28	"Naturgy Energy Group, S.A."	3	40.7767
18	"Fluidra, S.A."	4	53.0973

Tabla 30. Cartera generada usando k medias con variables medias y análisis mediante índice de fuerza relativa y medias móviles. Fuente: elaboración propia

Por último, usando análisis del sentimiento en las noticias publicadas durante los últimos días en Expansión y Cinco Días:

Hoja	Empresa	Cluster	RSI
9	"Banco de Sabadell, S.A."	1	51.1712
20	"Iberdrola, S.A."	2	53.7671
13	"Cellnex Telecom, S.A."	3	70.7234
10	"Banco Santander, S.A."	4	48.0587

Tabla 31. Cartera generada usando k medias con variables medias y análisis de noticias. Fuente: elaboración propia

Finalmente, el resultado de los distintos rendimientos comparado con el IBEX 35 es:

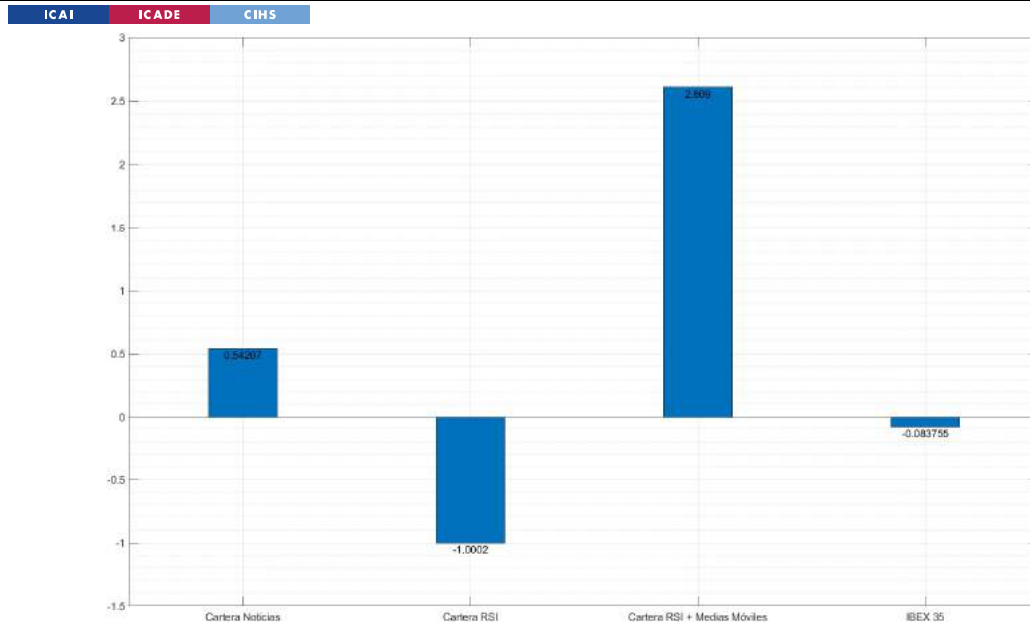


Ilustración 45. Comparación rendimiento para agrupación con k medias con variables medias. Fuente: elaboración propia.

4.5.3 JERÁRQUICO

Las tres carteras generadas mediante método jerárquico usando la distancia explicada en la Tabla 7 son:

Usando el indicador de fuerza relativa o RSI:

Hoja	Empresa	Cluster	RSI
8	"Banco Bilbao Vizcaya Argentaria, S.A."	1	41.1681
22	"Industria de Diseño Textil, S.A."	2	28.5068
29	"Pharma Mar, S.A."	3	37.0370
33	"Solaria Energía y Medio Ambiente, S.A."	4	41.1960
28	"Naturgy Energy Group, S.A."	5	40.7767

Tabla 32. Cartera generada usando método jerárquico y análisis mediante índice de fuerza relativa. Fuente : elaboración propia

Usando estrategias combinando indicador de fuerza relativa (RSI) junto con análisis de medias móviles:

Hoja	Empresa	Cluster	RSI
14	"CIE Automotive, S.A."	1	43.3962
27	"MERLIN Properties SOCIMI, S.A."	2	42.8571
5	"Almirall, S.A."	3	65.2174
18	"Fluidra, S.A."	4	53.0973
28	"Naturgy Energy Group, S.A."	5	40.7767

Tabla 33. Cartera generada usando método jerárquico y análisis mediante índice de fuerza relativa y medias móviles. Fuente: elaboración propia

Por último, usando análisis del sentimiento en las noticias publicadas durante los últimos días en Expansión y Cinco Días:

Hoja	Empresa	Cluster	RSI
10	"Banco Santander, S.A."	1	48.0587
34	"Telefónica, S.A."	2	38.2263
29	"Pharma Mar, S.A."	3	37.0370
13	"Cellnex Telecom, S.A."	4	70.7234
20	"Iberdrola, S.A."	5	53.7671

Tabla 34. Cartera generada usando método jerárquico y análisis de noticias. Fuente: elaboración propia

Finalmente, el resultado de los distintos rendimientos comparado con el IBEX 35 es:

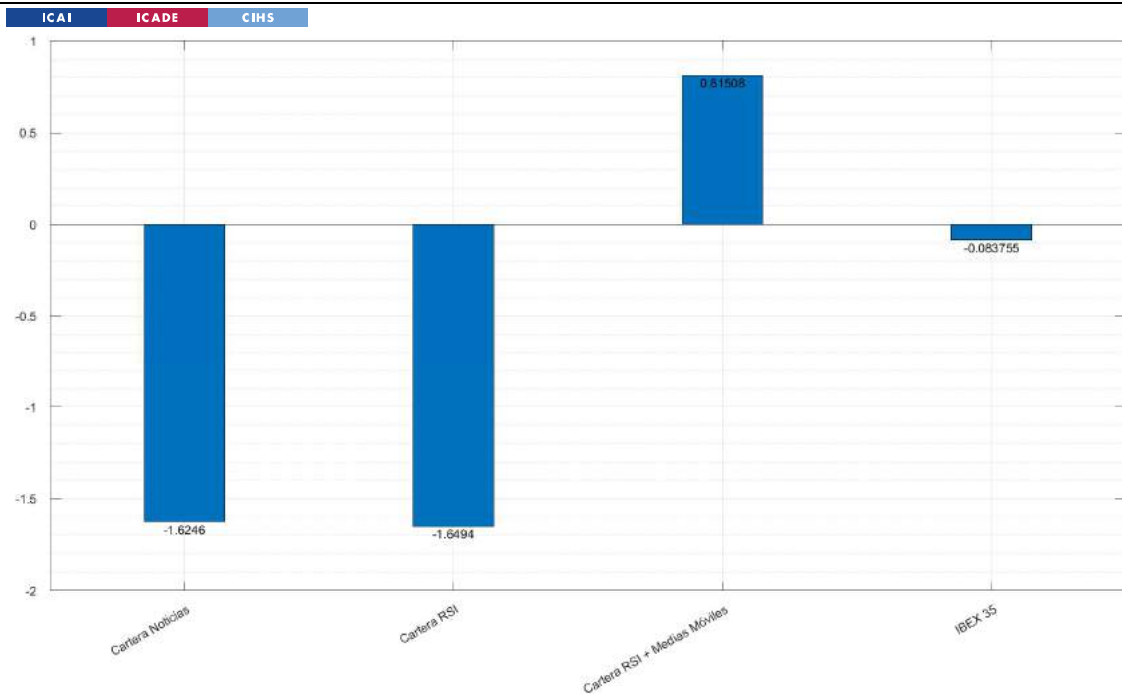


Ilustración 46. Comparación rendimiento para agrupación con método jerárquico. Fuente: elaboración propia.

Capítulo 5. CONCLUSIONES Y FUTUROS DESARROLLOS

Este capítulo proporciona un resumen del trabajo realizado haciendo hincapié en los principales objetivos alcanzados. Asimismo, se presenta una propuesta de mejoras y futuros desarrollos para el trabajo.

5.1 CONCLUSIONES

A lo largo de esta memoria se han recorrido múltiples aspectos relacionados con la creación de carteras, métodos de inversión, análisis de sentimiento y sistemas automatizados de análisis de inversiones. Se ha comenzado con un repaso de la teoría existente hasta el momento, revisando el análisis técnico clásico mediante medias móviles e índice de fuerza relativo, teorías de creación de carteras de Markowitz (premio nobel 1990) y la hipótesis de mercados eficientes de Eugene Fama (Premio Nobel 2013). A continuación se ha profundizado en aspectos innovadores relacionados con el agrupamiento de datos, la minería y extracción de información de internet y el análisis de sentimiento. Por último se han repasado algunas de las principales estrategias de trading y se han analizado los sistemas automatizados existentes.

Habiendo finalizado el estado del arte se ha desarrollado un sistema original combinando todos los elementos mencionados anteriormente. Para ello se ha creado un diagrama (Ilustración 28. Arquitectura del sistema desarrollado. Fuente: elaboración propia.) fruto de una serie de iteraciones en las que se han ido implementando mejoras. Este diagrama es la base que servirá como referencia para la creación de un sistema automatizado usando la aplicación MATLAB 2020. Este programa ofrece multitud de posibles funciones y se permite al usuario la elección de una serie de parámetros para afinar los análisis en función de las necesidades particulares de cada usuario. Además, el código se plantea como una herramienta para su uso a lo largo del tiempo, permitiendo la extracción de información financiera y de noticias en tiempo real. Gracias a esto el usuario puede ejecutar el programa

periódicamente para obtener resultados con la última información disponible. De esta forma se pueden crear distintas carteras a diario y analizar la variación de estas consiguiendo una ventaja competitiva respecto a otros inversores. También se ofrece la posibilidad de acceder a datos pasados (mientras que estén disponibles en Yahoo Finance) para el estudio del comportamiento de las acciones deseadas en el pasado.

En segundo lugar, el programa ofrece la creación de una base de datos de titulares de noticias ordenadas por empresas para su análisis. Este elemento del sistema solo está disponible para dos periódicos ubicados en España y por lo tanto, no permite esta funcionalidad para el trading basado en sentimiento en otros mercados.

Otro de los puntos fuertes del código creado es la posibilidad de aplicarlo a distintos mercados financieros. Mediante una serie de inputs (listado de empresas objetivo) el usuario puede elegir que empresas desea analizar. La información se extrae de Yahoo Finance por lo que el abanico de posibilidades abarca todas las disponibles en Yahoo.

Además, se incorporan una serie de visualizaciones que sirven para la monitorización de la cartera elegida o para la creación de lista de seguimientos. Estas visualizaciones no solo incluyen datos sobre la cotización de empresas sino que se incorporan análisis detallados como el cruce de medias móviles e índice de fuerza relativa que puedan servir como indicadores para promover una inversión activa por parte de los usuarios.

El proyecto realizado pretende facilitar la creación de carteras diversificadas para la creación de valor en el largo plazo. Mediante el sistema automatizado se crean recomendaciones para el usuario y se presentan una serie de visualizaciones. Estas visualizaciones ayudan a monitorizar al cartera y afinar los análisis para la una mejor decisión.

5.1.1 AGRUPACIÓN DE COMPAÑÍAS

En esta sección se han propuesto cuatro métodos para la agrupación de acciones cada uno con sus ventajas e inconvenientes.

La agrupación natural de datos bursátiles por sectores y actividad reduce la capacidad computacional del programa pero, al no tratarse de un algoritmo requiere la clasificación previa de las empresas. Como debilidad cabe destacar que se trata de un método estático y atemporal ya que no depende de los datos bursátiles sino de la naturaleza de las empresas.

La agrupación mediante matriz de distancias y árboles jerárquicos tiene como desventaja la necesidad de la creación de una medida de distancias para series temporales de intervalo. En este trabajo se propone una distancia explicada en el apartado 3.4.2. El problema de la distancia creada es que no tiene en cuenta el riesgo o volatilidad de las empresas estudiadas. Una vez creada la distancia la ejecución del programa tiene la desventaja de que requiere una elevada carga computacional. Como fortaleza este método permite la creación de carteras dinámicas que cambian en función de la cotización de las empresas. Además, el algoritmo desarrollado tiene en cuenta la similitud de acciones periódicamente ya que analiza los valores día a día comparando subidas y bajadas independientemente de la rentabilidad o volatilidad. Cabe destacar que mediante este análisis realizado en el IBEX 35 se llegan a resultados en los que compañías de naturalezas similares tiendan a estar juntas. Esto se explica con detenimiento en el apartado 4.1.

La agrupación mediante k medias, siguiendo la teoría de Markowitz usa volatilidad y rentabilidad como variables para de entrada al sistema. Al tratarse de series temporales de intervalo es necesario acotar la volatilidad y rentabilidad a la de un periodo de estudio determinado. Para ello se propone (i) análisis con variables absolutas del periodo y (ii) análisis con variables medias del periodo. La carga computacional de este método es media siendo un aspecto favorable a tener en cuenta.

5.1.2 ANÁLISIS DE NOTICIAS

En este ámbito se han probado distintos diccionarios encontrando un compromiso entre el uso de lexicón especializado en mercados financieros, pero con un reducido número de palabras o lexicón genérico con mayor número de términos. Habiendo realizado múltiples pruebas se optó por el uso de un lexicón genérico.

Debido al reducido número de palabras de cada titular es necesario un diccionario extenso para evitar tener noticias sin clasificar. El problema de un diccionario extenso es que se puede incurrir en errores de asignación de sentimiento por mala interpretación de los términos o términos demasiado genéricos. Por ello se hace un llamamiento a la comunidad científica y se expondrá en futuros desarrollos la necesidad de la creación de diccionarios especializados en castellano.

Además, usando el diccionario castellano genérico de LIWC se observó un sesgo positivo en los titulares analizados. A raíz de esto se llega a dos posibles hipótesis: (i) connotación positiva de los titulares de noticias en los periódicos Expansión y Cinco días, (ii) sesgo positivo en la creación del diccionario de LIWC.

Cobra más fuerza la hipótesis referente a la connotación positiva de los titulares de noticias debido a que la trayectoria de LIWC está contrastada con uso en otros trabajos de investigación. Además, se podría explicar el sesgo positivo en los titulares como una estrategia de marketing por parte de los periódicos para atraer a los lectores. Esto explica porque los resultados de rendimientos en carteras usando el análisis de noticias son peores.

Por último, mencionar la baja eficiencia del análisis de noticias para empresas del mercado continuo. A lo largo de este trabajo se realizaron una serie de pruebas para un conjunto de 50 empresas formado por empresas del mercado continuo y del IBEX 35. Debido a su menor tamaño y popularidad el nombre de las empresas del mercado continuo no aparecía en los titulares de noticias y por lo tanto, no se dispone de datos para llevar a cabo un análisis de sentimiento.

5.1.3 VISUALIZACIÓN Y MONITORIZACIÓN

Para la visualización, este proyecto incluye una serie de gráficos innovadores en los que combina datos bursátiles con análisis de sentimiento de noticias y análisis técnico explicado en el apartado 3.8 y que sirven como herramienta de soporte para las decisiones de los usuarios.

5.1.4 CONCLUSIONES SOBRE IBEX 35

Habiendo estudiado con mayor detalle el índice IBEX 35 se pueden extraer una serie de breves conclusiones referentes a este grupo de acciones.

En primer lugar, se observa que el IBEX 35 se trata de un índice más dispar que otros similares como el DAX o el CAC. Durante el periodo de estudio las distancias entre acciones alcanzan valores mayores. Además, el IBEX 35 tiene una componente muy fuerte en empresas del sector financiero.

Por otro lado, el número de noticias sobre las compañías del IBEX 35 varía mucho en función de la empresa. De esta forma, compañías como el Banco Santander son más mediáticas que otras como Viscofan. Por este motivo el trading mediante noticias no es uniforme y se dificulta la obtención de buenos resultados.

Por último, se puede concluir que el análisis del IBEX 35 no se comporta adecuadamente según los criterios de índice de fuerza relativa (RSI). Sin embargo, el análisis con medias móviles combinado con RSI es mucho más productivo obteniéndose rentabilidades de la cartera que baten al índice.

5.2 FUTUROS DESARROLLOS

Este proyecto es fruto de una idea ambiciosa de combinar la mayor cantidad de información para maximizar la diversificación de una cartera y maximizar la rentabilidad de las inversiones. Para ello ha sido necesario abarcar distintos ámbitos de estudio con la respectiva restricción de tiempo. Se ha logrado el diseño y creación de un sistema robusto pero es fundamental exponer una serie de elementos en los que profundizar como futuros desarrollos para la maximización del potencial del trabajo.

En primer lugar, en el ámbito del análisis de sentimiento de noticias en español es necesario la creación o ampliación del Lexicón financiero en castellano. Con esto se aumenta la precisión en análisis de noticias. Además, es conveniente ampliar el análisis de sentimiento

de los titulares al cuerpo de la noticia. Esta posee más información y supone una ventaja en el análisis en caso de disponer un buen diccionario. No obstante puede ser una desventaja en caso de disponer del lexicón inadecuado ya que puede introducir ruido en la calificación global que se le da a una compañía.

En este proyecto se usa un sistema no supervisado para la clasificación de noticias y se plantea una metodología innovadora para el uso de un sistema supervisado. Este análisis supervisado requiere la participación de expertos en mercados financieros y se puede implementar usando los pasos explicados en el apartado 3.2.3 y el cuestionario modelo del Anexo 1.

Por otro lado, se puede mejorar la calidad de los análisis mediante la extracción de información de más periódicos financieros. Esta ampliación permitiría no solo mejorar los resultados en carteras del IBEX 35 sino también extender esta funcionalidad a otros mercados como el DAX 30 o el CAC 40. Esta ampliación supondría un gran salto ya que requiere análisis de sentimiento en distintos idiomas. Para ello se recomienda el uso de Python ya que dispone de funciones creadas para la extracción de información de la web.

Otra de las limitaciones que tiene el programa generado es la cantidad de grupos de acciones que se pueden crear. Actualmente, el código permite la agrupación hasta en cinco grupos y es necesario que el usuario seleccione manualmente cuantos grupos quiere en función de los gráficos de centroides o diagrama de árbol. La automatización de esto se propone también para desarrollos futuros. Asimismo, se propone la incorporación de costes en las transacciones para la evaluación de los rendimientos.

Por último, este proyecto expone diversos códigos realizados en MATLAB 2020. De cara a mejorar la experiencia del usuario sería conveniente desarrollar una aplicación que englobe los códigos.

Capítulo 6. REFERENCIAS

¿Qué es el Indicador de Análisis Técnico MACD? (n.d.). Retrieved from bit2me Academy:
<https://academy.bit2me.com/que-es-indicador-analisis-tecnico-macd/>

(2021, marzo 19). Retrieved from Expansión:
https://www.expansion.com/empresas.html?intcmp=MENUHOM24101&s_kw=empresas

Albertos, Á. P. (2018). *Leyendo más allá de las palabras: Uso del análisis de textos para medir el sentimiento de mercado en España*. Madrid: Universidad Pontificia de Comillas ICADE-ICAI.

Anquan, W. (n.d.). *Stocks*. Retrieved from Webull: <https://app.webull.com/stocks>

Baviera, T. (2016). *Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength*. Dígitos: Revista de Comunicación Digital.

Bernal, Á. I. (2020). *Análisis de sentimiento del informe de estabilidad financiera*. Madrid: Banco de España.

C., R. C. (n.d.). *Estrategia de Cruces de tres medias móviles exponenciales*. Retrieved from tecnicasdetrading.com: <https://www.tecnicasdetrading.com/2014/01/estrategia-cruces-tres-medias-moviles-exponenciales.html>

Chong, T. T.-L. (2008). *Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30*. Applied Economics Letters.

Chong, T. T.-L. (n.d.). *Technical analysis and the London stock exchange: .*

Cinco días. (2021, febrero 27). Retrieved from Cinco días Compañías:
<https://cincodias.elpais.com/seccion/companias>

- CNAE Clasificación Nacional de Actividades Económicas. (2009). *CNAE Clasificación Nacional de Actividades Económicas*. Retrieved from <https://www.cnae.com.es/>
- Crespo, M. R. (2019). *Visual trading system based on moving averages*. Madrid: Universidad Pontificia de Comillas.
- Cruz, P. H. (2019). *Sistema analítico de agrupamiento de series temporales de intervalo y su aplicación en finanzas*. Madrid: Universidad Pontificia de Comillas.
- Cunningham, P. (n.d.). *Chapter 2 Supervised Learning*.
- Euronext Paris. (2021). Retrieved from <https://live.euronext.com/en/product/indices/FR0003500008-XPAR/market-information#index-composition>
- Fama, E. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. The Journal of Finance.
- Ferrari, F. J. (2016, enero 21). *Frontera de carteras eficiente*. Retrieved from Economipedia.com: <https://economipedia.com/definiciones/frontera-carteras-eficiente.html>
- Ferreruela, N. B. (2017). *Behavioral Finance: ¿Por qué los inversores se comportan como lo hacen y no como deberían?* Zaragoza: Estudios y Publicaciones Estudio.
- finanzas.com. (2016, septiembre 22). *Triple cruce de medias: menos frecuente pero más efectivo*. Retrieved from finanzas.com: https://www.finanzas.com/mercados/triple-cruce-de-medias-menos-frecuente-pero-mas-efectivo_13479458_102.html
- Gubu, L. (2019). *Classicak Portfolio Selection with Cluster Analysis: Comparison Between Hierarchical Complete Linkage and Ward Algorithm*. AIP Publishing.
- Hayes, A. (2021, Abril 30). *Investopedia*. Retrieved from <https://www.investopedia.com/terms/s/sma.asp>

- He-Shan Guan, Q.-S. J. (2007). *Cluster Financial Time Series for Portfolio*. Beijing: Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition.
- Liaño, B. G.-G. (2012). *La metodología Delphi como técnica de estudio de la validez de contenido*. Madrid: Anales de psicología: Ediciones de la Universidad de Murcia.
- LIWC. (2015). *LIWC*. Retrieved from <http://liwc.wpengine.com/>
- Madrid, B. d. (n.d.). *Listado de Empresas por Sectores*. Retrieved from Bolsa de Madrid: <https://www.bolsamadrid.es/esp/aspx/Empresas/EmpresasPorSectores.aspx?sector=01>
- Malkiel, B. G. (1973). *A random walk down Wall Street*. Nueva York: W.W. Norton & Company, Inc.
- Markets, A. (2020). *Qué es una Estrategia de trading de Momentum y cuándo utilizarla*. Madrid: Admiral Markets. Retrieved febrero 28, 2021, from <https://admiralmarkets.com/es/education/articles/forex-strategy/estrategia-momentum>
- Markowitz, H. M. (1991). *Foundations of Portfolio Theory*. Journal of Finance.
- Maté, C. (2012). El análisis de intervalos. Aplicaciones a la ingeniería. *Anales de mecánica y electricidad*.
- Maté, C. (2021). *Combining Interval Time Series Forecast. A First Step in a Long Way. (Research Agenda)*. Madrid: Revista Colombiana de Estadística - Applied Statistics.
- Mittermayer, M.-A. (2004). *Forecasting Intraday Stock Price Trends*. Hawaii: Proceedings of the 37th Hawaii International Conference on System Sciences.

- Montero, J. A. (2020). *Predicción de precios en mercados financieros usando la información de las noticias a través de métodos estadísticos*. Madrid: Universidad Pontificia de Comillas.
- MytripleA. (2017, marzo 01). Retrieved from MytripleA:
<https://www.mytriplea.com/blog/tipos-inversores-guia-infografia/>
- Nuij, W. (n.d.). *An Automated Framework for Incorporating*. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING.
- Pan, W. (2020). *Portfolio Learning Based on Deep Learning*. Shanghai: MDPI Journal: Future Internet.
- Quiroga, E. (2017). *Eficiencia en los mercados financieros y predicción de precios de los activos*. Ciencias Administrativas Revista Digital FCE UNLP.
- Renfree, J. (2021). *File Exchange*. Retrieved from MathWorks:
https://www.mathworks.com/matlabcentral/fileexchange/18458-hist_stock_data-start_date-end_date-varargin
- Rokach, L. (2014). *Chapter 15: Clustering Methods*. Tel-Aviv: Tel-Aviv University.
- Romero, J. (2019, junio 14). *Jorge Romero*. Retrieved from https://jorgeromero.net/analisis-cluster/#Metodos_de_particionamiento
- Schumaker, R. P. (2009). *Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System*. Tucson, Arizona: Artificial Intelligence Lab, Department of Management Information Systems.
- Stefan Feuerriegel, H. P. (July 19, 2018). *News-based trading strategies*. Tokyo & Freiburg.
- Takeuchi, L. (2013). *Applying Deep Learning to Enhance Momentum Trading Strategies in Stocks*.

Zhao, B. (2017). *Web Scraping*. College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA: Springer International Publishing AG.

ANEXO 1: ENCUESTA ANÁLISIS DE TITULARES.

A continuación se adjunta parte del conjunto de titulares de noticias a analizar por expertos siguiendo metodología Delphi. Se trata de un conjunto de 150 titulares de noticias en los que se menciona a alguna de las compañías target del estudio. Este formulario se comparte con los expertos en Excel a través de OneDrive.

MUCHAS GRACIAS POR LA COLABORACIÓN. POR FAVOR LEER CUIDADOSAMENTE LAS INSTRUCCIONES.

Instrucciones:

1/ Mirar columna B y buscar el número de las noticias que se le han asignado.

2/ Leer las noticias asignadas (columna C).

3/ Asignar una puntuación a cada noticia. Para ello escribir en la columna E correspondiente en función de lo positivo o negativo que es la noticia para la empresa de la columna D correspondiente.

4/ Se puntúa con 1 las noticias que tengan un efecto positivo para la empresa, 0 para noticias que tengan un efecto neutro y -1 para noticias que tengan efecto negativo.

** La negatividad, positividad o neutralidad se evalúa como el efecto en el corto plazo (hora-día-semana) que tiene esa noticia en el valor de la acción correspondiente.

- Positivo (1) = la noticia provoca un aumento del valor de la acción
- Negativo (-1) = provoca una disminución del valor de la acción
- Neutro (0) = la noticia es indiferente para el valor de la acción

Nº	Source	Source	Y	Source	Source
1	Repsol cumple 2 años en el negocio eléctrico con 1,1 millones de clientes y 2 GW de proyectos nuevos 'verdes'	Repsol		2020/11/01	Expansión
2	Amancio Ortega cobra mañana 646,8 millones de euros del dividendo de Inditex	Inditex		2020/11/01	Expansión
3	Los analistas mejoran las estimaciones para Fluidra tras presentar sus resultados hasta septiembre	Fluidra		2020/11/01	Expansión
4	ACS cancela un contrato de derivados y lo sustituye por uno 'forward' de 12 millones de acciones	ACS		2020/11/02	Expansión
5	Telefónica y Vértice 360 se alían para la digitalización y gestión de catálogos audiovisuales	Telefónica		2020/11/02	Expansión
6	PNM Resources (Iberdrola) acelera su salida del carbón en 2024, siete años antes de lo previsto	Iberdrola		2020/11/02	Expansión
7	Alantra crea una tituladora europea y ficha en Santander a su responsable	Santander		2020/11/03	Expansión
8	Juan Abelló triplica su inversión en Merlin en el tercer trimestre	Merlin		2020/11/03	Expansión
9	Santander arranca las conversaciones con los sindicatos previas al ERE	Santander		2020/11/03	Expansión
10	Santander y Norges entran en Soltec, estrella bursátil de las renovables en España	Santander		2020/11/03	Expansión
11	Alantra lanza una tituladora liderada por un ex-Santander	Santander		2020/11/04	Expansión
12	Cellnex pierde 84 millones hasta septiembre por los costes de crecimiento	Cellnex		2020/11/04	Expansión
13	Cellnex confirma "conversaciones avanzadas" con Hutchison para comprar sus torres en Europa valoradas en 10.000 millones	Cellnex		2020/11/04	Expansión
14	La compra de las torres de Hutchison aprovecharía al máximo la capacidad financiera actual de Cellnex	Cellnex		2020/11/04	Expansión
15	Endesa gana 1.511 millones tras estabilizar el impacto del Covid	Endesa		2020/11/04	Expansión
16	Santander Totta ganó 255 millones hasta septiembre, un 35 % menos	Santander		2020/11/04	Expansión
17	Cómo gestionan Mapfre y Catalana Occidente el Covid fuera de España	Mapfre		2020/11/04	Expansión
18	BBVA lanza con Amazon una plataforma en la nube para operar en Bolsa	BBVA		2020/11/05	Expansión
19	Cómo le va a Sabadell tras vender su gestora?	Sabadell		2020/11/05	Expansión
20	Santander: nuevos recortes de plantilla en Europa	Santander		2020/11/05	Expansión
21	Santander remodela la cúpula de Consumer para la fusión con Openbank	Santander		2020/11/05	Expansión
22	Siemens Gamesa tiene pérdidas récord de 918 millones en su año más "complejo"	Gamesa		2020/11/05	Expansión
23	ArcelorMittal triplica pérdidas	ArcelorMittal		2020/11/05	Expansión
24	Megaplán de Iberdrola con 75.000 millones de inversión y 19.000 millones de dividendos hasta 2025	Iberdrola		2020/11/05	Expansión
25	El juez que investiga los trabajos de Villarejo para BBVA cita de nuevo al exdirectivo Antonio Béjar	BBVA		2020/11/05	Expansión
26	Telefónica crea la nueva compañía Koa Health, una spin-out de Telefónica Alpha	Telefónica		2020/11/05	Expansión
27	Así será el plan de jubilaciones y bajas incentivadas de Sabadell	Sabadell		2020/11/05	Expansión
28	Meliá registra unas pérdidas récord de 470 millones por el Covid	Meliá		2020/11/05	Expansión
29	Ferrovial lanza una emisión de bonos de 500 millones a ocho años	Ferrovial		2020/11/05	Expansión
30	Cellnex lanza una emisión de bonos convertibles por 1.500 millones	Cellnex		2020/11/05	Expansión
31	El beneficio de Acciona cae un 63% hasta septiembre por la pandemia	Acciona		2020/11/05	Expansión
32	Han estafado a Endesa en Tarragona?	Endesa		2020/11/06	Expansión
33	Amadeus pierde 356,4 millones hasta septiembre por el Covid-19	Amadeus		2020/11/06	Expansión
34	El fondo Millennium declara una posición bajista del 0,52% en Mapfre	Mapfre		2020/11/06	Expansión
35	ACS cambia la cúpula de Cimic y sale Fernández Verdes	ACS		2020/11/06	Expansión
36	Indra propone a los sindicatos amortizar 240 puestos de trabajo en su filial Indra Sistemas	Indra		2020/11/06	Expansión
37	Acciona mantiene su compromiso de sumar 5 GW hasta 2024 y ve potencial para 1,3 GW adicionales	Acciona		2020/11/06	Expansión
38	Santander dará a conocer las cifras del ERE el 13 de noviembre, según los sindicatos	Santander		2020/11/06	Expansión
39	Santander contrata a Lee Hecht Harrison para recolocar a parte de los afectados por el ERE	Santander		2020/11/06	Expansión
40	El Frob confía en que la reestructuración en la fusión de Bankia y CaixaBank se realizará "de forma razonable"	Caixa		2020/11/06	Expansión
41	Barclays mejora sus previsiones sobre el beneficio de Sabadell y BBVA este año y rebaja las de 2021	BBVA		2020/11/06	Expansión
42	Barclays mejora sus previsiones sobre el beneficio de Sabadell y BBVA este año y rebaja las de 2021	Sabadell		2020/11/06	Expansión
43	Victor Grifols deja el Consejo de Administración de Criteria Caixa	Caixa		2020/11/06	Expansión
44	Telefónica explora la venta de sus cables submarinos por 2.000 millones	Telefónica		2020/11/07	Expansión
45	La OMT premia a CaixaBank por su apoyo a la industria turística	Caixa		2020/11/07	Expansión
46	Bankinter y Liberbank disparan un 10% el ahorro en cuentas y depósitos	Bankinter		2020/11/07	Expansión
47	El Banco de Inglaterra urge a Atom Bank, participado por BBVA, y a otros neobancos a ser rentables	BBVA		2020/11/07	Expansión
48	Santander entra en el capital del broker digital DriveWealth	Santander		2020/11/07	Expansión
49	Repsol aplica la meteorología y el 'big data' para mejorar la gestión de la energía	Repsol		2020/11/08	Expansión
50	Acciona, finalista en la adjudicación de una autopista en Noruega por 252 millones de euros	Acciona		2020/11/08	Expansión

	ICAI	ICADE	CIHS			
51	Antonio Béjar, exdirectivo de BBVA, declarará de nuevo este jueves ante el juez del caso 'Villarejo'			BBVA	2020/11/08	Expansión
52	BBVA pide cambios regulatorios para tener más libertad			BBVA	2020/11/08	Expansión
53	Naturgy, Telefónica y REE reducen a mínimos históricos sus costes de financiación			Naturgy	2020/11/09	Expansión
54	Naturgy, Telefónica y REE reducen a mínimos históricos sus costes de financiación			Telefónica	2020/11/09	Expansión
55	Qué riesgos climáticos tiene BBVA			BBVA	2020/11/10	Expansión
56	Cuándo habrá opas por Merlin, Colonial o Lar			Colonial	2020/11/10	Expansión
57	Santander entra en el capital de la plataforma de compraventa residencial Clikalia			Santander	2020/11/10	Expansión
58	S&P avala el plan estratégico de Iberdrola y confirma su rating 'BBB+', con perspectiva 'estable'			Iberdrola	2020/11/10	Expansión
59	Indra plantea bajar sueldos y jubilaciones para reducir los despidos a la mitad			Indra	2020/11/10	Expansión
60	Banco Sabadell acepta prejubilaciones a partir de los 56 años			Sabadell	2020/11/10	Expansión
61	CaixaBank emite su primer bono verde por un total de 1.000 millones de euros			Caixa	2020/11/10	Expansión
62	BBVA crea un área de negocio para lanzar índices de referencia			BBVA	2020/11/11	Expansión
63	IAG acelera en los recortes para parecerse a una 'low cost'			IAG	2020/11/11	Expansión
64	Santander repite como único banco español entre los grupos sistémicos mundiales			Santander	2020/11/11	Expansión
65	Acciona se adjudica el desarrollo de un sistema de transportes de Nueva Zelanda por 400 millones			Acciona	2020/11/11	Expansión
66	Aena blindará sus contratos contra futuras pandemias			Aena	2020/11/11	Expansión
67	Cómo va a recolocar Santander al 100% de los afectados por el ERE que lo pidan?			Santander	2020/11/12	Expansión
68	BBVA se incorpora a la red de bonos sostenibles del Nasdaq			BBVA	2020/11/12	Expansión
69	Cellnex cierra la compra de las torres de Hutchison en Europa			Cellnex	2020/11/12	Expansión
70	Ferrovial construirá un tramo del gasoducto que conecta Noruega, Dinamarca y Polonia por 70 millones			Ferrovial	2020/11/12	Expansión
71	Santander amplía capital en 361 millones para pagar el dividendo en acciones de 0,10 euros			Santander	2020/11/12	Expansión
72	Merlin reduce un 65% su beneficio y extiende otros seis meses la ayuda a inquilinos			Merlin	2020/11/12	Expansión
73	Naturgy vende Chile al gigante chino State Grid por 4.300 millones con enormes plusvalías			Naturgy	2020/11/13	Expansión
74	Gual (CaixaBank) afirma que la fusión con Bankia redundará en más valor para los accionistas			Caixa	2020/11/13	Expansión
75	Endesa lanza tarifas mensuales de recarga de vehículos eléctricos			Endesa	2020/11/13	Expansión
76	Santander España recortará 4.000 empleados, un 13% de la plantilla			Santander	2020/11/13	Expansión
77	BlackRock irrumpe en el capital de Siemens Gamesa con más del 3%			Gamesa	2020/11/13	Expansión
78	Santander, BBVA, CaixaBank y Bankinter disparan su colchón de liquidez para combatir el Covid			BBVA	2020/11/13	Expansión
79	Santander, BBVA, CaixaBank y Bankinter disparan su colchón de liquidez para combatir el Covid			Santander	2020/11/13	Expansión
80	Santander, BBVA, CaixaBank y Bankinter disparan su colchón de liquidez para combatir el Covid			Bankinter	2020/11/13	Expansión
81	Santander, BBVA, CaixaBank y Bankinter disparan su colchón de liquidez para combatir el Covid			Caixa	2020/11/13	Expansión
82	Aena saca a concurso por 155 millones los controles del Covid-19			Aena	2020/11/13	Expansión
83	Santander entra en el capital del bróker digital DriveWealth			Santander	2020/11/13	Expansión
84	Naturgy: tras State Grid vienen ventas masivas			Naturgy	2020/11/14	Expansión
85	Así será el ERE de Santander			Santander	2020/11/14	Expansión
86	ACS, incluida de nuevo en el Dow Jones Sustainability Index			ACS	2020/11/14	Expansión
87	BBVA lidera la representación española en el Dow Jones Sustainability Index			BBVA	2020/11/14	Expansión
88	BBVA pisará el freno en sus emisiones: le bastan 4.500 millones para cumplir su plan de 2021			BBVA	2020/11/14	Expansión
89	Santander, BBVA, CaixaBank y Bankia, en la elite digital en Europa			BBVA	2020/11/16	Expansión
90	Santander, BBVA, CaixaBank y Bankia, en la elite digital en Europa			Santander	2020/11/16	Expansión
91	Santander, BBVA, CaixaBank y Bankia, en la elite digital en Europa			Caixa	2020/11/16	Expansión
92	El golpe a las Sicav obliga a BBVA y Santander a revisar su banca privada			BBVA	2020/11/16	Expansión
93	El golpe a las Sicav obliga a BBVA y Santander a revisar su banca privada			Santander	2020/11/16	Expansión
94	BBVA vende su filial en EEUU a PNC por 9.700 millones de euros			BBVA	2020/11/16	Expansión
95	Carlos Torres valora la venta de la filial de BBVA a PNC			BBVA	2020/11/16	Expansión
96	Iberdrola se alía a SSE y National Grid para una macroinversión de 3.300 millones en Reino Unido			Iberdrola	2020/11/16	Expansión
97	BBVA tiene la llave para comprarse Sabadell			BBVA	2020/11/16	Expansión
98	BBVA tiene la llave para comprarse Sabadell			Sabadell	2020/11/16	Expansión
99	Glovo se alía con Naturgy para ofrecer reparaciones de electrodomésticos y cerrajería			Naturgy	2020/11/16	Expansión
100	Solaria gana 25,13 millones en los nueve primeros meses del año, un 79% más			Solaria	2020/11/16	Expansión
101	BBVA y Sabadell negocian su fusión			BBVA	2020/11/16	Expansión
102	BBVA y Sabadell negocian su fusión			Sabadell	2020/11/16	Expansión
103	El gigante financiero británico Schroders se suma a Norges y Santander en el desembarco en Soltec			Santander	2020/11/16	Expansión
104	La CNMV pide explicaciones a Repsol por su multimillonaria reconversión ecológica			Repsol	2020/11/16	Expansión
105	Santander compra el negocio de pagos para comercios de Wirecard por unos 100 millones			Santander	2020/11/16	Expansión
106	Torres Vila valora la venta de la filial de BBVA a PNC			BBVA	2020/11/17	Expansión
107	BBVA y Sabadell, al rebufo de CaixaBank			BBVA	2020/11/17	Expansión
108	BBVA y Sabadell, al rebufo de CaixaBank			Sabadell	2020/11/17	Expansión
109	BBVA y Sabadell, al rebufo de CaixaBank			Caixa	2020/11/17	Expansión
110	El proyecto de BBVA pasa por aumentar su peso en España			BBVA	2020/11/17	Expansión
111	Así sería el banco resultante de la fusión BBVA-Sabadell			BBVA	2020/11/17	Expansión
112	Así sería el banco resultante de la fusión BBVA-Sabadell			Sabadell	2020/11/17	Expansión
113	ACS gana un contrato de construcción en una base militar australiana			ACS	2020/11/17	Expansión
114	BBVA logra el capital para comerse a Sabadell			BBVA	2020/11/17	Expansión
115	BBVA logra el capital para comerse a Sabadell			Sabadell	2020/11/17	Expansión
116	Santander estudia la salida a Bolsa de su negocio de pagos para comercios Getnet			Santander	2020/11/17	Expansión
117	Guardiola: "Las próximas semanas sabremos el resultado de la negociación entre BBVA y Sabadell"			BBVA	2020/11/17	Expansión
118	Guardiola: "Las próximas semanas sabremos el resultado de la negociación entre BBVA y Sabadell"			Sabadell	2020/11/17	Expansión
119	Telefónica recibirá unos 39 millones de su filial brasileña como pago por los intereses de capital propio			Telefónica	2020/11/17	Expansión
120	S&P confirma el rating 'BB' para IAG con perspectiva negativa por la pandemia			IAG	2020/11/17	Expansión
121	Naturgy ofrece gratis el servicio de reparación a los profesionales de educación y centros docentes			Naturgy	2020/11/17	Expansión
122	Sabadell y los sindicatos suspenden temporalmente las negociaciones del ERE para reflexionar sobre la fusión			Sabadell	2020/11/17	Expansión
123	TSB, el verso suelto de BBVA-Sabadell			BBVA	2020/11/17	Expansión
124	TSB, el verso suelto de BBVA-Sabadell			Sabadell	2020/11/17	Expansión
125	BBVA y Sabadell, suma de fuerzas			BBVA	2020/11/17	Expansión
126	BBVA y Sabadell, suma de fuerzas			Sabadell	2020/11/17	Expansión
127	Cuáles son los mayores accionistas en BBVA-Sabadell?			BBVA	2020/11/17	Expansión
128	Cuáles son los mayores accionistas en BBVA-Sabadell?			Sabadell	2020/11/17	Expansión
129	Sabadell, el banco centenario de la burguesía catalana			Sabadell	2020/11/18	Expansión
130	BBVA-Sabadell: prima de hasta el 30%			BBVA	2020/11/18	Expansión
131	BBVA-Sabadell: prima de hasta el 30%			Sabadell	2020/11/18	Expansión
132	Mapfre y Macquarie compran la antigua sede de BBC Radio			Mapfre	2020/11/18	Expansión
133	Genç: "BBVA no está forzado y hará la fusión con Sabadell si crea valor"			BBVA	2020/11/18	Expansión
134	Genç: "BBVA no está forzado y hará la fusión con Sabadell si crea valor"			Sabadell	2020/11/18	Expansión
135	Siemens Gamesa recupera a Marc Becker para liderar su negocio de eólica marina			Gamesa	2020/11/18	Expansión
136	Santander lanza un fondo de inversión alternativa para financiar pymes y aporta 300 millones			Santander	2020/11/18	Expansión
137	Iberdrola busca sitio para una gran fábrica española de máquinas de hidrógeno			Iberdrola	2020/11/18	Expansión
138	BBVA-Sabadell se hace fuerte financieramente en El Corte Inglés, Telefónica, Aena y los Matutes			Aena	2020/11/18	Expansión
139	BBVA-Sabadell se hace fuerte financieramente en El Corte Inglés, Telefónica, Aena y los Matutes			BBVA	2020/11/18	Expansión
140	BBVA-Sabadell se hace fuerte financieramente en El Corte Inglés, Telefónica, Aena y los Matutes			Sabadell	2020/11/18	Expansión
141	BBVA-Sabadell se hace fuerte financieramente en El Corte Inglés, Telefónica, Aena y los Matutes			Telefónica	2020/11/18	Expansión
142	CaixaBank y BBVA: el 45% del sector asegurador en la diana			BBVA	2020/11/18	Expansión
143	CaixaBank y BBVA: el 45% del sector asegurador en la diana			Caixa	2020/11/18	Expansión
144	BBVA amarra el negocio con PepsiCo, AT&T o BlackRock pese a su salida de Estados Unidos			BBVA	2020/11/19	Expansión
145	Así convirtió Bill Demchak PNC, comprador de la filial de BBVA en EEUU, en una poderosa entidad bancaria			BBVA	2020/11/19	Expansión
146	Onur Genç: "La operación con Sabadell competirá con otras opciones"			Sabadell	2020/11/19	Expansión
147	Endesa o Naturgy necesitan permiso del Gobierno para compras en España			Endesa	2020/11/19	Expansión
148	Endesa o Naturgy necesitan permiso del Gobierno para compras en España			Naturgy	2020/11/19	Expansión
149	BBVA: una historia de lucha por el poder			BBVA	2020/11/19	Expansión
150	Novartis se alía a Acciona, EDP y Endesa para hacer de España su gran proveedor de electricidad verde para toda Europa			Acciona	2020/11/19	Expansión

ANEXO 2: OBJETIVOS DE DESARROLLO SOSTENIBLE

Los objetivos de desarrollo sostenible son un conjunto de metas que buscan erradicar la pobreza, proteger el planeta y asegurar la prosperidad. Con ello se pretende crear una agenda de desarrollo sostenible y, para alcanzarla con éxito, todo el mundo debe involucrarse y colaborar. Con este TFM se pretende crear un sistema para la inversión sostenible y diversificada ayudando a la toma de decisiones de los individuos. Con este tipo de inversión se fomenta el crecimiento económico y sostenido que es capaz de impulsar el progreso ayudando a crear empleos y, mejorando el nivel de vida de las personas.

Por lo tanto, este trabajo de fin de máster se enmarca en el entorno del octavo objetivo de desarrollo sostenible: “Trabajo decente y crecimiento económico”.



Ilustración 47. Objetivo desarrollo sostenible n°8. Fuente: (Objetivos de desarrollo sostenible, n.d.)