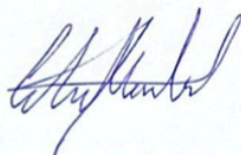


Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título **Análisis de la correlación entre movimiento bursátiles y las emociones en Twitter en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el curso académico 2019/2020** es de mi autoría, original e inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.

Fdo.: Francisco José Ortuño Montel Fecha: 2020/06/05



Autorizada la entrega del proyecto
EL DIRECTOR DEL PROYECTO



Fdo.: Atilano Fernández-Pacheco Sánchez-Migallón Fecha: 2020/06/05



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

Análisis de la correlación entre movimientos bursátiles y las emociones en Twitter

Autor: Francisco José Ortuño Montel

Director: Atilano Fernández-Pacheco Sánchez-Migallón

Madrid

AGRADEMIENTOS

Mi primer agradecimiento es para el Dr. Atilano Fernández-Pacheco no solo por dirigir el proyecto, si no, además, por su apoyo durante muchos años a lo largo de la carrera

A mi familia, a mi madre, a mi padre y a mis hermanos Jacobo y Tomas por su amor, paciencia y comprensión.

Por último, al Dr. Rafael Palacios, por su apoyo y ánimo para finalizar mis estudios.

ANÁLISIS DE LA CORRELACIÓN ENTRE MOVIMIENTOS BURSÁTILES Y LAS EMOCIONES EN TWITTER

Autor: Ortuño Montel, Francisco José.

Director: Fernández-Pacheco Sánchez-Migallón, Atilano.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

En este proyecto hemos estudiado la relación entre los tweets que se publican en Twitter sobre una empresa y la evolución del valor de la acción de dicha empresa. Para nuestro estudio hemos elegido la empresa Morgan Stanley. Los meses en los que hemos realizado el estudio han sido febrero, marzo y abril de 2020. Encontramos por una parte que los tweets que más afectaron a su cotización en bolsa en febrero fueron los de polaridad neutra con una correlación de -0.5. Por otra parte, durante el periodo de la pandemia COVID-19 (marzo y abril), resultó que ninguno de los diferentes tipos de tweets tuvo una correlación mayor a ± 0.11 con las acciones de la bolsa.

1. Introducción

Determinar con antelación la evolución del valor de una acción siempre ha sido un interrogante para operadores profesionales y particulares, exceptuando aquellos que tienen información privilegiada. Para poder invertir con cierta seguridad se han propuesto diferentes tipos de análisis, como puede ser el macroeconómico, técnico y fundamental. El objetivo de nuestro proyecto es dar un punto más de seguridad a la hora de invertir.

2. Definición del Proyecto

El desarrollo de este proyecto se ha dividido en dos fases, una de almacenamiento de datos y la otra de análisis y resultados. La parte de almacenamiento la hemos realizado con scripts de Python y bases de datos MongoDB. Para la fase de análisis de datos y resultados hemos usado la aplicación EXCEL de Microsoft y Python para el cálculo de correlaciones.

3. Descripción del modelo/sistema/herramienta

Fase de almacenamiento:

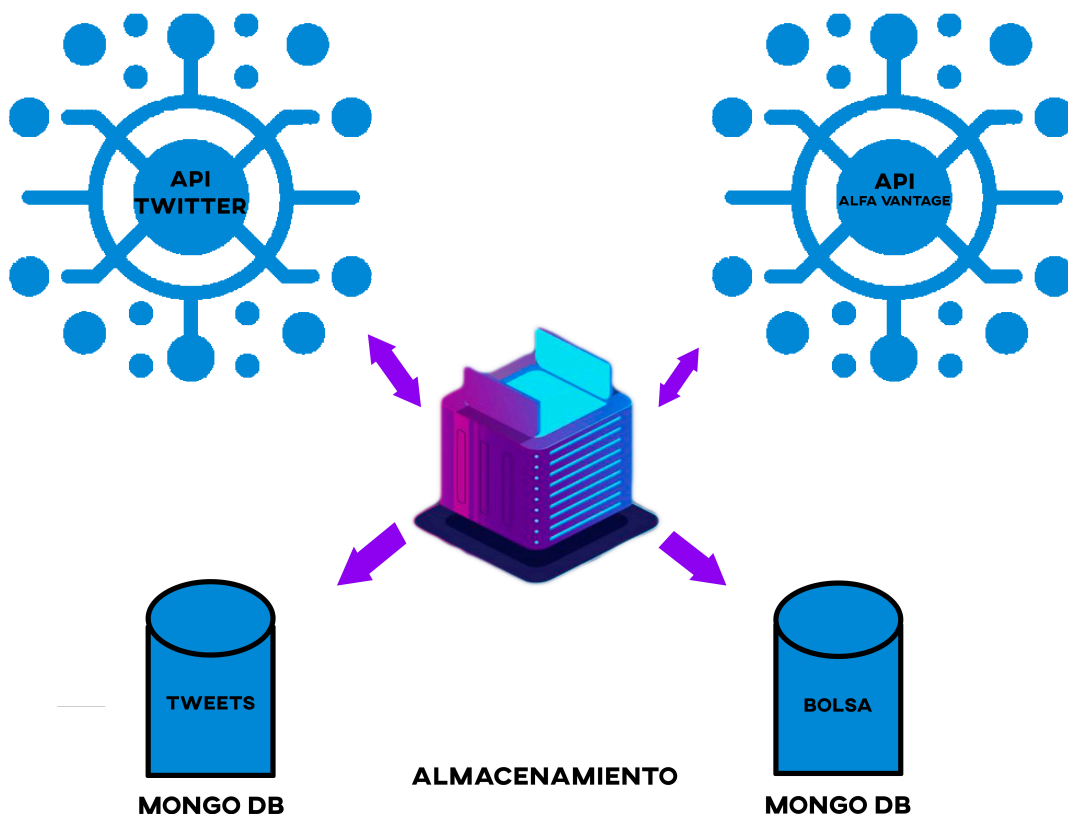


Figura 1 Fase de almacenamiento de datos

Los datos que hemos obtenido de Twitter son los tweets. Los hemos recopilado a través de las APIs de Twitter y los datos de la acción de bolsa a través de la API de Alpha Vantage. Estos datos los hemos almacenado en dos bases de datos distintas de MongoDB.

Fase de análisis y resultados:

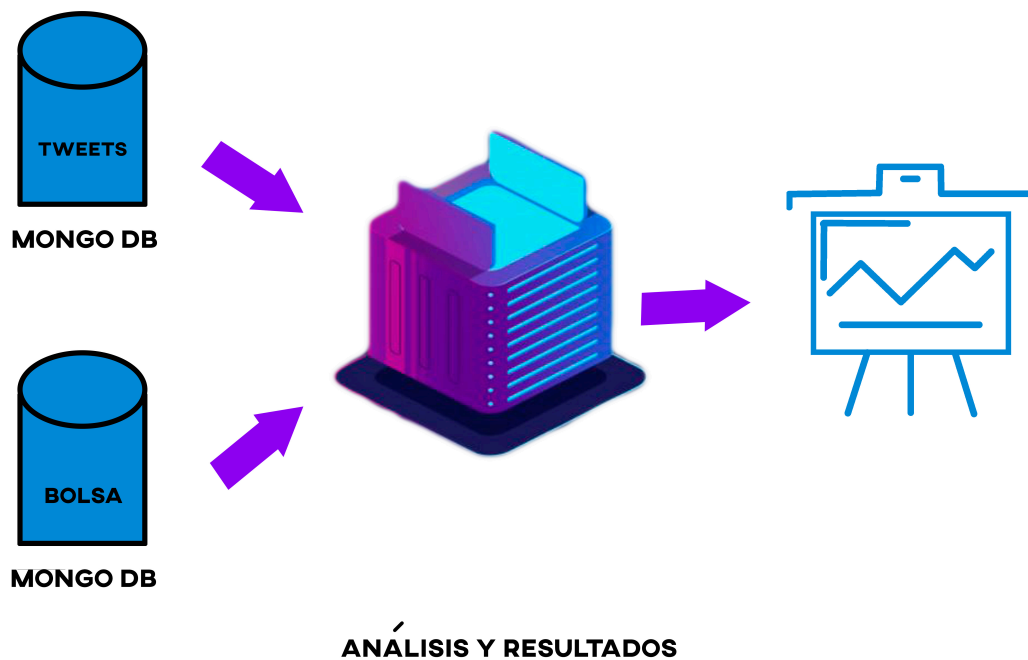


Figura 2 Fase de análisis y resultados

Hemos recuperado los datos con un script de Python, los hemos pasado a un EXCEL y con Python hemos calculado las correlaciones de los diferentes tipos de tweets con el valor de acción en Bolsa.

4. Resultados

Hemos analizado febrero, marzo y abril independientemente y combinaciones de dichos datos en diferentes periodos.

En febrero mostramos que la mayor correlación fue con los tweets con polaridad neutra, obteniendo un coeficiente de correlación indirecto de $-0,5$, siendo la línea representada en la figura 3, la línea de tendencia

Relación de tweets con polaridad neutra y valor de bolsa en febrero:

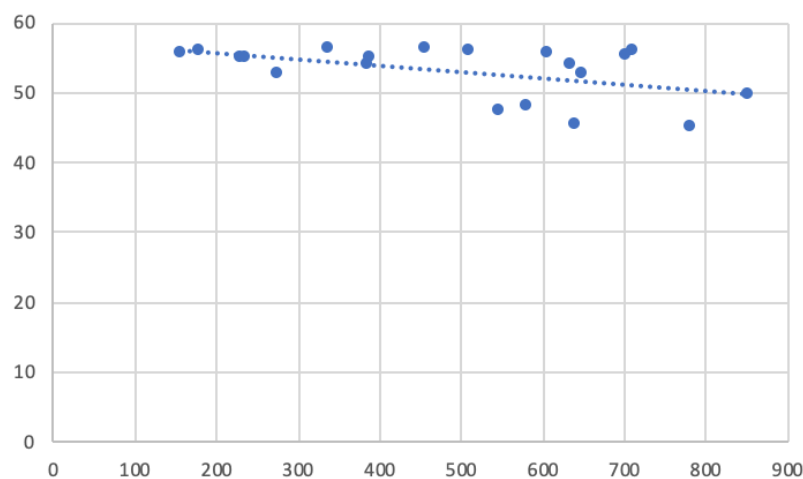


Figura 3 Correlación Tweets polaridad neutra y valor acción febrero

En marzo encontramos que la máxima correlación resultó ser con los tweets con polaridad negativa con una correlación de $-0,27$.

Relación de tweets con polaridad negativa y valor de bolsa en marzo:

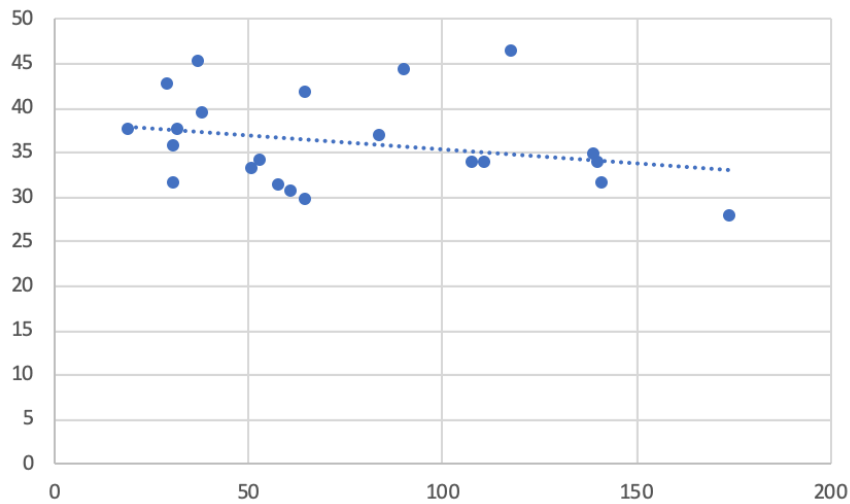


Figura 4 Correlación Tweets polaridad negativa y valor acción marzo

En abril encontramos que la máxima correlación resultó ser con los tweets con polaridad negativa con una correlación de $-0,1432$.

Relación de tweets con polaridad negativa y valor de bolsa en abril:

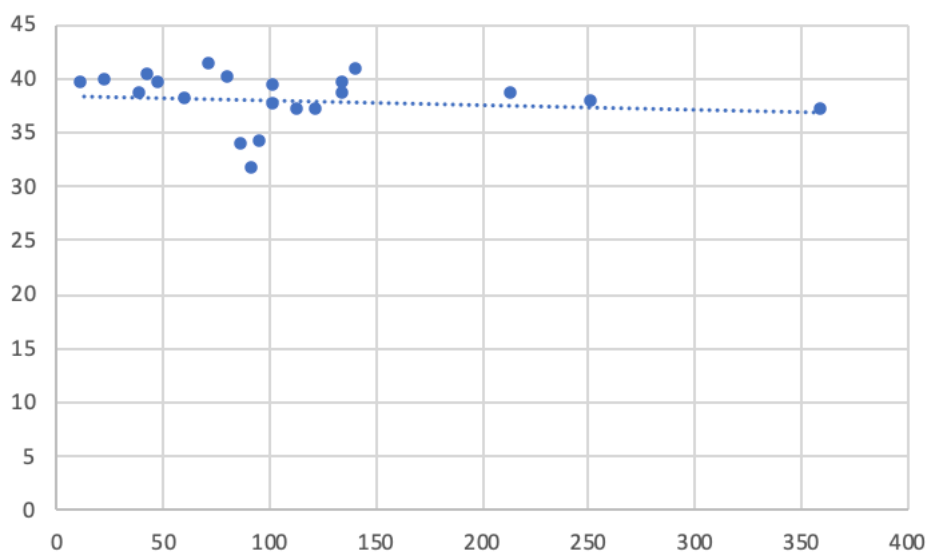


Figura 5 Correlación Tweets polaridad negativa y valor de acción abril

El conjunto de marzo y abril encontramos que apenas hubo correlación entre cualquier tipo de tweet y el valor de cotización.

Tabla 1 Correlación de marzo y abril

	Positivo	Negativo	Neutro	Valor de bolsa
Positivo	1,0000	-0,1507	-0,7777	0,0859
Negativo	-0,1507	1,0000	-0,2357	-0,1004
Neutro	-0,7777	-0,2357	1,0000	-0,0432
Valor de bolsa	0,0859	-0,1004	-0,0433	1,000

5. Conclusiones

Nuestros resultados objetivan que existe una relación entre los tweets que se publican sobre una empresa y su cotización en bolsa durante periodos “normales”. Con ello, sabiendo que la relación entre los tweets de polaridad neutra tiene una relación negativa con su cotización, asumimos que cuantos menos tweets neutros se publiquen, hay más posibilidades de obtener beneficios con su inversión.

También hemos podido comprobar que durante periodos excepcionales, como lo ha sido la pandemia COVID-19, ya no existe esa relación, en cuyo caso los tweets no aportarían información útil para invertir en el mercado de valores en términos mensuales.

ANALYSIS OF THE CORRELATION BETWEEN STOCK MARKET MOVEMENTS AND EMOTIONS IN TWITTER

Author: Ortuño Montel, Francisco José.

Supervisor: Fernández-Pacheco Sánchez-Migallón, Atilano.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

In the herein project we carried out a study to foresee what is the relationship between the tweets that are published on Twitter about a company and the market evolution of its shares. For our study, we chose Morgan Stanley. We found that the tweets that most affected their stock evolution along February were those with neutral polarity with a correlation of -0.5; On the other hand, during the COVID-19 pandemic period (March and April) it turned out that none of the different types of tweets had a greater than ± 0.11 correlation with stocks.

Project summary

Forecast the evolution of a stock exchange has always been a major question for stock traders and particulars, except those with privileged information. In order to improve investment security, different types of analysis have been proposed, such as macroeconomic, technical and fundamental. With our project we herein intend to further upgrade the security of stock investments.

1. Project Definition

To develop this project, we divided it into two phases, one of data storage then followed by their analysis and results. We carried out the storage part with Python scripts and MongoDB databases and for the analysis of data and results we used the Microsoft EXCEL application.

2. Description of the model

Storage phase:

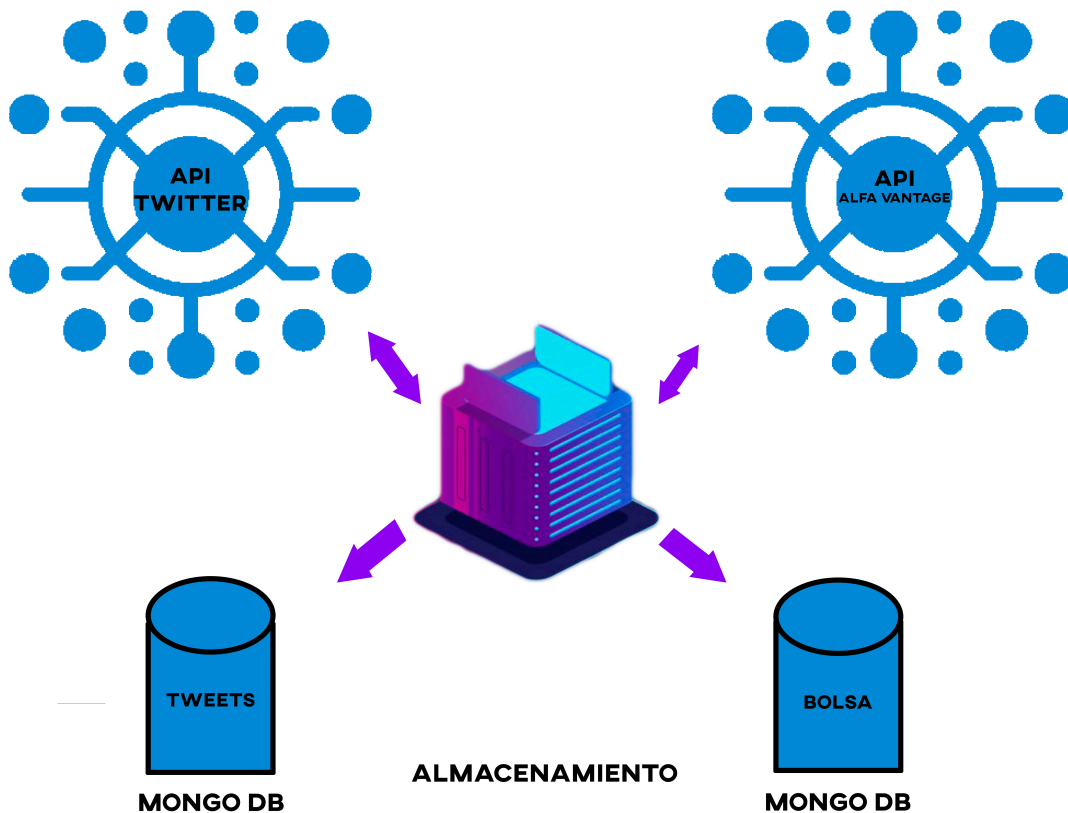


Figura 6 Data storage

We have extracted the tweet data from the Twitter API and the data of the stock exchange, from the Alpha Vantage API; these data have been stored in two different databases of MongoDB.

Analysis and results phase:



Figura 7 Analysis and results

We recovered the data with Python script and transferred them to an EXCEL. With the functions of the latter we performed the analysis and shown the results

3. Results

We analyzed February, March and April independently and combinations of these data in different periods.

In February we showed that the highest correlation was with tweets with neutral polarity, with an indirect correlation coefficient of -0.5, with the line represented in the figure being the trend line.

Ratio of tweets with neutral polarity and stock market value in February:

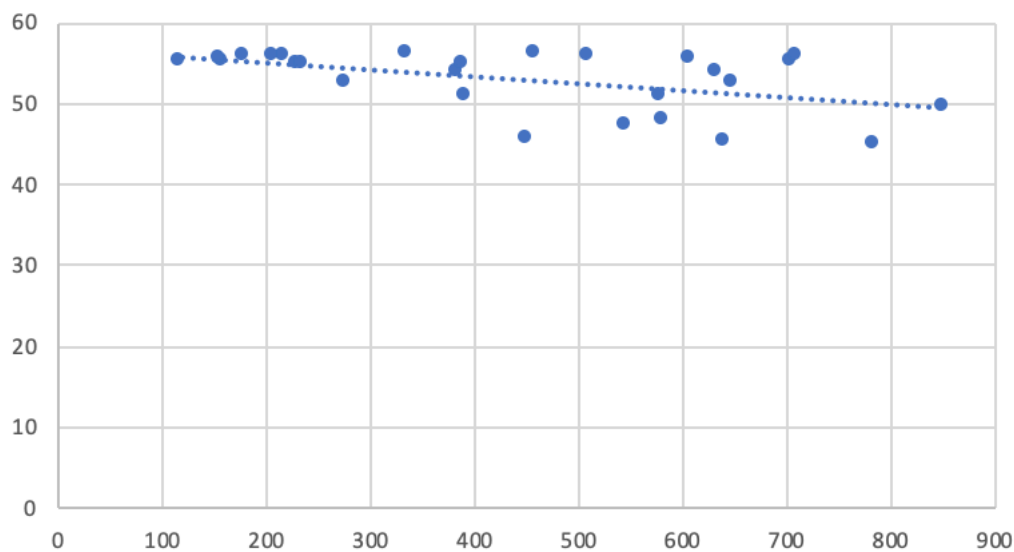


Figura 8 Correlation Neutral tweets and stock value Februar

In March we showed that the highest correlation was with tweets with negative polarity, we obtained a -0.27 indirect correlation coefficient; the line represented in the figure was the trend line.

Ratio of tweets with negative polarity and stock market value in March:

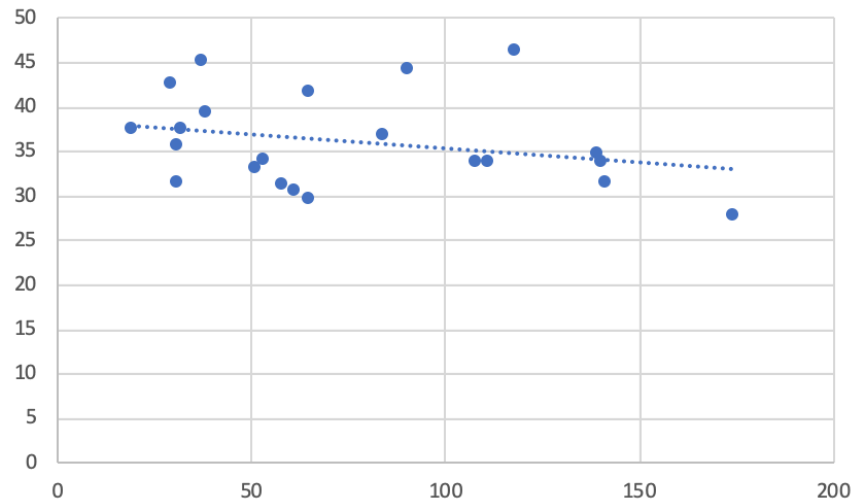


Figura 9 Correlation negative tweets and stock value March

In April we showed that the highest correlation was with tweets with negative polarity, with a -0.1432 indirect correlation coefficient; the line represented in the figure being the trend line.

Ratio of tweets with negative polarity and stock market value in April:

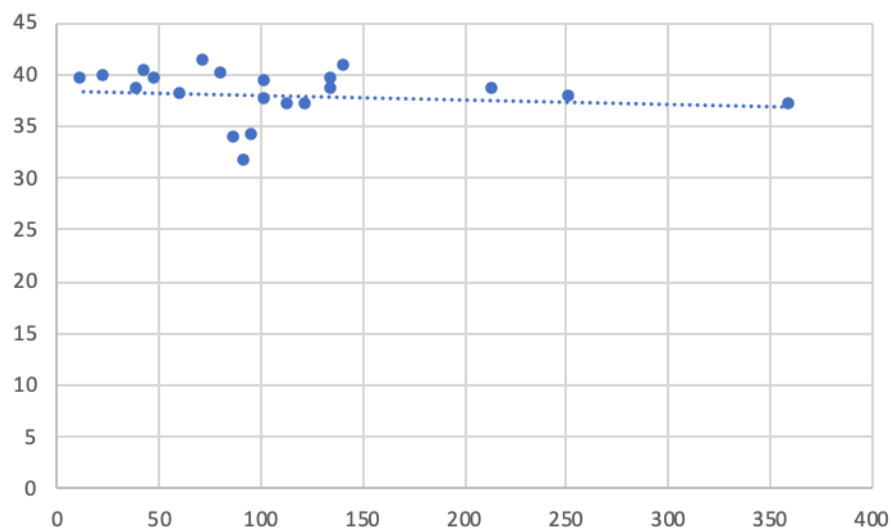


Figura 10 Correlation negative tweets and stock value April

There were only subtle correlations among any type of tweet and the stock market value when considering the combined April and March period.

Tabla 2 Correlation of march and april with the value of stock

	Positivo	Negativo	Neutro	Valor de bolsa
Positivo	1,0000	-0,1507	-0,7777	0,0859
Negativo	-0,1507	1,0000	-0,2357	-0,1004
Neutro	-0,7777	-0,2357	1,0000	-0,0432
Valor de bolsa	0,0859	-0,1004	-0,0433	1,000

4. Conclusion

We herein show that, during a “normal” period there is a relationship between the tweets that are published on a company and its stock value, by demonstrating the association between tweets of neutral polarity with a negative evolution of stocks; that is: the less neutral tweets are published on a company, the more chances of optimizing money investment.

We have also been able to verify that during exceptional periods such as the COVID-19 pandemic, that relationship no longer exists, in which case, the tweets would not provide useful information to invest in the stock market in monthly terms.

Índice

CAPÍTULO 1.	INTRODUCCIÓN	20
CAPÍTULO 2.	DESCRIPCIÓN DE LAS TECNOLOGÍAS	21
2.1	PHP	21
2.2	PYTHON	22
2.3	ANACONDA	23
2.4	MONGODB.....	24
2.5	ALPHA VANTAGE	25
2.6	EXCEL	26
CAPÍTULO 3.	ESTADO DE LA CUESTIÓN.....	27
3.1	MERCADO DE VALORES.....	27
3.2	TWITTER	29
3.3	ESTUDIOS QUE RELACIONAN LA BOLSA Y TWITTER	30
CAPÍTULO 4.	DEFINICIÓN DEL TRABAJO.....	31
4.1	MOTIVACIÓN	31
4.2	OBJETIVOS	32
4.3	METODOLOGÍA	33
4.4	PLANIFICACIÓN Y ESTIMACIÓN ECONÓMICA	34
CAPÍTULO 5.	SISTEMA/MODELO DESARROLLADO	35
5.1	ARQUITECTURA DEL SISTEMA	35
5.2	API DE TWITTER.....	38
5.2.1	<i>Conexión a la api</i>	<i>39</i>
5.2.2	<i>Descarga de datos de Twitter.....</i>	<i>40</i>
5.2.3	<i>Definir nuestro patrón de búsqueda.....</i>	<i>41</i>
5.2.4	<i>Determinar la polaridad de un tweet</i>	<i>41</i>
5.3	APIS DE BOLSA.....	43
5.3.1	<i>Conexión a Quandl</i>	<i>43</i>
5.3.2	<i>Conexión a SimFin</i>	<i>44</i>
5.3.3	<i>Conexión a Alpha Vantage</i>	<i>45</i>

ÍNDICE DE LA MEMORIA

5.4	BASES DE DATOS.....	47
5.4.1	Almacenamiento de tweets.....	47
5.4.2	Almacenamiento de datos de bolsa	48
5.4.3	Recuperación de datos	49
CAPÍTULO 6.	ANÁLISIS DE RESULTADOS	53
6.1	CORRELACIONES DE DATOS EN FEBRERO	53
6.2	CORRELACIONES DE DATOS EN MARZO	55
6.3	CORRELACIONES DE DATOS EN ABRIL.....	57
6.4	CORRELACIONES DE DISTINTOS PERIODOS	59
6.5	CORRELACIONES EN TIEMPO REAL	61
CAPÍTULO 7.	CONCLUSIONES	65
CAPÍTULO 8.	TRABAJOS FUTUROS.....	68
CAPÍTULO 9.	BIBLIOGRAFÍA.....	69
CAPÍTULO 10.	ANEXO A GUÍA DE INSTALACIÓN	72
10.1	MONGODB	72
10.2	ANACONDA	72
CAPÍTULO 11.	ANEXO B MANUAL DE USUARIO	73
11.1	MONGODB	73
11.2	API DE TWITTER.....	78
11.3	DOMINIO WEB.....	81
11.4	ANACONDA	83
11.5	ALPHA VANTAGE	85
CAPÍTULO 12.	ANEXO C ODS.....	87

ÍNDICE DE FIGURAS

Figura 1 Fase de almacenamiento de datos	V
Figura 2 Fase de análisis y resultados	VI
Figura 3 Correlación Tweets polaridad neutra y valor acción febrero.....	VII
Figura 4 Correlación Tweets polaridad negativa y valor acción marzo.....	VIII
Figura 5 Correlación Tweets polaridad negativa y valor de acción abril.....	VIII
Figura 6 Data storage	XI
Figura 7 Analysis and results	XII
Figura 8 Correlation Neutral tweets and stock value Februar.....	XIII
Figura 9 Correlation negative tweets and stock value March	XIV
Figura 10 Correlation negative tweets and stock value April	XIV
Figura 11 Logo de PHP	21
Figura 12 Logo de Python.....	22
Figura 13 Logo de Anaconda.....	23
Figura 14 Logo MongoDB.....	24
Figura 15 Logo Alpha Vantage.....	25
Figura 16 Logo de EXCEL	26
Figura 17 Diagrama de GRANT	33
Figura 18 Almacenamiento de datos.....	35
Figura 19 Análisis y resultados	38
Figura 20 Respuesta Twitter formato JSON	40
Figura 21 Ejemplo Quandl.....	43
Figura 22 Ejemplo SimFin.....	44
Figura 23 Ejemplo Alpha Vantage.....	46
Figura 24 febrero correlación tweets positivos y valor de bolsa.....	54
Figura 25 febrero correlación tweets negativos y valor de bolsa.....	54
Figura 26 febrero correlación tweets neutros y valor de bolsa	55
Figura 27 marzo correlación tweets positivos y valor de bolsa	56
Figura 28 marzo correlación tweets negativos y valor de bolsa	56
Figura 29 marzo correlación tweets neutros y valor de bolsa	57

ÍNDICE DE FIGURAS

Figura 30 abril correlación tweets positivos y valor de bolsa	58
Figura 31 abril correlación tweets negativos y valor de bolsa	58
Figura 32 abril correlación tweets neutros y valor de bolsa.....	59
Figura 33 Instalación Anaconda y Python	72
Figura 34 Conectandose a MongoDB	73
Figura 35 Mostrar bases de datos.....	73
Figura 36 Ver que en base de datos estamos.....	74
Figura 37 Crear base de datos	74
Figura 38 Agregar colección.....	74
Figura 39 Mostrar colecciones	74
Figura 40 Eliminar base de datos	75
Figura 41 Crear colección	75
Figura 42 Eliminar colección.....	76
Figura 43 Insertar colección.....	76
Figura 44 Buscar elementos de una colección	77
Figura 45 Conectarse a MongoDB.....	78
Figura 46 Ejemplo JSON	79
Figura 47 Ejemplo ARRAY.....	80
Figura 48 Ejemplo modificar formato.....	81
Figura 49 000webhost página inicio	82
Figura 50 000webhost página de operaciones.....	82
Figura 51 000webhost página de archivos	83
Figura 52 Anaconda página principal	84
Figura 53 Página principal de SPYDER	85
Figura 54 Página principal Alpha Vantage	86

Capítulo 1. INTRODUCCIÓN

En nuestros días, hay varios tipos de tecnologías que están en auge, uno de los ejemplos más importantes es el uso de base de datos no relacionales. Cuando hablamos de este tipo, queremos decir que son bases de datos que no poseen una estructura. Son denominadas, NoSQL. Estas bases de datos crecieron cuando las principales redes sociales tenían que enfrentarse al tratamiento de datos que las tradicionales bases de datos relacionales no solucionaban. La ventaja principal de esta tecnología es que puede manejar enormes cantidades de datos sin generar cuellos de botellas.

Que mejor que estudiando un Grado en Ingeniería en Tecnologías de la Telecomunicación, realicemos un TFG referente a esta tecnología. Por otra parte, esta podría ser una herramienta de utilidad para empresas que están usando este tipo de tecnología para análisis de datos.

Otro de los aspectos que vamos a estudiar en nuestro TFG, son las redes sociales, en concreto vamos a estudiar Twitter. Este es un servicio de comunicación bidireccional con el que se puede compartir información de diverso tipo de una forma rápida, sencilla y gratuita, es decir, es una red de *microblogging*. En este ámbito, es de las más populares en la actualidad.

Otra de las partes de nuestro TFG, va a estar relacionada con el mercado de valores. La bolsa es un mercado, un lugar donde compradores y vendedores se reúnen para intercambiar un bien, las acciones, por un precio acordado entre ellos. Una de sus funciones más importantes es que permite a las empresas capitalizarse para desarrollar sus actividades.

Capítulo 2. DESCRIPCIÓN DE LAS TECNOLOGÍAS

2.1 PHP



Figura 11 Logo de PHP

PHP es un lenguaje de programación de código abierto, pensado para el desarrollo web, es lenguaje que puede ir embebido en HTML; muchas de las páginas web actuales están desarrolladas en PHP por su facilidad de uso. Se utiliza para producir páginas web dinámicas, es decir, cuyo contenido puede modificarse, a diferencia de las estáticas, normalmente realizadas con HTML y CSS en las que no pueden producirse cambios.

El funcionamiento de PHP es sencillo, el usuario realiza una petición, se procesa y prepara y se envía una respuesta. Normalmente esta información está guardada en una base de datos.

Lo que distingue a PHP de Javascript es que el código es ejecutado en el servidor, generando un HTML y enviándolo al cliente como si se tratara de una página web estática. Además, en PHP se puede procesar información de formularios, generar páginas con contenidos dinámicos, o enviar y recibir cookies, crear imágenes a partir de datos, autogenerar archivos y almacenar en el sistema. También puede interactuar con otros servidores usando diversos protocolos.

2.2 PYTHON



Figura 12 Logo de Python

Python es un lenguaje de programación presente en multitud de aplicaciones y sistemas operativos, podemos encontrar en aplicaciones de IOS, Android, Linux o Windows. Se puede encontrar en estos diversos sistemas operativos debido a que cuenta con una curva de aprendizaje moderada.

Es un lenguaje de programación versátil, multiplataforma y multiparadigma que se destaca por su código legible y limpio. Como es un lenguaje de código abierto tiene mucho éxito; además hace que muchas personas se inicien en la programación con este lenguaje.

Por otra parte, es adecuado para trabajar con grandes cantidades de datos ya que favorece su extracción y procesamiento, por eso se usa este lenguaje para técnicas Big Data.

Las ventajas de programar en Python son las siguientes: simplificado, rápido, flexible, ordenado, limpio, portable y con una gran comunidad de usuarios.

2.3 ANACONDA



Figura 13 Logo de Anaconda

Anaconda es una plataforma de programación fundamentalmente diseñada para desarrollar lenguaje de Python, nos da muchas funcionalidades que permitirán que desarrollemos aplicaciones de una manera más eficiente, rápida y sencilla.

Es una Suite de código abierto que posee unas aplicaciones, librerías y conceptos diseñados para el desarrollo de la ciencia de datos con Python. Funciona como un gestor de entorno, un gestor que posee una colección de más de 720 paquetes de código abierto.

Se divide en cuatro sectores, Anaconda Navigator, Anaconda Project, las librerías de ciencia de datos y Conda. Se puede usar tanto dentro de la aplicación como en la consola del terminal.

Las características principales de esta aplicación son las siguiente: libre, de código abierto, documentación detallada, multiplataforma, usa diversos IDE como Jupiter, Sypder y RSstudio, permite visualizar datos, tiene una interfaz bastante sencilla, elimina problemas de dependencia de paquetes y control de versiones y facilita la escritura de algoritmos complejos paralelos para la ejecución de tareas.

2.4 MONGODB



Figura 14 Logo MongoDB

MongoDB es una base de datos de documentos que ofrece una gran escalabilidad y flexibilidad, y un modelo de consultas e indexación avanzado.

Es una base de datos NoSQL, es decir que, en lugar de guardar los datos en registros, se guardan en BSON que es una representación del JSON, esto conlleva a que no se necesite un esquema, si no se guardan los datos en una colección.

MongoDB esta escrito en c++, aunque las consultas se hacen pasando objetos en formato JSON y se almacenan en BSON. Para acceder a él, se hace desde la consola, que está construida con JavaScript. También se puede usar otros lenguajes de programación para entrar en MongoDB usando diferentes drivers.

Es recomendable usar MongoDB para almacenar datos semiestructurados y que requieran escalabilidad porque tiene diferentes opciones de replicación y *sharding*, que son sencillas de configurar y podemos conseguir un sistema que escale horizontalmente sin demasiados problemas.

2.5 *ALPHA VANTAGE*



Figura 15 Logo Alpha Vantage

Alpha Vantage son una serie de APIs que ofrecen datos gratuitos accesibles de la bolsa, se puede acceder a estos datos usando Python o R.

La descarga de datos es gratuita, pero para acceder a los datos, primero tienes que darte de alta, para obtener una clave de acceso personal.

Podemos obtener datos al final del día, semanales o mensuales y para estos *time frames* podemos obtener los datos ajustados o sin ajustar. Además, también podemos coger datos en tiempo real.

También incluyen la posibilidad de descargar directamente los valores de muchos indicadores; estos pueden ser simples como las medias móviles o también pueden ser indicadores de cálculo más complejo.

2.6 EXCEL



Figura 16 Logo de EXCEL

Excel es un programa de Microsoft del tipo de hoja de cálculo que permite realizar operaciones con números organizados en cuadrículas. Es útil para realizar cualquier tipo de contabilidad.

La hoja de cálculo de Excel es un programa capaz de trabajar con números de forma sencilla e intuitiva. Para ello usa una cuadrícula donde, en cada celda, se pueden introducir números, letras, gráficos y fórmulas.

Es útil cuando tenemos que hacer cálculos de una cantidad de datos importante, además si modificamos cualquier celda individual, el resultado global cambia automáticamente.

Excel se puede utilizar para multitud de funciones, tanto en el plano particular como en el plano profesional, como, por ejemplo, llevar a cabo contabilidades de manera sencilla.

Capítulo 3. ESTADO DE LA CUESTIÓN

3.1 MERCADO DE VALORES

El mercado de valores se ha convertido en los últimos tiempos en una de las fórmulas de inversión más populares a la hora de destinar los ahorros.

En el siglo XIV los comerciantes de Brujas se reunían para sus negocios en un edificio en cuyo escudo se apreciaban tres bolsas de piel. Los edificios donde se realizaban transacciones comerciales se empezaron a llamar “bolsa” en toda esa región. En 1460 en Amberes se creó la primera institución bursátil de la modernidad, donde mercaderes de diferentes lugares cotizaban con sus distintos objetos de valor. El primer lugar donde se pusieron acciones a la venta fue en a bolsa de valores de Ámsterdam en 1602, donde la Compañía de la Indias Orientales, que fue la primera sociedad anónima de la historia, empezó a vender participaciones en sus negocios para financiar su expansión comercial. De esta forma podían captar grandes cantidades de dinero a través de pequeñas inversiones de los ahorros de un gran número de personas a cambio de acciones de la empresa.

La institución de bolsa termino de desarrollarse con la revolución industrial y la expansión del capitalismo. Así, hoy en día, las bolsas de valores son organizaciones que proveen un mercado donde las compañías ofrecen sus acciones y los inversores, a través de intermediarios o incluso directamente, pueden comprarlas.

Han surgido hasta la fecha muchas técnicas de análisis para este mercado bursátil, como el análisis de gráficos en los que hay que detectar los soportes y resistencias más relevantes para ver que posibilidades de *trading* existen realmente en el valor que analizas, las velas japonesas, en que las que hay que identificar a que grupo pertenecen, es decir, las denominadas coloquialmente *arrolladoras*, las *quieroynopuedo*, las *nifunifá* y el resto. Otro de los análisis es la tendencia en las que hay que distinguir entre la tendencia de alcista y bajista.

ESTADO DE LA CUESTIÓN

Otro sistema de *trading* que está cogiendo mucha fuerza, es el sistema automático, este consiste en realizar un *script* de programación en el cual, se contenga una serie de órdenes. Este tipo de trading se está aplicando de forma masiva en el mercado de Forex. Las ventajas de este tipo de *trading* son que se elimina la presión psicológica del *trader*, permite operar sin limitación de órdenes ni horarios y es posible realizar un test para comprobar si el robot es fiable.

El tipo de órdenes que se tienen en el *script* de programación, son una serie de reglas de entrada y salida de operaciones que se pueden basar en condiciones simples o complejas.

3.2 TWITTER

Twitter se creó en 2006 por Jack Dorsey, Noah Glass, Biz Stone y Evan Williams. Surgió como proyecto de investigación dentro de Obvious, una compañía de San Francisco. En sus comienzos era una red donde comunicarse con otros usuarios a través de 140 caracteres y hoy en día sigue comportándose de manera similar, pero con unas pequeñas mejoras. Hoy Twitter permite la promoción de cuentas y Tweets a través de los llamados “Promoted”, un sistema por el que las empresas pueden anunciarse pagando. También Twitter creó los *trending topic* que son fundamentalmente los temas más hablados del momento, los *Hashtags*, que son palabras clave para seguir cierto tema.

Para poder compartir información con empresas, desarrolladores y los usuarios de acceso programáticos, Twitter desarrolló unas APIs propias. Una API es un conjunto de definiciones y protocolos que se utilizan para desarrollar e integrar el software de las aplicaciones. API significa interfaz de programación de aplicaciones y esto permite que sus productos y servicios se comuniquen con otros, sin necesidad de saber cómo están implementados. Con ello se permite acceder a partes del servicio de Twitter mediante estas APIs. La información que puedes solicitar a estas APIs es pública y para acceder a ella, tienes que registrar una aplicación. Los servicios o puntos de conexión que ofrece las APIs de Twitter son fundamentalmente: cuentas y usuarios, tweets y respuestas, mensajes directos, anuncios y herramientas y SDK del editor.

3.3 ESTUDIOS QUE RELACIONAN LA BOLSA Y TWITTER

Ha habido diversos estudios sobre cómo afecta lo que se publica en Twitter y la bolsa. Uno de los más importantes fue un estudio que realizó la Universidad de Munich que tras analizar más de 250.000 tweets sobre bolsa, llegó a la conclusión de que podía predecir la evolución de la bolsa con un 87% de fiabilidad. Otro de los más importante es una aplicación del BBVA, Stockbuzz, que analiza la opinión en Twitter de los valores que forman el Ibex 35, con el propósito de conocer las expectativas sobre el mercado a raíz de la información que se publica en la red social.

Nuestro proyecto podría apoyar estos estudios prospectivos, mediante la implementación en las reglas de trading automático, del analizador sintáctico utilizado, es decir, *textblob* que explicaremos en el apartado 5.2.4. Con ello añadiríamos un punto de seguridad, que es el fin de este proyecto.

Capítulo 4. DEFINICIÓN DEL TRABAJO

4.1 *MOTIVACIÓN*

Como hemos comentado, actualmente el uso de bases de datos no relacionales está en auge en muchas empresas. En este sentido, hemos considerado que proponer una herramienta útil para evaluar la evolución del mercado bursátil y aumentar la seguridad de las inversiones nos podría dar una ventaja importante a la hora de ser candidato profesional para una empresa.

También se pretende adquirir conocimientos de, tanto poder comunicarse con una API ágilmente, como poder desarrollar las habilidades de programación y aplicar conceptos teóricos impartidos a lo largo del Grado y su aplicación de una manera práctica y real en nuestro proyecto; ejemplo de esto va a ser el uso de análisis estadísticos utilizando la correlación de variables o calculo de la “p” de Person y el desarrollo en varios lenguajes de programación de los diferentes *scripts* del proyecto. Además, hemos intentado profundizar en la gestión ágil de los procesamientos decidiendo no utilizar un sistema de *clusters* para guardar la información.

4.2 OBJETIVOS

Los objetivos principales que se van a tratar en este proyecto son:

1. Determinar desde un punto de vista de las emociones, si el texto en un tweet es de carácter positivo, negativo o neutro.
2. Determinar la relación que hay entre estos tweets y la evolución de la cotización de una empresa en el mercado de valores.
3. Disponer de información diferenciada a la hora de invertir en bolsa.

4.3 METODOLOGÍA

La definición de metodología: es el conjunto de métodos que se siguen en una investigación científica o en una exposición doctrinal.

Es muy importante decidir que método se va a usar al comienzo de cualquier proyecto y hacer una planificación de las tareas que se van a realizar a lo largo de este. Esto nos permite a lo largo del proyecto tener un *feed back* continuo. La metodología que vamos a usar en nuestro proyecto va a ser *waterfall* o también llamada modelo de desarrollo en cascada, este tipo de metodología consiste en realizar los requisitos o tareas de un proyecto de manera secuencial. Hemos elegido esta metodología porque nuestro proyecto tiene múltiples componentes y en caso de error no hace falta volver al comienzo del proyecto.

Una vez decidida la metodología que vamos a usar, realizamos un diagrama de GANTT con las tareas que se van a efectuar a lo largo del proyecto:

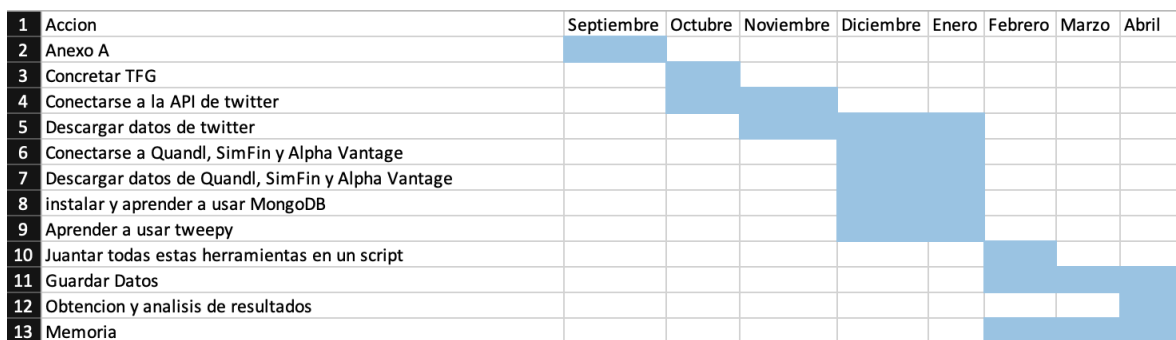


Figura 17 Diagrama de GRANT

4.4 PLANIFICACIÓN Y ESTIMACIÓN ECONÓMICA

Para realizar cualquier proyecto hay que tener en cuenta siempre el riesgo económico y la viabilidad de éste.

En nuestro caso concreto todas las herramientas definidas en la sección de descripción de tecnologías son gratuitas, por tanto, tendríamos el coste del equipo donde se van a usar estas herramientas y el coste de tiempo (en horas) de un ingeniero técnico *junior* que en promedio son 1800 euros mensuales (10,6 €/h) y el de un ingeniero superior con experiencia que ha estado supervisando el proyecto, que de media son 4000 euros mensuales (23,5 €/h).

Tabla 3 Costes del proyecto

<i>Concepto</i>	<i>Justificación</i>	<i>Coste</i>
<i>Ordenador</i>	Es el ordenador portátil donde se ha realizado el proyecto, MACBOOK PRO-2013	1000 €
<i>Trabajador</i>	Se han necesitado dos trabajadores para realizar el proyecto, un ingeniero técnico junior y un ingeniero superior con experiencia durante 6 meses	33800 €
<i>Herramientas</i>	Definidas en sección de descripción de tecnologías	0 €
<i>Total</i>		34800 €

Capítulo 5. SISTEMA/MODELO DESARROLLADO

5.1 ARQUITECTURA DEL SISTEMA

El proyecto se compone de dos fases, la fase de almacenamiento y la fase de análisis y resultados.

Fase de almacenamiento:

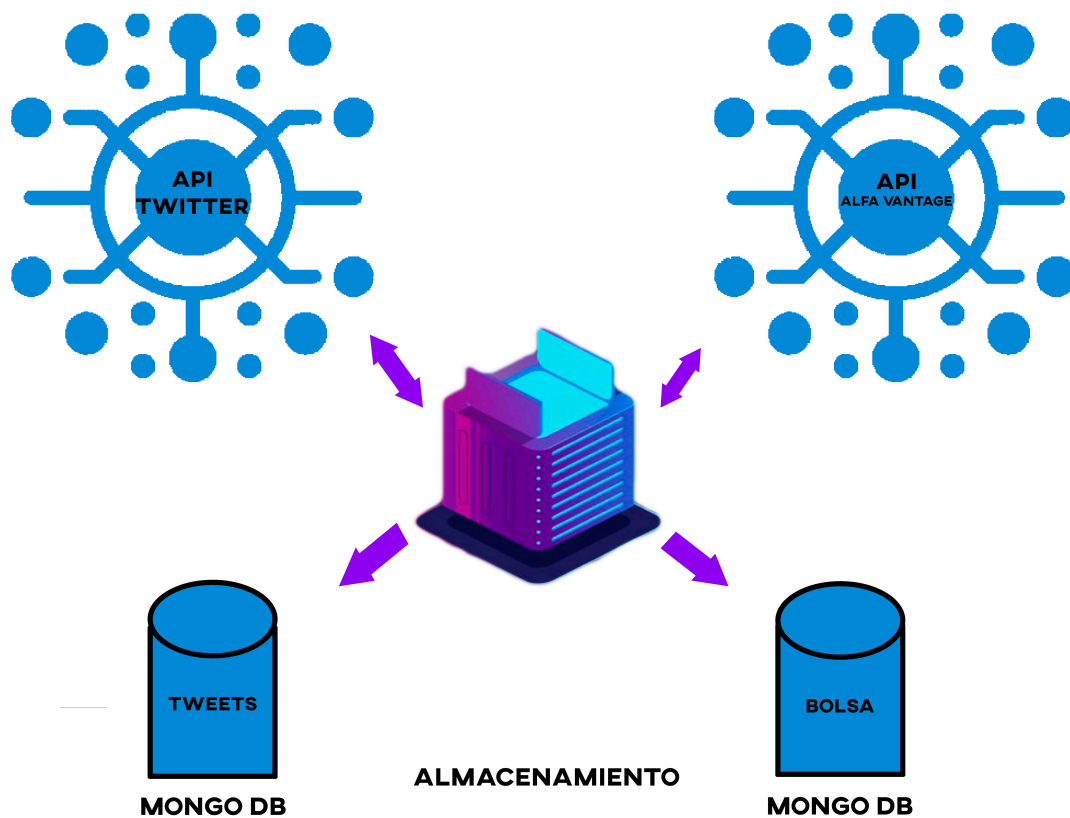


Figura 18 Almacenamiento de datos

En esta fase vamos a distinguir dos áreas: la de twitter (a la izquierda del servidor en la figura 18) y la de bolsa (a la derecha del servidor en la figura 18).

Vamos a comenzar hablando sobre el área de twitter. Aquí solicitaremos información a la API de twitter; esta información se tratará y archivará en una base de datos.

En este punto teníamos que decidir si usar un tipo de almacenamiento distribuido o centralizado. El almacenamiento centralizado se basa en que la totalidad de la información se almacena en una sola maquina y una sola CPU; está no posee múltiples elementos de procesamiento y tiene difícil recuperación en caso de perdida, pero evita la redundancia y la inconsistencia, además pueden aplicarse restricciones de seguridad y puede conservar la integridad; por último, el procesamiento de datos ofrece rendimiento y confiabilidad.

El almacenamiento de datos distribuidos se basa en que una colección de datos pertenece lógicamente a un mismo sistema, aunque no estén físicamente en el mismo espacio. Las ventajas de las bases de datos distribuidos son el compartimiento de datos, la autonomía y la disponibilidad y los inconvenientes son el coste de desarrollo del software, la mayor probabilidad de errores y la mayor sobrecarga de procesamiento.

Una de las tecnologías basada en los sistemas distribuidos más importante es *Big Data*. Cuando hablamos de *Big Data*, queremos decir un conjunto de datos o combinación de conjunto de datos cuyo volumen, variabilidad y velocidad de crecimiento hacen que su análisis sea más complejo con tecnologías y herramientas tradicionales. Esta complejidad reside en la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas. La importancia de *Big Data* atañe a que proporciona a las empresas financieras una capacidad de identificar problemas de una forma más comprensible, a movimientos de negocios más inteligentes, operaciones más eficientes y por ende a mejorar los beneficios.

Big Data es un sistema distribuido que posee una serie de *clusters* que se accede a ellos desde un servidor remoto. Pero para este proyecto hemos intentado profundizar en la gestión ágil de los procesamientos decidiendo no montar un sistema de distribuido, eligiendo así el

SISTEMA/MODELO DESARROLLADO

centralizado para guardar la información, ya que era innecesario, porque el tamaño de la información almacenada no ha superado el gigabyte, y eso puede ser procesado por un ordenador sin mayor problema. Por otra parte, una vez decidido el sistema de almacenamiento ahora pasaríamos a elegir si usar una base de datos relacional o no relacional. La diferencia principal entre ambas es que las relacionales guardan los datos de forma estructurada y la no relacionales no. Dentro de las no relacionales, tenemos una base de datos llamada MongoDB que guarda los datos en formato documento y que tiene la estructura de un JSON, que es el mismo con el que las APIs nos devuelven la información y por dicho motivo la hemos elegido.

El área de Bolsa es similar a la de Twitter. Aquí se ha de decidir, dentro de las diferentes APIs de bolsa cuál escoger y debatiremos por qué finalmente escogimos Alpha Vantage. La diferencia entre ambas áreas, Twitter y bolsa es que, en esta última, no se ha realizado tratamiento de datos; simplemente se guardaron valores de bolsa (*closing stock*) y la fecha en una base de datos.

Fase de Análisis y resultados:



Figura 19 Análisis y resultados

En esta fase vamos a recuperar los datos de sus respectivas bases. Una vez realizado esto, y transferidos a Excel, realizaremos los cálculos estadísticos usando la correlación o “p” de Pearson mediante el uso de un script de Python y la función *corr(method=pearson)*.

5.2 API DE TWITTER

En primer lugar, antes de conectarnos a la API es necesario crear un servidor Web, para poder realizar las pruebas de búsqueda más adelante. Para crear este servidor hemos utilizado un proveedor de dominios web llamado 000webhost, aquí nos hemos registrado y hemos

creado una página web cuyo dominio es <https://tfgfrjpom.000webhostapp.com/>. Este paso tiene dos fines: a/ porque es obligado para crear un usuario en Twitter para desarrolladores (lo que explicaremos más adelante) y b/ para efectuar pruebas.

5.2.1 CONEXIÓN A LA API

Para conectarse a la API, hay que crear una cuenta en twitter para desarrolladores en <https://developer.twitter.com/en/apps>. Si no se tiene una cuenta de usuario normal, está debe crearse para, en un segundo paso, crear una cuenta para desarrolladores, en los que se solicitan una serie parámetros como nombre, apellido, fecha de nacimiento y motivo de uso de la API; en nuestro caso, uso para estudio.

Una vez la cuenta esta aprobada, hay que crear una aplicación para desarrolladores de Twitter. Para dar de alta la aplicación hay que tener un dominio web como anteriormente hemos mencionado. Con la aplicación activa, hay que crear un *token* de acceso que tiene los siguientes parámetros: *consumer_key*, *consumer_secret*, *Access_token* y *Access_token_secret*. Estos parámetros nos permitirán autenticarnos para poder conectarnos a la API.

Por tanto, lo primero que haremos al inicio de cada conexión será autenticarnos, y lo haremos de la siguiente forma:

```
$settings = array(  
    'oauth_access_token' => "-----",  
    'oauth_access_token_secret' => "-----",  
    'consumer_key' => "-----",  
    'consumer_secret' => "-----"  
);
```

5.2.2 DESCARGA DE DATOS DE TWITTER

Para probar que podemos descargar datos de Twitter, solicitamos a la url que aparece en la primera línea del código la búsqueda que aparece en la segunda línea, es decir, buscaremos los tweets que posean la palabra “necesito” y descargaremos la cantidad =1, indicando que el método que vamos a usar es “get”, como se va a poder apreciar, el lenguaje utilizado a sido PHP.

```
1-$url = 'https://api.twitter.com/1.1/search/tweets.json';
2-$getfield = '?q=necesito&count=1';
3-$requestMethod = 'GET';
4-
5-/*json_decode(, $assoc=TRUE)*/
6-$twitter = new TwitterAPIExchange($settings);
7-$response= $twitter->setGetfield($getfield)
8-                                     ->buildOauth($url, $requestMethod)
9-                                     ->performRequest
12- print($response);
```

El formato de respuesta de Twitter es de tipo JSON



```
{
  "statuses": [
    {
      "created_at": "Mon Dec 23 10:09:46 +0000 2019",
      "id": 1209053474144559104,
      "id_str": "1209053474144559104",
      "text": "RT @jpenalver: \u00d7Amigos, estoy seriamente enfermo. Necesito con urgencia alguien me lleve a un m\u00e1\u00f1ico. Preferiblemente a emergencia de hosp\u00ed\u00e1l",
      "truncated": false,
      "entities": {
        "hashtags": [],
        "symbols": [],
        "user_mentions": [
          {
            "screen_name": "jpenalver",
            "name": "Jes\u00fas Penalver",
            "id": 85251929,
            "id_str": "85251929",
            "indices": [
              3,
              13
            ]
          }
        ],
        "media": [
          {
            "id": 1209053474144559104,
            "id_str": "1209053474144559104",
            "media_url": "https://t.co/Hw1X4dWjIQ",
            "media_url_https": "https://t.co/Hw1X4dWjIQ",
            "display_url": "https://t.co/Hw1X4dWjIQ",
            "expanded_url": "https://t.co/Hw1X4dWjIQ",
            "type": "photo",
            "url": "https://t.co/Hw1X4dWjIQ",
            "indices": [
              117,
              140
            ]
          }
        ]
      },
      "retweet_count": 1,
      "retweeted_by": [
        {
          "id": 1209053474144559104,
          "id_str": "1209053474144559104",
          "name": "Jes\u00fas Penalver",
          "screen_name": "jpenalver",
          "location": "Venezuela",
          "description": "Abogado- Propiedad Intelectual. Derecho Administrativo. Poes\u00eda. Humor. Democracia. libertades pu\u00f1ificas. Solu\u00f1ando un pa\u00f1o con dos soles. @EINacionalWeb",
          "url": "https://t.co/cOFLAvU4p",
          "entities": {
            "url": "https://t.co/cOFLAvU4p"
          }
        }
      ],
      "retweet_results": [
        {
          "completed_in": 0.021,
          "max_id": 1209053474144559104,
          "max_id_str": "1209053474144559104",
          "next_results": "?max_id=1209053474144559103&q=necesito&count=1&include_entities=1",
          "query": "necesito",
          "refresh_url": "?since_id=1209053474144559104&q=necesito&include_entities=1",
          "count": 1,
          "since_id": 0,
          "since_id_str": "0"
        }
      ]
    }
  ]
}
```

Figura 20 Respuesta Twitter formato JSON

5.2.3 DEFINIR NUESTRO PATRÓN DE BÚSQUEDA

Para nuestro proyecto queremos recoger los tweets que se publican sobre la empresa Morgan Stanley. Dado que no era posible recogerlos en su totalidad, hemos cogido la máxima muestra permitida con acceso libre, esto es 1000 tweets por día. Este paso lo hemos realizado con Python, que es el lenguaje que vamos a usar para el resto de proyecto.

```
1-from PrivateTemp import
consumer_key,consumer_secret,access_token,access_token_secret
2-auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
3-auth.set_access_token(access_token, access_token_secret)
4-api = tweepy.API(auth)
5-query='Morgan Stanley'
6-searched_tweets=searched_tweets = [status for status in
tweepy.Cursor(api.search, q=query, until="2020-02-22").items(1000)]
8-
9-for tweet in searched_tweets:
10-     print(tweet.text)
```

En la línea uno importamos nuestro token de acceso, en la línea cinco decidimos cual va a ser nuestra petición, que es “Morgan Stanley”, en la línea seis realizamos la petición y en la línea nueve y diez mostramos todos los tweets por la consola para comprobar que todo ha ido bien.

5.2.4 DETERMINAR LA POLARIDAD DE UN TWEET

Para determinar la polaridad del tweet, usamos una librería de procesamiento de lenguaje natural que se llama *textBlob*. Al aplicarle un texto a esta librería, esta devuelve una métrica de la polaridad y subjetividad, la polaridad tiene un rango de -1 hasta +1 y la subjetividad de 0 a 1. En nuestro proyecto nos vamos a fijar solo en la polaridad, pero antes tendremos que tratar el tweet, eliminando los caracteres extraños como pueden ser (@[A-Za-z0-9+])((^[0-9A-Za-z\t])|(w+:\V\S+)) valiéndonos de la función de Python Split().

Una vez eliminados esos caracteres, aplicaremos la función *textBlob* al texto de los tweets. Si la polaridad es menor que cero, es porque es una polaridad negativa, si es igual a cero, significa que es neutra y si es mayor que cero es positiva. Además, esta librería genera estos resultados independientemente del idioma en el que se ha escrito

5.3 APIS DE BOLSA

Las APIs de bolsa en las que vamos a hacer pruebas en nuestro proyecto son Quandl, SimFin y Alpha vantage. La alternativa a no usar las APIs, sería utilizar *web scrapping*; esto conllevaría que en el caso de que se cambiase el html de la página, habría que cambiar el código, lo cual no es eficiente.

5.3.1 CONEXIÓN A QUANDL

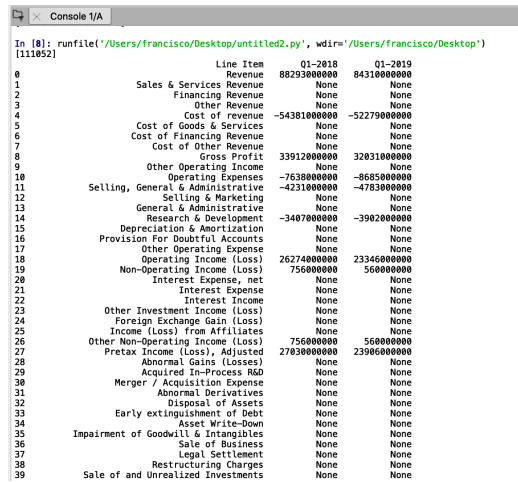
Quandl es una plataforma financiera y económica de datos alternativos que sirve para inversores profesionales. Para poder trabajar con esta plataforma hay que tener las librerías pandas, quandl y matplotlib previamente instaladas. Una vez realizado esto, hay que configurar la fecha de inicio y fin de la muestra y usar la función `quandl.get()`. Un ejemplo de una petición a Quandl sería el mostrado en la figura 21.

Date	Unit described in title
2015-03-31	0.37
2015-09-30	0.23
2016-03-31	0.32
2016-09-30	0.23
2017-03-31	0.25
2017-09-30	0.26
2018-03-31	0.28
2018-09-30	0.38

Figura 21 Ejemplo Quandl

5.3.2 CONEXIÓN A SIMFIN

Simfin también es una plataforma que proporciona información de la bolsa, pero funciona de manera distinta. Esta plataforma guarda los datos en ficheros csv. Por tanto, en lugar de usar una función `get()`, debemos usar una función `read_csv()`. Este fichero que hemos leído se puede inspeccionar mediante la función `head()` y obtendríamos un resultado como mostramos en la figura 22:



```

In [8]: runfile('/Users/francisco/Desktop/untitled2.py', wdir='/Users/francisco/Desktop')
[111852]

```

Line Item	01-2018	01-2019
Revenue	8825300000	8431000000
Sales & Services Revenue	None	None
Financing Revenue	None	None
Other Revenue	None	None
Cost of revenue	-54381000000	-52279000000
Cost of Goods & Services	None	None
Cost of Financing Revenue	None	None
Cost of Other Revenue	None	None
Gross Profit	33912000000	32031000000
Other Operating Income	None	None
Operating Expenses	-7638000000	-8685000000
Selling, General & Administrative	-4231000000	-4783000000
Selling & Marketing	None	None
General & Administrative	None	None
Research & Development	-3487000000	-3982000000
Depreciation & Amortization	None	None
Provision For Doubtful Accounts	None	None
Other Operating Expense	None	None
Operating Income (Loss)	26274000000	23346000000
Non-Operating Income (Loss)	750000000	500000000
Interest Expense, net	None	None
Interest Expense	None	None
Other Investment Income (Loss)	None	None
Foreign Exchange Gain (Loss)	None	None
Income (Loss) from Affiliates	None	None
Other Non-Operating Income (Loss)	750000000	500000000
Pretax Income (Loss), Adjusted	27030000000	23900000000
Abnormal Gains (Losses)	None	None
Acquired In-Process R&D	None	None
Merger / Acquisition Expense	None	None
Abnormal Derivatives	None	None
Disposal of Assets	None	None
Early extinguishment of debt	None	None
Asset Write-Down	None	None
Impairment of Goodwill & Intangibles	None	None
Sale of Business	None	None
Legal Settlement	None	None
Restructuring Charges	None	None
Sale of and Unrealized Investments	None	None

Figura 22 Ejemplo SimFin

5.3.3 CONEXIÓN A ALPHA VANTAGE

Alpha Vantage es otra plataforma que permite recuperar datos de bolsa y es la que finalmente se ha seleccionado para realizar el proyecto. Los motivos de esta selección fueron que Quandl era de pago y SimFin no dejaba claro cual era el *closing stock*, mientras que Alpha Vantage si que lo hacía. Para ejecutar las preguntas a dicha API hemos realizado el siguiente código:

```
1-from alpha_vantage.timeseries import TimeSeries
2-from PrivateTemp import key_alpha
3-ts = TimeSeries(key_alpha)
4-aapl, meta = ts.get_daily(symbol='MS')
5-fecha = datetime(2020,2,3)
6-for i in range(5):
7-    fecha_s= str(fecha)
8-    dia =fecha_s[0:10]
9-    print(dia)
10-    jsonData=aapl[dia]
11-    print(jsonData)
12-    fecha = fecha + timedelta(days=1)
```

En el código anterior, lo primero es autenticamos con nuestra clave de acceso y la variable *ts* en la línea tres. Mediante la función *get_daily()* en línea cuatro introducción el símbolo de la empresa que queremos obtener la información, en este caso MS es el símbolo de Morgan Stanley. Por último, mediante la variable *aapl* en la línea diez le pasamos la fecha que queremos solicitar. El bucle de la línea seis es para pasar a la variable *aapl* las fechas de una semana completa.

El resultado de la ejecución del código anterior es:

```
2020-02-03
{'1. open': '52.6400', '2. high': '53.5900', '3. low': '52.5800', '4. close': '5
2.6000', '5. volume': '8375131'}
2020-02-04
{'1. open': '53.7500', '2. high': '54.6000', '3. low': '53.7300', '4. close': '5
4.0200', '5. volume': '9373374'}
2020-02-05
{'1. open': '54.8600', '2. high': '55.1500', '3. low': '54.3829', '4. close': '5
4.9900', '5. volume': '8231751'}
2020-02-06
{'1. open': '55.4300', '2. high': '55.4900', '3. low': '54.6100', '4. close': '5
5.0100', '5. volume': '6974962'}
2020-02-07
{'1. open': '54.6100', '2. high': '55.1100', '3. low': '54.4200', '4. close': '5
5.0300', '5. volume': '7586746'}
```

Figura 23 Ejemplo Alpha Vantage

El dato que nos interesa es el *closing stock* y procederemos a almacenarlo en una base de datos.

5.4 BASES DE DATOS

Como hemos comentado en el apartado de arquitectura del sistema, el almacenamiento tanto de los tweets como el valor de bolsa lo vamos a realizar mediante una base de datos de MongoDB.

5.4.1 ALMACENAMIENTO DE TWEETS

Para almacenar los tweets lo vamos a hacer con Python, vamos a crear una Base de Datos que se llame “Tweets” y la colección se llamará “tweets”

```
1-myclient = pymongo.MongoClient("mongodb://localhost:27017/")
2-mydb = myclient["Tweets"]
3-mycol=mydb["tweets"]
4-
5-for tweet in searched_tweets:
6-     if (dia=="2020-02-21"):
7-         lista_semanal ={'texto':tweet.text, 'fecha':dia,'hora':hora,
'sentimiento':sentimiento}
8-         x=mycol.insert_one(lista_semanal)
```

En la línea uno nos conectamos a nuestra aplicación MongoDB del ordenador, en las líneas dos y tres nos conectamos a la base de datos y a la colección que queremos modificar, en nuestro caso, ampliar. Cada vez que ejecutamos el código ampliaremos nuestra base con los tweets de un determinado día.

5.4.2 ALMACENAMIENTO DE DATOS DE BOLSA

Para almacenar el valor de *closing stock* también lo vamos a realizar con Python y en este caso crearemos una base de datos denominada “Bolsa” y una colección llamada “valores”

```
1-import pymongo
2-
3-myclient = pymongo.MongoClient("mongodb://localhost:27017/")
4-
5-mydb = myclient["Bolsa"]
6-mycol=mydb["valores"]
7-
8-for i in range(5):
9-     datobolsamongo={'dia':dia,'valor': jsonData.get('4. close')}
10-     #x=mycol.insert_one(datobolsamongo)
```

5.4.3 RECUPERACIÓN DE DATOS

Para recuperar los datos seleccionados hemos realizado en un pequeño script que nos devuelva cantidad de tweets de cierta polaridad, por ejemplo, el número de tweets positivos, negativos y neutros de un determinado día, en este caso el día 2 de febrero de 2020 como se muestra a continuación:

```
1-for fut in cursor:
2-   if fut['fecha']== "2020-02-22" :
3-       cont_total=cont_total+1
4-   if fut['fecha']== "2020-02-22" and fut['sentimiento']=="positive":
5-       cont_positivo=cont_positivo+1
6-   if fut['fecha']== "2020-02-22" and fut['sentimiento']=="negative":
7-       cont_negativo=cont_negativo+1
8-   if fut['fecha']== "2020-02-22" and fut['sentimiento']=="neutral":
9-       cont_neutro=cont_neutro+1
```

Y el resultado de esa ejecución nos da 277 positivos, 74 negativos, 649 neutros

La recuperación de datos se va a realizar durante todo el periodo de estudio, es decir, febrero, marzo y abril llegando a los siguientes datos:

Datos de Febrero.

Tabla 4 Datos de febrero

Día	Positivo	Negativo	Neutro	Valor bolsa
3/2/20	121	120	277	52,6
4/2/20	344	69	385	54,02
5/2/20	437	72	389	54,99
6/2/20	191	84	232	55,01
7/2/20	173	58	236	55,03
10/2/20	244	52	704	55,35
11/2/20	233	56	711	55,78
12/2/20	146	74	337	56,18
13/2/20	804	16	180	55,93
14/2/20	431	58	511	55,84
17/2/20	149	57	157	55,58
18/2/20	265	127	608	55,52
19/2/20	211	168	458	56,31
20/2/20	261	105	634	53,75
21/2/20	277	74	649	52,42
24/2/20	99	49	852	49,68
25/2/20	184	62	581	48,03
26/2/20	294	47	547	47,24
27/2/20	170	110	641	45,41
28/2/20	128	89	783	45,03

Datos de Marzo:

Tabla 5 Datos de marzo

Día	Positivo	Negativo	Neutro	Valor bolsa
2/3/20	248	118	634	46,48
3/3/20	323	90	445	44,4
4/3/20	106	37	857	45,24
5/3/20	128	29	843	42,59
6/3/20	188	65	747	41,86
9/3/20	199	19	782	37,5
10/3/20	444	38	518	39,5
11/3/20	277	84	406	36,87
12/3/20	257	58	498	31,31
13/3/20	303	32	335	37,5
16/3/20	147	141	712	31,65
17/3/20	117	140	743	33,79
18/3/20	379	61	560	30,75
19/3/20	173	31	796	31,58
20/3/20	185	65	750	29,67
23/3/20	124	174	702	27,81
24/3/20	225	51	724	33,22
25/3/20	269	108	623	33,99
26/3/20	314	31	655	35,71
27/3/20	319	53	628	34,01
30/3/20	239	139	622	34,84
31/3/20	241	111	566	34

Datos de abril:

Tabla 6 Datos de abril

Día	Positivo	Negativo	Neutro	Valor bolsa
1/4/20	296	92	612	31,62
2/4/20	89	96	815	33,9
3/4/20	414	88	498	33,85
6/4/20	171	360	469	37,01
7/4/20	189	123	688	36,94
8/4/20	232	135	529	39,37
9/4/20	274	72	654	41,08
13/4/20	806	23	171	39,58
14/4/20	303	81	616	39,86
15/4/20	211	40	749	38,4
16/4/20	308	135	557	38,36
17/4/20	382	103	515	39,09
20/4/20	131	215	654	38,36
21/4/20	365	114	521	36,95
22/4/20	327	252	397	37,81
23/4/20	330	102	377	37,47
24/4/20	275	61	375	37,97
27/4/20	9	12	979	39,31
28/4/20	271	43	686	40,13
29/4/20	289	142	569	40,79
30/4/20	170	49	781	39,43

Capítulo 6. ANÁLISIS DE RESULTADOS

Para el análisis de los datos vamos a estudiar de forma independiente los meses de febrero, marzo y abril y después realizaremos el análisis de diferentes tramos de tiempos dentro de estos meses.

6.1 CORRELACIONES DE DATOS EN FEBRERO

En febrero hemos obtenido las siguientes correlaciones entre el número de tweets con una determinada polaridad y el valor de Bolsa

Tabla 7 Correlaciones en febrero

	Positivo	Negativo	Neutro	Valor bolsa
Positivo	1,0000	-0,4155	-0,2955	0,3330
Negativo	-0,4156	1,0000	0,0092	-0,0107
Neutro	-0,2955	0,0929	1,0000	-0,5071
Valor bolsa	0,3330	-0,010	-0,5071	1,000

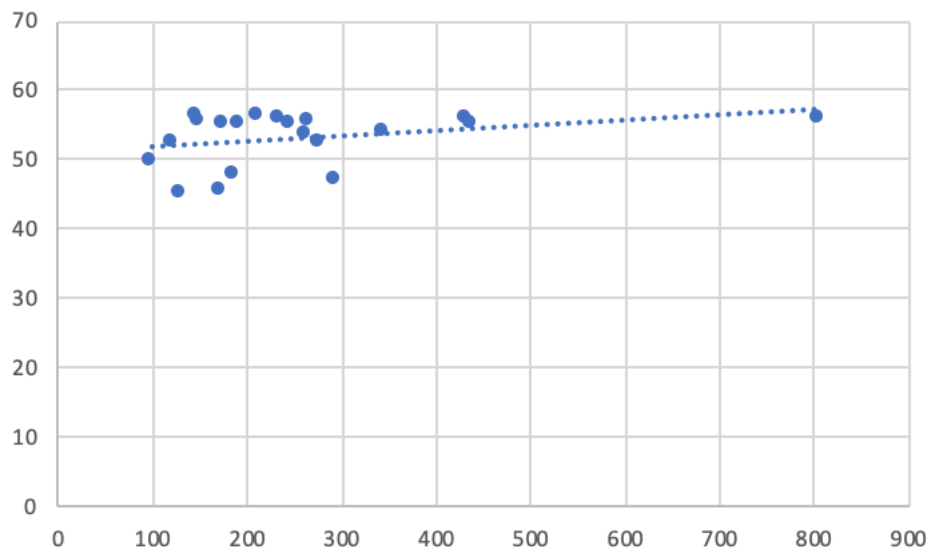


Figura 24 febrero correlación tweets positivos y valor de bolsa

Coefficiente de correlación de número de tweets negativos y valor de bolsa = -0,010

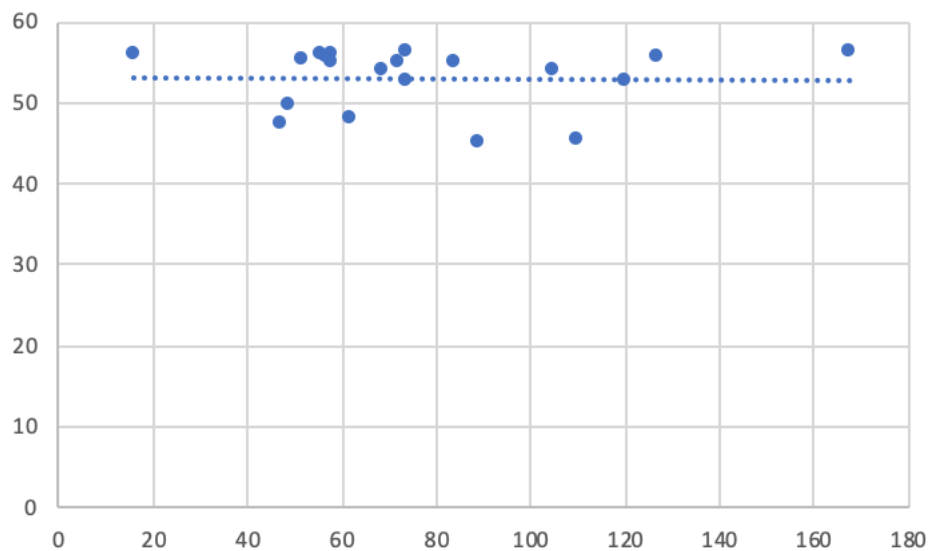


Figura 25 febrero correlación tweets negativos y valor de bolsa

Coeficiente de correlación de número de tweets neutros y valor de bolsa = -0,5071

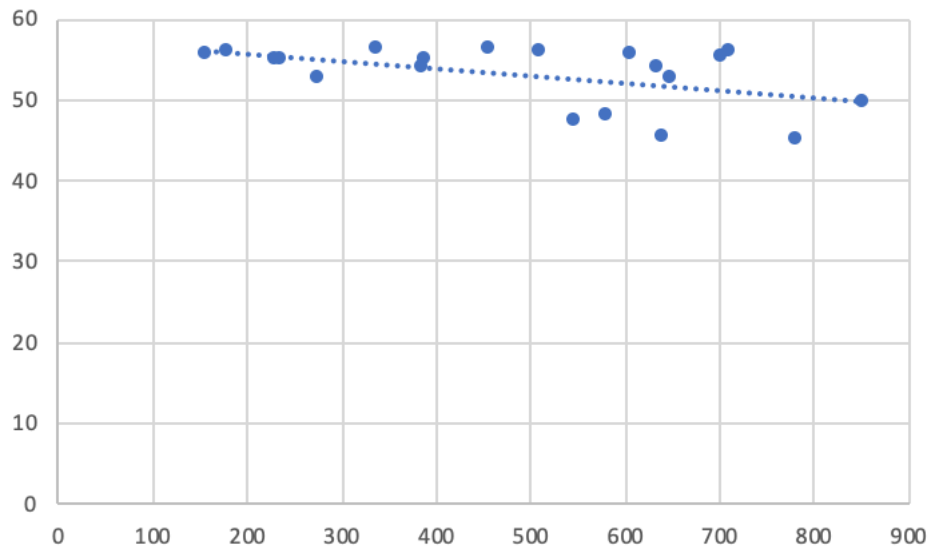


Figura 26 febrero correlación tweets neutros y valor de bolsa

6.2 CORRELACIONES DE DATOS EN MARZO

En marzo hemos obtenido las siguientes correlaciones entre el número de tweets con una determinada polaridad y el valor de Bolsa

Tabla 8 Correlaciones en marzo

	Positivo	Negativo	Neutro	Valor bolsa
Positivo	1,0000	-0,2769	-0,7240	0,0562
Negativo	-0,2769	1,0000	0,0059	-0,1286
Neutro	-0,7240	-0,0599	1,0000	-0,0017
Valor bolsa	0,0562	-0,2671	-0,0017	1,000

Coefficiente de correlación de número de tweets positivos y valor de bolsa = 0,0562

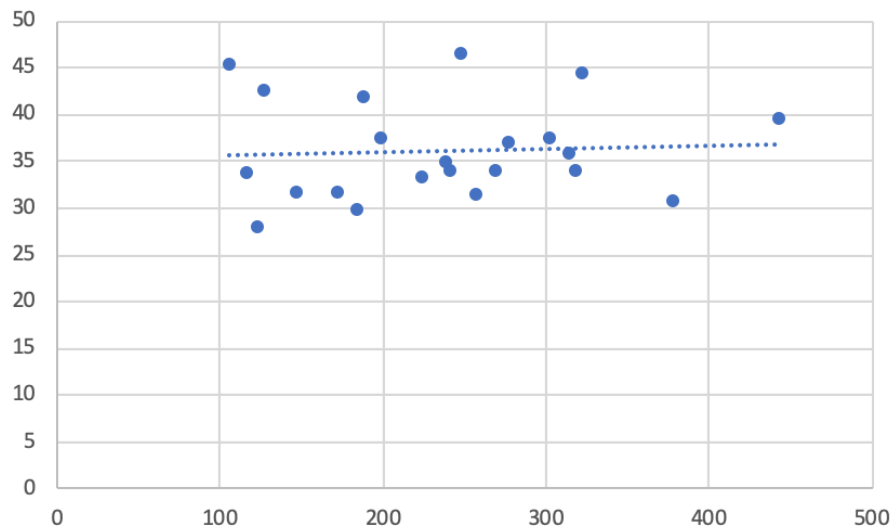


Figura 27 marzo correlación tweets positivos y valor de bolsa

Coefficiente de correlación de número de tweets negativos y valor de bolsa = -0,2671

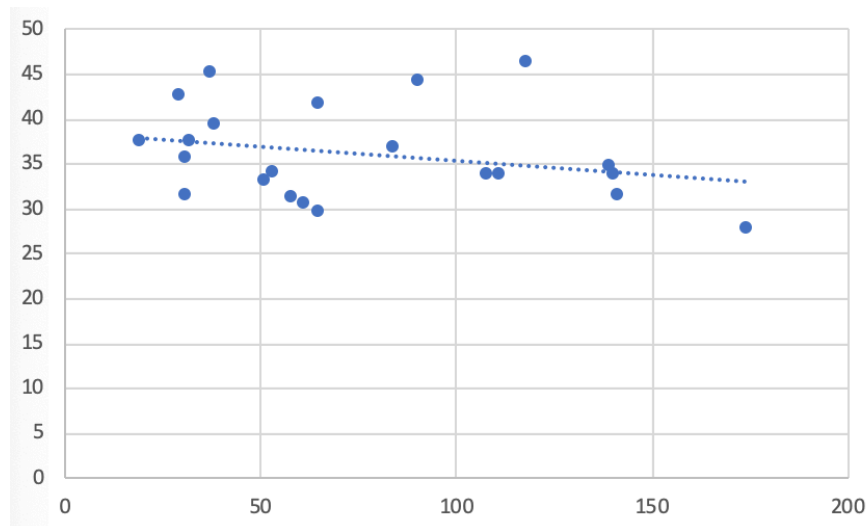


Figura 28 marzo correlación tweets negativos y valor de bolsa

Coefficiente de correlación de número de tweets neutros y valor de bolsa = -0,0017

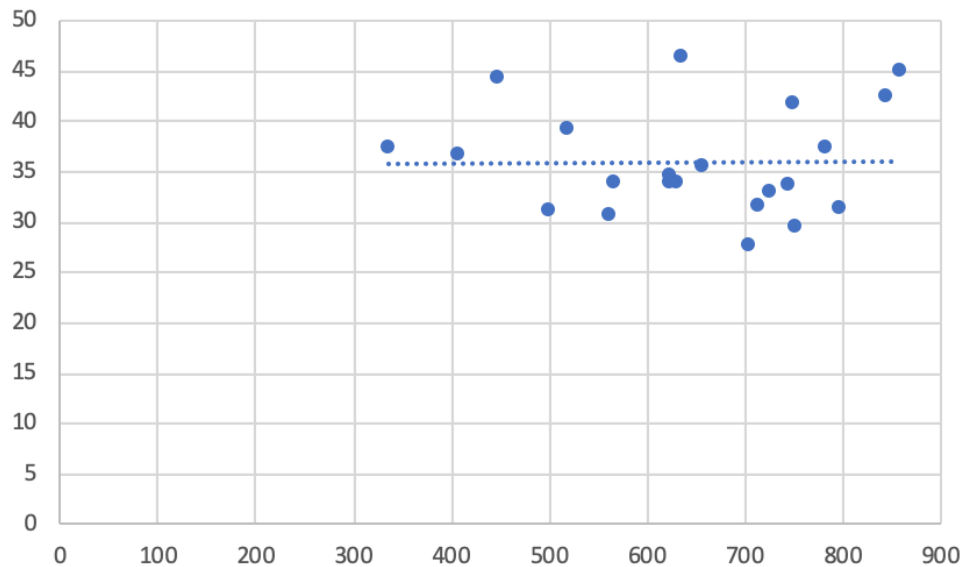


Figura 29 marzo correlación tweets neutros y valor de bolsa

6.3 CORRELACIONES DE DATOS EN ABRIL

En abril hemos obtenido las siguientes correlaciones entre el número de tweets con una determinada polaridad y el valor de Bolsa

Tabla 9 Correlaciones en abril

	Positivo	Negativo	Neutro	Valor bolsa
Positivo	1,0000	-0,1813	-0,8063	0,0706
Negativo	-0,1813	1,0000	-0,2610	-0,1432
Neutro	-0,8063	-0,2610	1,0000	-0,0025
Valor bolsa	0,0706	-0,1432	-0,0025	1,000

Coefficiente de correlación de número de tweets positivos y valor de bolsa = 0,0706

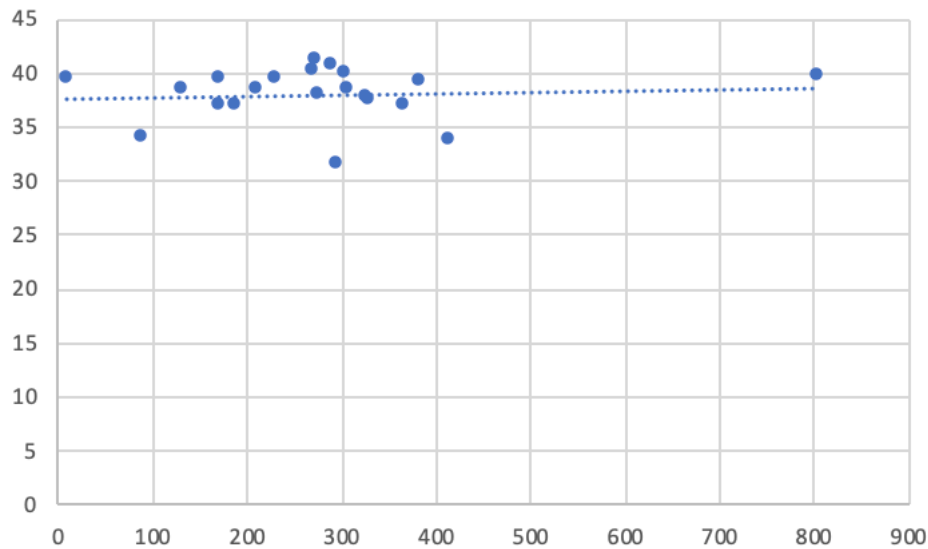


Figura 30 abril correlación tweets positivos y valor de bolsa

Coefficiente de correlación de número de tweets negativos y valor de bolsa = -0,1432

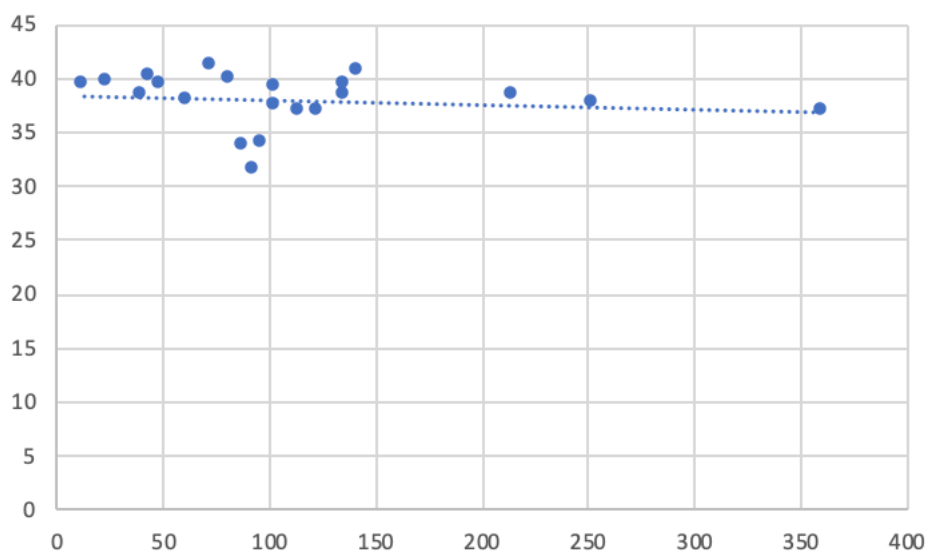


Figura 31 abril correlación tweets negativos y valor de bolsa

Coefficiente de correlación de número de tweets neutros y valor de bolsa = -0,0025

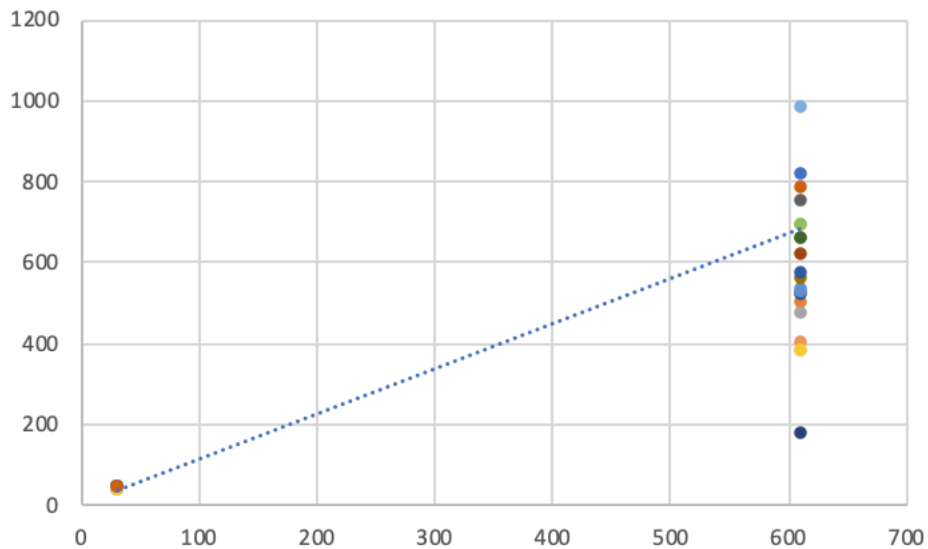


Figura 32 abril correlación tweets neutros y valor de bolsa

6.4 CORRELACIONES DE DISTINTOS PERIODOS

En el conjunto de marzo y abril hemos obtenido las siguientes correlaciones entre el número de tweets con una determinada polaridad y el valor de Bolsa

Tabla 10 Correlaciones en marzo y abril

	Positivo	Negativo	Neutro	Valor bolsa
Positivo	1,0000	-0,1507	-0,7777	0,0859
Negativo	-0,1507	1,0000	-0,2357	-0,1004
Neutro	-0,7777	-0,2357	1,0000	-0,0432
Valor bolsa	0,0859	-0,1004	-0,0433	1,000

En el conjunto de febrero, marzo y abril hemos obtenido las siguientes correlaciones entre el número de tweets con una determinada polaridad y el valor de Bolsa

Tabla 11 Correlaciones en febrero, marzo y abril

	Positivo	Negativo	Neutro	Valor bolsa
Positivo	1,0000	-0,1955	-0,5468	0,0827
Negativo	-0,1955	1,0000	-0,1010	-0,1513
Neutro	-0,5468	-0,1010	1,0000	-0,3573
Valor bolsa	0,0827	-0,1513	-0,3573	1,000

En el conjunto de febrero y la mitad de marzo hemos obtenido las siguientes correlaciones entre el número de tweets con una determinada polaridad y el valor de Bolsa

Tabla 12 Correlaciones en febrero y mitad de abril

	Positivo	Negativo	Neutro	Valor bolsa
Positivo	1,0000	-0,2871	-0,4011	0,1648
Negativo	-0,2871	1,0000	-0,0366	-0,1154
Neutro	-0,4011	-0,0366	1,0000	-0,3544
Valor bolsa	0,1648	-0,1154	-0,3441	1,000

6.5 CORRELACIONES EN TIEMPO REAL

Este análisis lo hemos realizado el día 26/05/2020 cogiendo los valores de bolsa en periodos de 10 minutos y viendo los tweets que se publican entre esas horas.

Para realizar hemos tenido que modificar el código de recopilación de datos de la bolsa, quedando así:

```
1-while True:
2-     ts = TimeSeries(key_alpha)
3-     aapl, meta = ts.get_daily(symbol='MS')
4-     daya = "2020-05-26"
5-     now = datetime.now()
6-     print("Hora: "+str(now.hour)+" "+str(now.minute)+"")
7-     jsonData=aapl[daya]
8-     print(jsonData.get('4. close'))
9-     time.sleep(60)
```

Las diferencias fundamentalmente están en la línea uno, en la que hacemos un bucle infinito en el que solicitamos la información y en la línea nueve, esperamos un minuto para volver a pedirla.

Para la recuperación de datos también hemos realizado unas modificaciones en el código que habíamos realizado en el apartado de recuperación de datos, usando ahora la librería *datetime*

```
1-if fut['fecha']== "2020-05-26" and fut['sentimiento']=="positive":
2-     hora = str(fut['fecha'])+" "+str(fut['hora'])
3-     hora =str(hora)
4-     hora_tweet =datetime.strptime(hora,'%Y-%m-%d %H:%M:%S')
5-     hora_abajo=datetime.strptime("2020-05-26 21:50:00", '%Y-%m-%d
%H:%M:%S')
7-     hora_arriba=datetime.strptime("2020-05-26 22:00:00", '%Y-%m-%d
%H:%M:%S')
8-
9-     if hora_tweet>= hora_abajo and hora_tweet<= hora_arriba:
10-         cont_positivo=cont_positivo+1
```

En código anterior es un ejemplo de como recuperar los tweets publicados entre 21:50 y 22:00 con polaridad positiva de nuestra base de datos.

Resultado de análisis en tiempo real del día 26/05/2020

Tabla 13 Datos en tiempo real 26/05/2020

Hora	Positivo	Negativo	Neutro	Total	Valor bolsa
15:40	0	1	2	3	41,95
15:50	1	0	1	2	42,24
16:00	2	0	3	5	42,32
16:10	1	0	2	3	42,185
16:20	1	0	2	3	42,23
16:30	0	0	4	4	42,52
16:40	1	0	3	4	42,685
16:50	1	0	0	1	42,66
17:00	0	0	3	3	42,845
17:10	0	0	0	0	42,89
17:20	2	0	0	2	42,885
17:30	0	1	2	3	42,76
17:40	0	0	1	1	42,82
17:50	0	1	0	1	42,86
18:00	0	0	1	1	42,83
18:10	1	1	1	3	42,72
18:20	1	0	3	4	42,79
18:30	0	0	0	0	42,7501
18:40	0	0	1	1	42,765
18:50	0	0	1	1	42,78

ANÁLISIS DE RESULTADOS

19:00	1	0	1	2	42,95
19:10	7	0	6	13	43,25
19:20	3	1	5	9	43,325
19:30	0	0	7	7	43,39
19:40	0	0	12	12	43,395
19:50	3	0	4	7	43,53
20:00	0	1	5	6	43,53
20:10	1	0	6	7	43,495
20:20	1	0	2	3	43,405
20:30	2	0	7	9	43,465
20:40	2	1	3	6	43,335
20:50	1	1	4	6	43,335
21:00	0	1	4	5	43,31
21:10	1	0	1	2	43,35
21:20	0	0	3	3	43,49
21:30	1	0	3	4	43,6
21:40	2	0	2	4	43,685
21:50	0	0	4	4	43,58
22:00	1	0	0	1	43,68

Tabla de correlaciones

Tabla 14 correlaciones en tiempo real

	Positivo	Negativo	Neutro	C. total	V. de bolsa
Positivo	1,0000	-0,0720	0,1978	0,5815	0,1910
Negativo	-0,0720	1,0000	0,0210	0,1256	-0,0013
Neutro	0,1978	0,0210	1,0000	0,9007	0,4087
C. total	0,5815	0,1256	0,9007	1,0000	0,4142
V. de bolsa	0,1910	-0,0013	0,4087	0,4142	1,0000

Capítulo 7. CONCLUSIONES

Los objetivos principales que se han a tratado en este proyecto son

1. Determinar desde un punto de vista de las emociones, si el texto en un tweet es de carácter positivo, negativo o neutro.

Este punto lo hemos logrado mediante la función *Textblob*

2. Determinar la relación que hay entre estos tweets y la evolución del valor de la acción del mercado de valores de una empresa.

Este punto lo hemos logrado calculando de la correlación entre ellos

Tabla resumen de correlaciones:

Tabla 15 Resumen correlaciones

	Positivo + B	Negativo + B	Neutro + B
Febrero	0,3330	-0,010	-0,5071
Marzo	0,0562	-0,2671	-0,0017
Abril	0,0706	-0,1432	-0,0025
Marzo + Abril	0,0859	-0,1004	-0,0433
Febrero + Marzo + Abril	0,0827	-0,1513	-0,3573
Febrero + ½ Abril	0,1648	-0,1154	-0,3441
Tiempo Real	0,1910	-0,0013	0,4087

CONCLUSIONES

El objetivo global del proyecto ha sido la utilización del análisis informático de las tendencias de opinión en Twitter para su aplicación en la prospección de los movimientos bursátiles. Creemos que ese objetivo ha sido conseguido ya que hemos objetivado una correlación entre los tweets positivo, negativos y neutros, siendo la correlación más grande en los tweets con polaridad neutra.

Debe hacerse hincapié en que si bien los datos de febrero pueden corresponder a los de un mes “habitual”, los de marzo y abril tienen unas condiciones especiales, ya que la bolsa descendió considerable y bruscamente, en “pánico” se podría considerar, debido a la pandemia mundial por COVID19, un escenario de estrés.

En definitiva, creemos que la herramienta que se presenta en este trabajo, por la información diferencial que obtiene, podría jugar un papel integrada de manera complementaria con los algoritmos de análisis de diferentes operadores profesionales.

En este sentido, hemos confirmado que existe que una correlación inversa en febrero con los tweets de polaridad neutra de $-0,5$, esto quiero decir a mayor número de tweets publicados con polaridad neutra, más bajará la bolsa. Por otra parte, en el análisis conjunto de los meses de marzo y abril, ninguna combinación de datos proporcionaba una correlación mayor a $\pm 0,11$, es decir, que apenas había correlación entre los tweets que se publicaron en esa época y el valor de bolsa.

Por último, hemos hecho el estudio en tiempo real en un momento donde la bolsa empezaba a remontar. Lo que objetivamos en el análisis del día 26/05/2020 es que existió una correlación de $+0,4$ con los tweets de polaridad neutra y de $+0,41$ con la cantidad total de tweets publicados. Es decir, una relación contraria a la obtenida en el mes de febrero, en este caso encontrábamos que a mayor numero de tweets publicados con polaridad neutra hay más probabilidad de crecer la acción de bolsa en tiempo real.

Capítulo 8. TRABAJOS FUTUROS

En nuestra opinión, un trabajo que podría complementar el presente debería tener las siguientes premisas

- 1- Usuario de Twitter para desarrolladores Premium: podría trabajar con todos los datos que se publican sobre una empresa en Twitter sin tener limitación de cantidad
- 2- La captura de datos durante un mayor periodo de tiempo: el presente proyecto consta de tres meses de almacenamiento de datos. Creemos que su validación o refutación debería llevarse a cabo con una muestra más representativa; debería considerarse los datos de al menos un año.
- 3- Captura de datos en condiciones normales: este proyecto se ha realizado en febrero, marzo y abril de 2020, en momento donde existía había una pandemia.

Capítulo 9. BIBLIOGRAFÍA

Big Data: ¿En que consiste? Su importancia, desafíos y gobernabilidad. Lugar de publicación: Powerdata. Recuperado de: <https://www.powerdata.es/big-data>

Enrique González Gutiérrez (2019) ¿Qué es PHP? ¿Para que sirve? Lenguaje de programación para crear paginas. Lugar de publicación: APRENDERAPROGRAMAR. Recuperado de: <https://www.aprenderaprogramar.com/attachments/article/492/CU00803B%20Que%20es%20PHP%20para%20sirve%20potente%20lenguaje%20programacion%20paginas%20web.pdf>

FCA (2019) el mejor software de trading automatico. Lugar de publicación: admiral market Recuperado de: <https://admiralmarkets.es/education/articles/trading-software/el-mejor-software-de-trading-automatico>

emontero (2019) ¿Qué es el trading automatico?. Lugar de publicación: invertir en bolsaweb.net Recuperado de: <https://www.invertirenbolsaweb.net/que-es-el-trading-automatico/>

(1/1/2020).¿Que es un dominio en internet? Lugar de publicación: MiPosicionamientoWeb. Recuperado de <https://miposicionamientoweb.es/que-es-un-dominio/>

Angel Robledano(23/11/2019). Que es Python. Lugar de publicación: OpenWebinars. Recuperado de <https://openwebinars.net/blog/que-es-python/>

Anaconda Distribution: la Suit más completa para la ciencia de datos con python. Lugar de publicación: DesdeLinux. Recuperado de <https://blog.desdelinux.net/ciencia-de-datos-con-python/>

¿Qué es mongoDB?. Lugar de publicación: mongoDB. Recuperado de <https://www.mongodb.com/es/what-is-mongodb>

Rubenf(3/2/2014).MongoDB: que es, como funciona y cuando podemos usarlo. Lugar de publicación: genbeta. Recuperado de <https://www.genbeta.com/desarrollo/mongodb-que-es-como-funciona-y-cuando-podemos-usarlo-o-no>

Duk2 Datos financieros para Python- test con Alpha Vantage. Lugar de publicación: Estrategias de trading. Recuperado de <https://estrategiastrading.com/datos-financieros-para-python-alpha-vantage/>

Eduardo Simon Hernandez (27/10/2012) Bases de datos distribuidas vs centralizadas Lugar de publicación: SlideShare. Recuperado de <https://es.slideshare.net/EduardoSimonHernandez/base-de-datos-distribuidas-vs-centralizadas>

Angel (13/12/2012) ¿Qué es y para que sirve Excel?. Lugar de publicación: accessyexcel. Recuperado de <http://www.accessyexcel.com/que-es-y-para-que-sirve-excel/>

La bolsa de valores: historia y funcioonn Lugar de publicación: gedesco. Recuperado de <https://www.gedesco.es/blog/la-bolsa-de-valores-historia-y-funcion-2/>

(1/3/2018).¿Qué es twitter?¿Como funciona?¿como puedo usarlo para mi organización. Lugar de publicación: webempresa. Recuperado de <https://www.webempresa.com/blog/que-es-twitter-como-funciona-2.html>

¿Qué es la bolsa?. Lugar de publicación: caixabank. Recuperado de: https://www.caixabank.com/deployedfiles/caixabank/Estaticos/PDFs/Espacio_accionista/junior_accionistas/Curso_de_bolsa_f01.pdf

Jimenez, L. (2009). El mercado Latinamericano de valores. Lugar de publicación: eumed. Recuperado de: <http://www.eumed.net/libros-gratis/2009a/510/HISTORIA%20Y%20ORIGEN%20DE%20LA%20BOLSA.htm>

BIBLIOGRAFÍA

Salinas, A. (10/24/2017). Origen historia e información completa sobre la red social twitter .
Lugar de publicación: Mott. Recuperado de <https://help.twitter.com/es/rules-and-policies/twitter-api>

Twitter. Información sobre las API de Twitter. Lugar de publicación:help.Twitter. Recuperado de: <https://mott.marketing/origen-historia-e-informacion-completa-sobre-la-red-social-twitter/>

El Mundo. (4/6/2011). ¿puede twitter predecir los vaivenes de la bolsa? . Lugar de publicación:
El Mundo. Recuperado de: <https://www.elmundo.es/elmundo/2011/04/06/navegante/1302089813.html>

(9/2/2013).BBVA triunfa con StockBuzz, una app que analiza el IBEX35 en twitter . Lugar de
publicación: Ticbeat. Recuperado de : <https://www.ticbeat.com/economia/bbva-stockbuzz/>

Objetivos de desarrollo sostenible. Lugar de publicación: UNDP.ORG. Recuperado de : <https://www.undp.org/content/undp/es/home/sustainable-development-goals.html>

Capítulo 10. ANEXO A GUÍA DE INSTALACIÓN

10.1 MONGODB

Para instalarnos MongoDB en el ordenador hay que acceder a la página oficial de Mongo, <https://www.mongodb.com/download-center/community> en la que seleccionas la versión, el sistema y el paquete, pero en el caso de IOS no puedes escoger ningún paquete, los requisitos para su instalación es tener descargado el programa XCODE y HOMEBREW, una vez tienes estos programas hay que escribir una serie de comandos en el terminal o consola

- Brew tap mongodb/brew
- Brew install [mongodb-community@4.2](#)
- Brew services start [mongodb-comunity@4.2](#)

Una vez realizados estos comandos, MongoDB ya esta funcionando

10.2 ANACONDA

Para descargar anaconda hay que ir a su página de descargas: <https://www.anaconda.com/download/>

Para ver que todo se a instalado correctamente, realizamos los siguientes comandos

```
(base) MacBook-Pro-de-Francisco-2:~ francisco$ conda -V
conda 4.7.12
(base) MacBook-Pro-de-Francisco-2:~ francisco$ python -V
Python 3.7.3
(base) MacBook-Pro-de-Francisco-2:~ francisco$ █
```

Figura 33 Instalación Anaconda y Python

Capítulo 11. ANEXO B MANUAL DE USUARIO

11.1 MONGODB

Para entrar en mongodb tenemos que ir a la consola del ordenador y es escribir mongo

```
(base) MacBook-Pro-de-Francisco-2:~ francisco$ mongo
MongoDB shell version v4.2.1
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("5f9f9229-078e-4e61-ad35-1c9bb50c28e9") }
MongoDB server version: 4.2.1
Server has startup warnings:
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten]
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten] **           Read and write access to data and configuration is unrestricted.
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display metrics about your deployment (disk utilization, CPU, operation statistics, etc)

The monitoring data will be available on a MongoDB website with a unique URL accessible to you and anyone you share the URL with. MongoDB may use this information to make product improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---
>
```

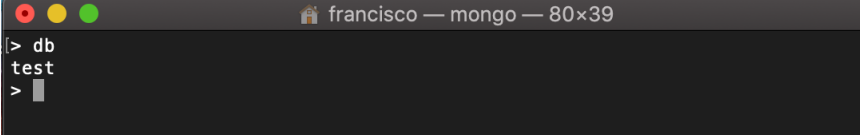
Figura 34 Conectandose a MongoDB

Para ver que bases de datos tenemos

```
francisco — mongo — 80x39
> show dbs
admin 0.000GB
config 0.000GB
local 0.000GB
>
```

Figura 35 Mostrar bases de datos

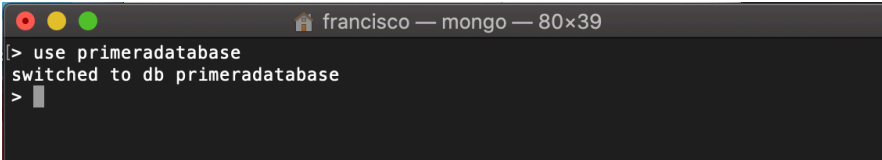
En que database estamos



```
francisco — mongo — 80x39
> db
test
>
```

Figura 36 Ver que en base de datos estamos

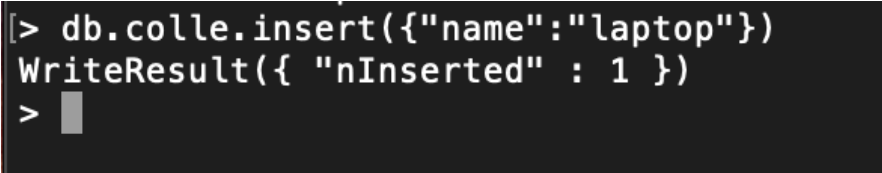
Crear una database



```
francisco — mongo — 80x39
> use primeradatabase
switched to db primeradatabase
>
```

Figura 37 Crear base de datos

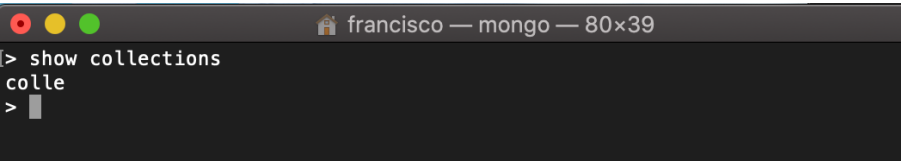
Agregar un json a la database



```
[> db.colle.insert({"name":"laptop"})
WriteResult({ "nInserted" : 1 })
>
```

Figura 38 Agregar colección

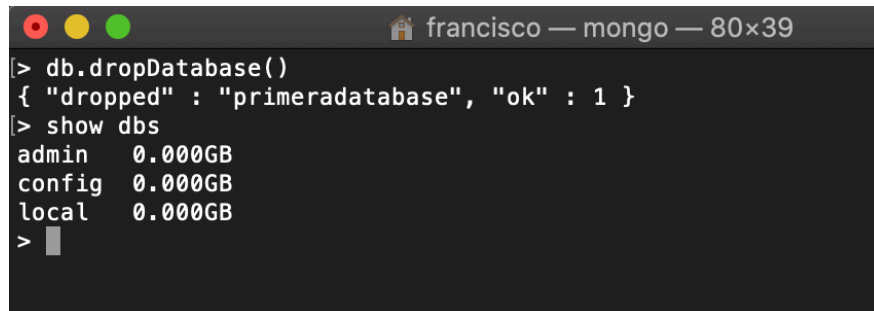
Ver que colecciones tenemos



```
francisco — mongo — 80x39
> show collections
colle
>
```

Figura 39 Mostrar colecciones

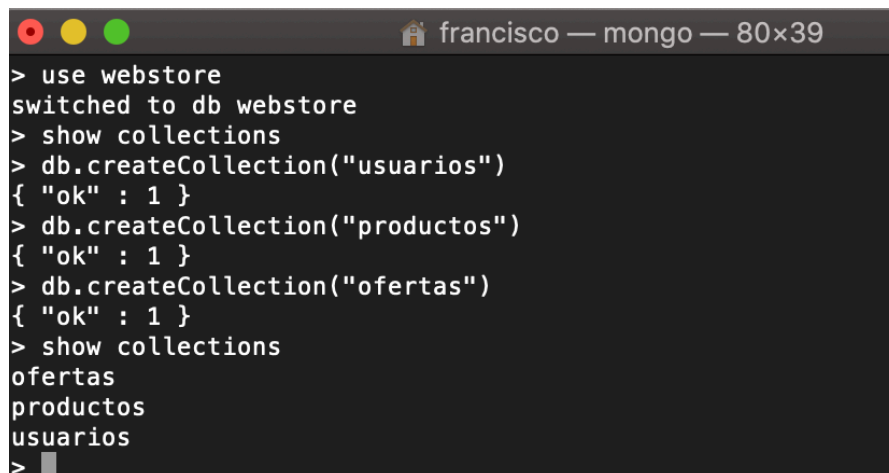
Eliminar database



```
francisco — mongo — 80x39
> db.dropDatabase()
{ "dropped" : "primeradatabase", "ok" : 1 }
> show dbs
admin    0.000GB
config  0.000GB
local   0.000GB
>
```

Figura 40 Eliminar base de datos

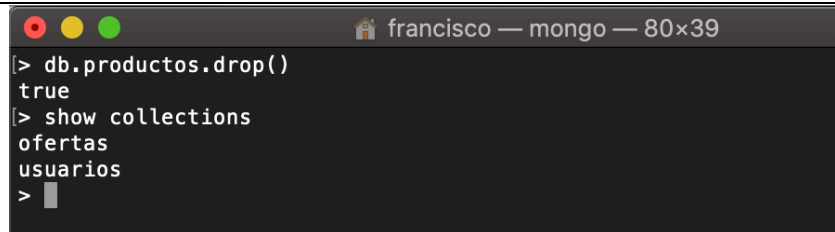
Crear colección



```
francisco — mongo — 80x39
> use webstore
switched to db webstore
> show collections
> db.createCollection("usuarios")
{ "ok" : 1 }
> db.createCollection("productos")
{ "ok" : 1 }
> db.createCollection("ofertas")
{ "ok" : 1 }
> show collections
ofertas
productos
usuarios
>
```

Figura 41 Crear colección

Eliminar colección



```
francisco — mongo — 80x39
> db.productos.drop()
true
> show collections
ofertas
usuarios
>
```

Figura 42 Eliminar colección

Añadir json a las colecciones



```
francisco — mongo — 80x39
> db.ofertas.insert({
... "nombre": "mac",
... "precio": 40,
... "activo": true,
... "somedata": [1,"a",["b",1,2]],
... "facturacion":{
... "nombre":"antonio",
... "residencia" :"madrid"
... }
... })
WriteResult({ "nInserted" : 1 })
>
```

Figura 43 Insertar colección

Para ver las colecciones:



```
francisco — mongo — 80x39
[> db.ofertas.find().pretty()
{
  "_id" : ObjectId("5e00a262970ecb2f957b83de"),
  "nombre" : "mac",
  "precio" : 40,
  "activo" : true,
  "somedata" : [
    1,
    "a",
    [
      "b",
      1,
      2
    ]
  ],
  "facturacion" : {
    "nombre" : "antonio",
    "residencia" : "madrid"
  }
}
```

Figura 44 Buscar elementos de una colección

s

Notas

- El método count te devuelve el número de colecciones de donde aparece cierta información

```
(base) MacBook-Pro-de-Francisco-2:~ francisco$ mongo
MongoDB shell version v4.2.1
connecting to: mongod://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongod
Implicit session: session { "id" : UUID("5f9f9229-078e-4e61-ad35-1c9bb50c28e9") }
MongoDB server version: 4.2.1
Server has startup warnings:
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten]
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten] **           Read and write access to data and configuration is unrestricted.
2019-12-23T09:33:20.268+0100 I CONTROL [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display metrics about your deployment (disk utilization, CPU, operation statistics, etc)

The monitoring data will be available on a MongoDB website with a unique URL accessible to you and anyone you share the URL with. MongoDB may use this information to make product improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---
```

Figura 45 Conectarse a MongoDB

11.2 API DE TWITTER

Una vez obtenido Access token y la contraseña

Hay que tener el fichero TwitterAPIExchange.php que se descarga de aquí:

<https://github.com/J7mbo/twitter-api-php/blob/master/TwitterAPIExchange.php>

Para Autenticarse

```
include("TwitterAPIExchange.php");
require 'simple_html_dom.php';

$settings = array(
    'oauth_access_token' => "_____",
    'oauth_access_token_secret' => "_____ ",
    'consumer_key' => "_____ ",
    'consumer_secret' => "_____ "
);
```

Solicitar un tweet según palabra de búsqueda y respuesta JSON

Palabra de búsqueda en getfield y número de tweet en count

```
$url = 'https://api.twitter.com/1.1/search/tweets.json';
$getfield = '?q=necesito&count=1';
$requestMethod = 'GET';

/*json_decode(, $assoc=TRUE)*/
$twitter = new TwitterAPIExchange($settings);
$response= $twitter->setGetfield($getfield)
            ->buildOAuth($url, $requestMethod)
            ->performRequest();

print($response);
```



```
{
  "statuses": [
    {
      "created_at": "Mon Dec 23 10:09:46 +0000 2019",
      "id": 120905347144559104,
      "id_str": "120905347144559104",
      "text": "RT @jpenalver: \u00c3\u00b1 Amigos, estoy seriamente enfermo. Necesito con urgencia alguien me lleve a un m\u00e100e9dico. Preferiblemente a emergencia de hospiu2026",
      "truncated": false,
      "entities": {
        "hashtags": [],
        "symbols": [],
        "user_mentions": [
          {
            "screen_name": "jpenalver",
            "name": "Jes\u00fas Pe\u00falver",
            "id": 85251929,
            "id_str": "85251929",
            "indices": [3, 13]
          }
        ],
        "urls": [
          {
            "expanded_url": "http://v.t.co/download/vandroid",
            "display_url": "http://v.t.co/download/vandroid",
            "url": "http://v.t.co/download/vandroid",
            "rel": "nofollow"
          }
        ]
      },
      "retweets_count": 0,
      "in_reply_to_status_id": null,
      "in_reply_to_status_id_str": null,
      "in_reply_to_user_id": null,
      "in_reply_to_user_id_str": null,
      "in_reply_to_screen_name": null,
      "user": {
        "id": 72110248785185280,
        "id_str": "72110248785185280",
        "name": "Gabriela Castillo",
        "screen_name": "laperrisedm",
        "location": "Venezuela",
        "description": "pensar",
        "url": null,
        "entities": {
          "urls": []
        },
        "protected": false,
        "followers_count": 196,
        "friends_count": 1094,
        "listed_count": 1,
        "created_at": "Fri Apr 15 22:26:45 +0000 2016",
        "favourites_count": 16987,
        "utc_offset": null,
        "time_zone": null,
        "geo_enabled": false,
        "verified": false,
        "statuses_count": 5963,
        "lang": null,
        "contributors_enabled": false,
        "is_translator": false,
        "is_translation_enabled": false,
        "profile_background_color": "f5f5dc",
        "profile_background_image_url": "http://abs.twimg.com/images/themes/theme1/bg.png",
        "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme1/bg.png",
        "profile_banner_url": "https://pbs.twimg.com/profile_banners/72110248785185280/1449610470",
        "profile_image_url": "http://pbs.twimg.com/profile_images/1201023812164358144V4QWbQ-p_normal.jpg",
        "profile_image_url_https": "https://pbs.twimg.com/profile_images/1201023812164358144V4QWbQ-p_normal.jpg",
        "profile_link_color": "1B95E0",
        "profile_sidebar_border_color": "f5f5dc",
        "profile_sidebar_fill_color": "f5f5dc",
        "profile_text_color": "333333",
        "profile_use_background_image": true,
        "has_extended_profile": true,
        "default_profile": false,
        "default_profile_image": false,
        "following": false,
        "follow_request_sent": false,
        "notifications": false,
        "translator_type": "none"
      },
      "geo": null,
      "coordinates": null,
      "place": null,
      "contributors": null,
      "is_quote_status": false,
      "lang": "es",
      "source": "android"
    }
  ]
}
```

Figura 46 Ejemplo JSON

Nota

- Si quieres buscar un hastag solo lo tienes que poner en q

Para Transformar el JSON en un array

```
$decodeficaco = json_decode($response, $assoc=TRUE);
```

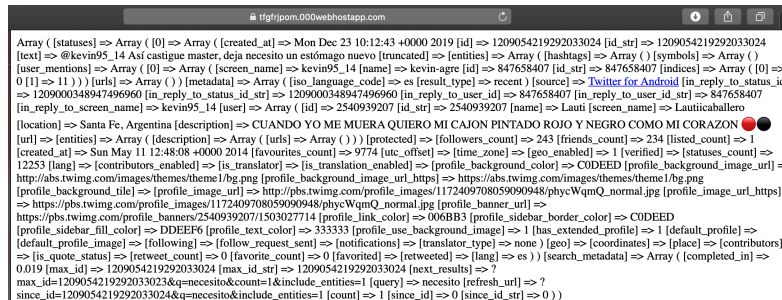


Figura 47 Ejemplo ARRAY

Para buscar los últimos tweets de un usuario

```
$url = 'https://api.twitter.com/1.1/statuses/user_timeline.json';
$getfield = '?screen_name=pablocasado_&count=2';
$requestMethod = 'GET';

$twitter = new TwitterAPIExchange($settings);
$response= json_decode( $twitter->setGetfield($getfield)
->buildOauth($url, $requestMethod)
->performRequest(), $assoc=TRUE);

foreach($response as $key)
{
    $profilepic= $key['user']['profile_image_url'];

    echo '<img src=' . $profilepic . '></img>';
    echo "<br>";
}
```

```
echo "Time and date: ".$key['created_at']."<br>";  
echo "Tweet: ".$key['text']."<br>";  
  
echo "Screen Name: ".$key['user']['screen_name'];  
echo " Followers: ".$key['user']['followers_count'];  
echo " Friends: ".$key['user']['friends_count']. "<br>";  
  
//echo "description" . $key[];  
//echo "<hr>";  
  
}
```

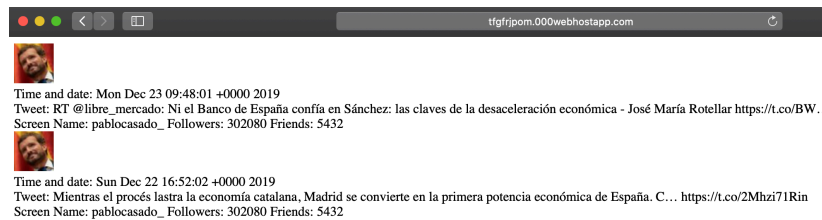


Figura 48 Ejemplo modificar formato

Notas

- Si te fijas ha cambiado la URL
- El texto no esta completo, no se puede forzar la variable truncate a false

11.3 DOMINIO WEB

Para crear el dominio web lo vamos a hacer desde 000webhost

<https://www.000webhost.com>

Una vez registrado, le damos a create new site, y allí registras el dominio

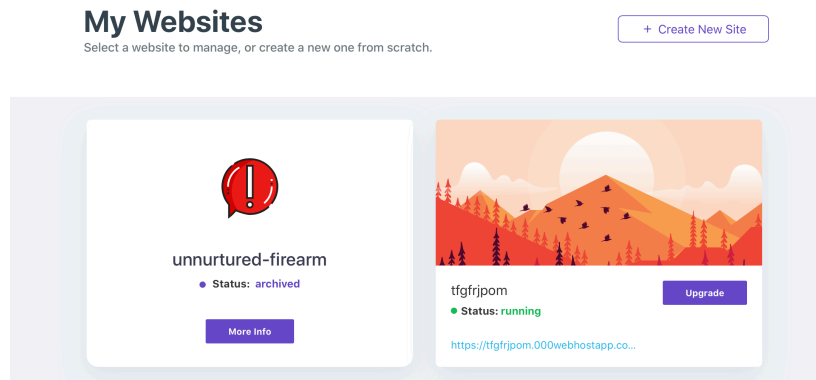


Figura 49 000webhost página inicio

Para subir archivos hay que entrar en upgrade y ir a upload files

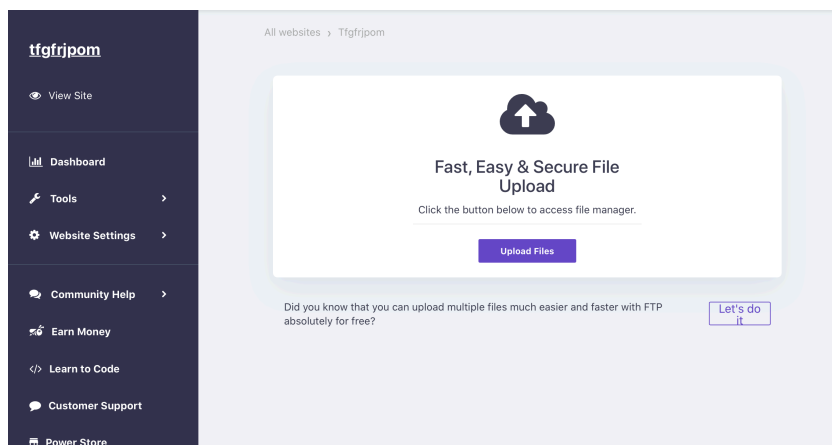


Figura 50 000webhost página de operaciones

Y aquí ya se pueden subir los archivos que quieras, pero en php

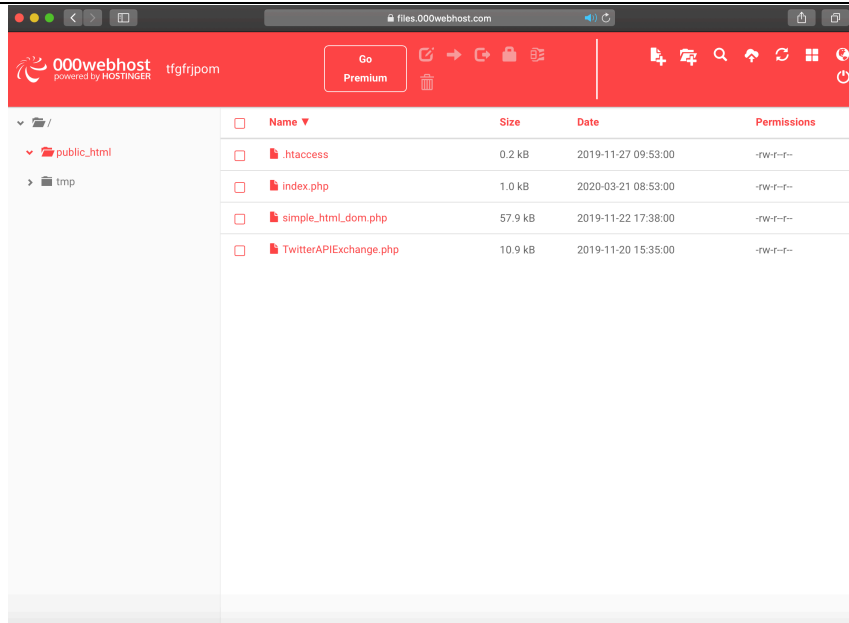


Figura 51 000webhost página de archivos

11.4 ANACONDA

Para realizar scripts de Python, simple tenemos que irnos a la aplicación de spyder

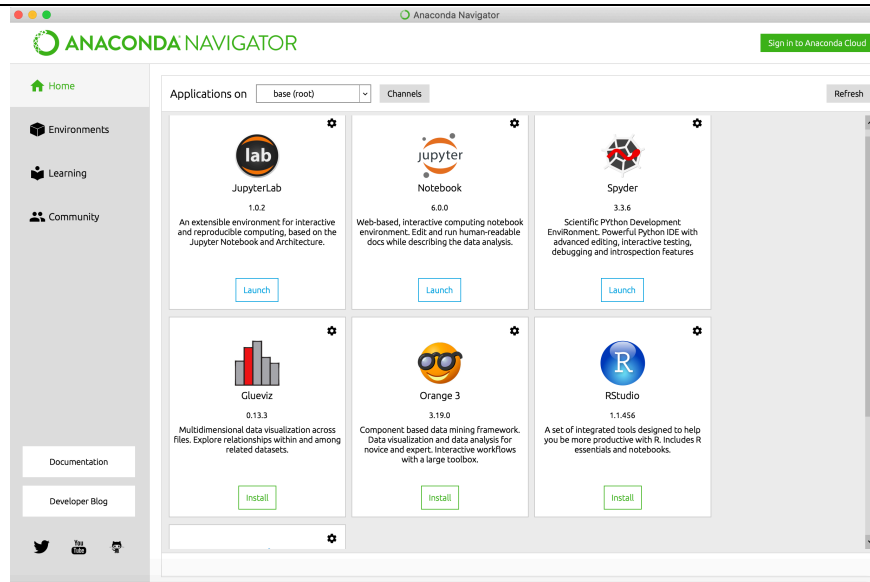


Figura 52 Anaconda página principal

En spyder escribimos el código y pulsamos el botón de play

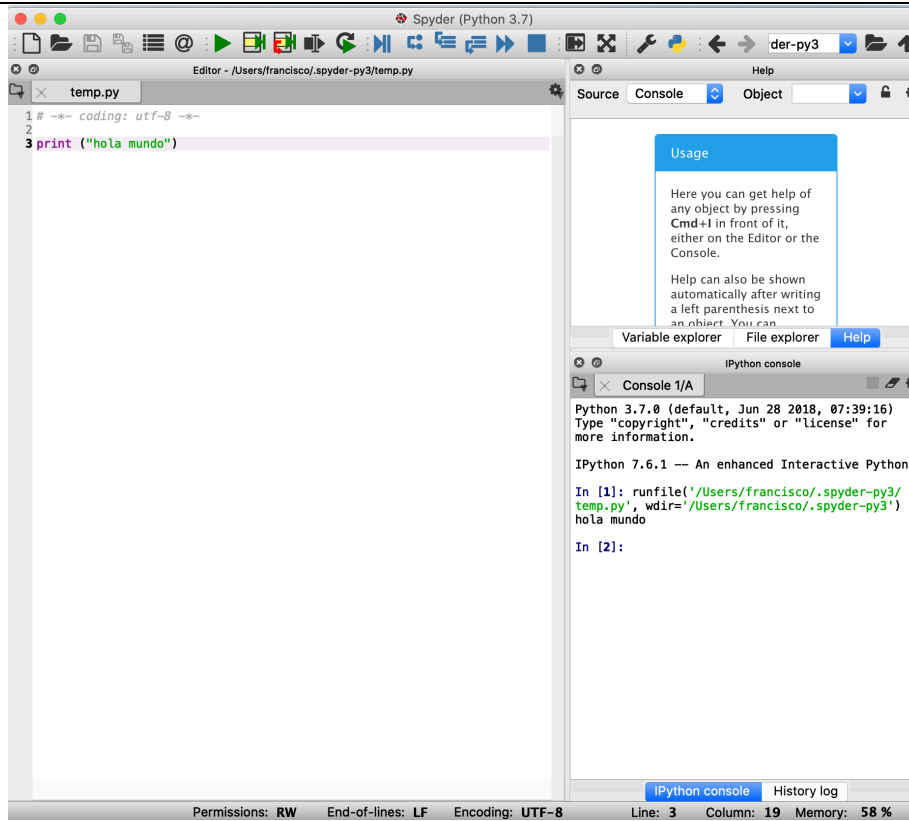


Figura 53 Página principal de SPYDER

11.5 ALPHA VANTAGE

Para registrarse tienes que ir a un página web : <https://www.alphavantage.co>

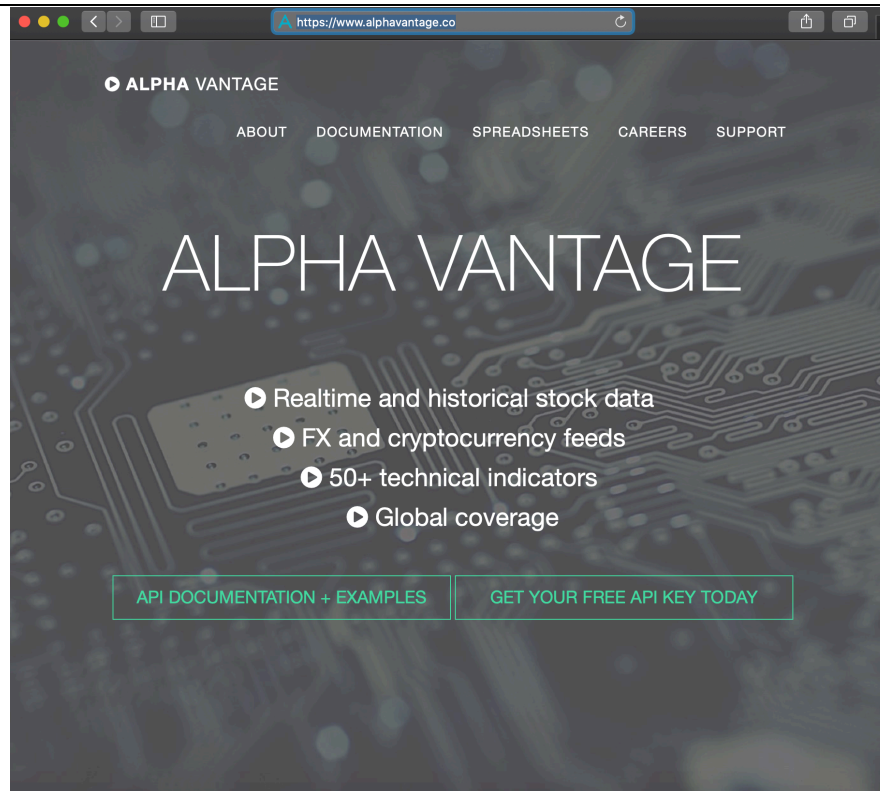


Figura 54 Página principal Alpha Vantage

Aquí pulsas GET YOUR FREE APY KEY TODAY, pones tus datos y te mandan tu clave.

Para realizar una petición

```
from PrivateTemp import key_alpha
ts = TimeSeries(key_alpha)
aapl, meta = ts.get_daily(symbol='MS')
jsonData=aapl[dia]
```

Capítulo 12. ANEXO C ODS

Las siglas ODS significan objetivos de desarrollo sostenible y hace un llamamiento mundial para lograr los siguientes objetivos : Fin de la pobreza, hambre cero, salud y bienestar, educación de calidad, igualdad de genero, agua limpia y saneamiento, Energía asequible y no contaminante, trabajo decente y crecimiento económico, industria, innovación e infraestructura, reducción de las desigualdades, ciudades y comunidades sostenible, producción y consumo responsables, acción por clima, vida submarina, vida de ecosistemas terrestres, paz, justicia e instituciones solidad y alianzas para lograr los objetivos.

Nuestro proyecto aplica al objetivo trabajo decente y crecimiento económico ya que proporciona una herramienta gratuita que puede estar operativa en un ordenador personal que ayudará a invertir con mayor seguridad.

Además, nuestro proyecto también aplica en el hecho de que se está democratizando la información, ya que, con estas aplicaciones todo el mundo puede acceder a mucha más información que de otro modo no sería posible. Además, que el sector financiero está últimamente apostando por la financiación verde, a través, de emisiones verde y este tipo de aplicativos hace que más gente puede acceder a productos complejos que ayuda a la sostenibilidad del planeta.