

# **Prototipado Rápido de un Sistema de Ayuda a la Toma de Decisiones para la Gestión de Carteras de Bolsa.**

Trabajo final de grado

Rodríguez García, Víctor (741414)

<b>1. Introducción</b>	<b>4</b>
1.1 Motivación	4
1.2 Objetivos	4
1.3 Alcance	5
1.4 Contenido del documento	5
<b>2. Estado del arte</b>	<b>6</b>
2.1 Funcionamiento de los mercados de valores[27]	6
2.2 Metodologías y Herramientas de Análisis	7
2.3 Fuentes de Datos y Entornos Software de Prototipado Rápido	8
2.4 Conclusiones	8
<b>3. Propuesta de solución</b>	<b>10</b>
3.1 Requisitos	10
3.2 Diseño del Sistema	10
3.2.1 Métodos de Análisis Elegidos	12
Regresión Lineal[19]	12
Exponential Moving Average (EMA)[17, pág. 30]	13
Double Exponential Moving Average (DEMA)[17, pág 43]	13
Triple Exponential Moving Average[17, pág 43]	13
Zero Lag Exponential Moving Average[17, pág 39]	14
3.3 Implementación	14
General	14
Regresión Lineal	15
Double Exponential Moving Average	16
Triple Exponential Moving Average	16
Zero Lag Exponential Moving Average	17
<b>4. Experimentos</b>	<b>19</b>
IBEX	19
IXIC	20
GSPC	22
AAPL	23
AMZN	24
GOOG	25
Conclusiones de los experimentos	27
<b>5. Conclusiones y trabajo futuro</b>	<b>28</b>
5.1 Conclusiones sobre el trabajo	28
5.2 Dedicación	28
5.3 Trabajo futuro	31
<b>Bibliografía</b>	<b>32</b>

<b>Anexo I</b>	<b>35</b>
Guia de instalación	35
Entrada y manipulación de datos	35
Google	35
Quandl	36
Tiingo	36

# **1. Introducción**

En esta sección voy a explicar los motivos que me han llevado a realizar este trabajo, los objetivos que quiero conseguir con el desarrollo de este proyecto y el alcance que puede tener el proyecto. Por último, se va a describir brevemente el contenido de este documento.

## **1.1 Motivación**

La motivación principal que tengo para realizar este Trabajo Fin de Grado desde el punto de vista académico es explorar entornos software de prototipado rápido para análisis de datos dinámicos de tipo económico. Quiero poder explorar la potencia y la viabilidad de uso de estos entornos software en un caso real, en este caso, en el análisis de datos dinámicos de tipo económico.

La motivación que tengo desde una perspectiva más personal es que soy estudiante de ingeniería informática y siempre me han interesado soluciones sencillas, rápidas y potentes a problemas con una cierta complejidad ya que esto ahorra una gran cantidad de tiempo al programador a la hora de programar, quiero encontrar un balance adecuado entre el tiempo necesario para desarrollar funcionalidades o actualizar las existentes utilizando recursos que permitan conseguir una eficiencia razonable en las soluciones. Creo que los entornos software de prototipado rápido pueden proporcionarme todo esto y por eso he decidido explorarlos, además he decidido analizarlos en el ámbito económico debido a mi gran interés por este mundo y en concreto me he decantado por la bolsa ya que siempre me he preguntado cómo invertir de la manera más inteligente posible y una posible opción es mediante la previsión de un valor en base a como se ha comportado ya previamente.

## **1.2 Objetivos**

El objetivo general del proyecto es el prototipado rápido de sistemas de ayuda a la toma de decisiones para inversión en bolsa.

Los objetivos concreto serían los siguientes:

- 1) Se quiere estudiar cómo generar herramientas sencillas para invertir con criterios sólidos.

- 2) Explorar herramientas informáticas de prototipado rápido que permitan predecir valores futuros en bolsa.
- 3) Realizar diferentes prototipos basándose en diferentes criterios.
- 4) Realizar experimentos que prueben los modelos y determinen cuál es el modelo que mejor se adapta a la realidad y que por lo tanto será más fiable.

## **1.3 Alcance**

En este proyecto se aborda el uso de herramientas software de prototipado rápido para la creación de carteras de inversión. La idea es hacer una primera aproximación al problema para evaluar la facilidad para incorporar y ampliar el prototipo con nuevos criterios de análisis de inversión en bolsa.

El alcance que puede tener esta herramienta no va ir más allá de un uso personal de la herramienta. En un futuro se podrían ampliar las funcionalidades de la herramienta para cubrir más necesidades que surjan y evolucionar el prototipo en consecuencia.

## **1.4 Contenido del documento**

En una primera parte se describe el funcionamiento de los mercados de valores y las diferentes herramientas de análisis y software que existen a día de hoy para analizar mercados.

En la siguiente parte se introduce la propuesta de solución en la que en primer lugar se explica como es el sistema, a continuación se introducen las diferentes herramientas utilizadas, cada una junto con su debida explicación. Y finalmente se explica como se ha implementado cada herramienta matemática en la herramienta de prototipado rápido.

A continuación se expone el fin de los experimentos realizados así como los resultados obtenidos y una conclusión en la que se explica para qué casos se deberían de utilizar las diferentes herramientas.

Por último, se exponen las conclusiones extraídas sobre el trabajo del autor, la dedicación de este y el trabajo futuro que se podría realizar para expandir y mejorar este trabajo.

## 2. Estado del arte

En esta sección se explica el funcionamiento de un mercado de valores y se introducen las diferentes herramientas de análisis y software que existen para realizar predicciones sobre estos mercados. Por último se incluyen unas pequeñas conclusiones de la información recopilada.

### 2.1 Funcionamiento de los mercados de valores<sup>[27]</sup>

"El mercado de valores es un tipo de mercado de capitales de los que operan alrededor del mundo en el que se negocia la renta variable y la renta fija de una forma estructurada, también la compra y venta de bienes que tiene un plan fijo, o un negocio que tiene renta variable a través de la compraventa de valores negociables. Permite la canalización de capital a medio y largo plazo de los inversores a los usuarios. Tanto como para poder tener dinero o a la vez no, podemos tener una amplia compra o una amplia venta."<sup>[1]</sup>

En los mercados de valores existen cuatro roles principalmente<sup>[2]</sup>:

- **Emisores:** Entidades que emiten los valores negociados en bolsa.
- **Reguladores:** Se encargan de supervisar el funcionamiento del mercado y de redactar las normas para su control.
- **Intermediarios:** Agentes mediante los cuales el inversionista puede comprar o vender instrumentos financieros.
- **Inversionistas:** Son los que invierten el dinero esperando obtener algún beneficio de este.

"Los Mercados de Valores en la actualidad son automatizados y la mayoría de las operaciones se realizan mediante los ECN, que son simplemente un conjunto de servidores (computadoras) que contienen el sistema o programa donde confluyen todos los compradores y todos los vendedores de manera virtual. Usualmente las corredoras proporcionan las cotizaciones utilizando acciones de su propio inventario para venderlas y comprarlas a sus clientes. Es por ello que se les denomina los "hacedores del mercado" (market makers), pues son ellos los responsables de mantener la liquidez de los mismos.

Para cotizar sus valores en la Bolsa, las empresas primero deben hacer públicos sus estados financieros, puesto que a través de ellos se pueden determinar los indicadores que permiten saber la situación financiera de las compañías. El vendedor puede ser la propia empresa, uno, varios o todos los accionistas mayoritarios. Los activos financieros ofrecidos en venta comúnmente son representados por acciones o obligaciones. Las acciones ofrecidas pueden ser parte o la totalidad de la empresa. Los vendedores deberán hacer presentación de un folleto donde se explican las condiciones de la oferta pública de venta, que son registradas y verificadas por el órgano regulador de la bolsa correspondiente. El folleto informativo se pone a disposición de los interesados, quienes podrán aceptarlo o no, y se abren los períodos de aceptación y finalización de la oferta pública de venta.”<sup>[1]</sup>

## **2.2 Metodologías y Herramientas de Análisis**

En cuanto a la literatura científica sobre análisis de carteras en bolsa, el abanico de estudios es muy amplio. Por un lado, es posible encontrar trabajo clásico relacionado con la cuantificación del riesgo en los mercados <sup>[20][21][23][17]</sup> e incluso patentes <sup>[22]</sup>. Si nos centramos en trabajo reciente, podemos encontrar métodos que se acercan a la cuantificación del riesgo de los mercados desde la perspectiva de la estimación de la densidad de núcleos (Kernel Density Estimation) <sup>[24]</sup> y, en general, de otros métodos que han tenido éxito en otras disciplinas científicas, como los sistemas basados en la toma de decisiones y otras áreas <sup>[25][26]</sup>.

En cuanto a las herramientas empleadas para el análisis de mercados, una de las más utilizadas es SPSS, esto es gracias a su facilidad de uso, flexibilidad y escalabilidad. Esta herramienta ofrece análisis estadístico avanzado, una amplia biblioteca de algoritmos de machine learning, análisis de texto, extensibilidad de código abierto, integración con big data y un fácil despliegue en las aplicaciones.<sup>[3]</sup>

Otra herramienta a destacar es el lenguaje de programación R. “Se trata de uno de los lenguajes de programación más utilizados en investigación científica, siendo además muy popular en los campos de aprendizaje automático (machine learning), minería de datos, investigación biomédica, bioinformática y matemáticas financieras.”<sup>[4]</sup>

Por último, una herramienta matemática muy utilizada también son las medias móviles, estas son indicadores de tendencia que se utilizan en análisis para invertir en diferentes instrumentos financieros. Son muy

utilizados debido a que actúan como suavizadores de los movimientos de sus precios, ya sea en el corto, medio o largo plazo.<sup>[17]</sup>

## **2.3 Fuentes de Datos y Entornos Software de Prototipado Rápido**

Se encuentran diferentes fuentes de las que extraer datos, una fuente de datos para las empresas del mercado español es yahoo<sup>[5]</sup>, de esta se pueden obtener los datos históricos de las empresas. Estos datos van desde la fecha en la que la empresa entró en el mercado de valores hasta el dia de hoy. Para obtener los datos en tiempo real, a parte de yahoo finanzas, también se encuentran las páginas web iberbolsa<sup>[6]</sup> e infobolsa<sup>[7]</sup>.

En cuanto a una fuente para los mercados a nivel mundial, además de yahoo se encuentra investing<sup>[8]</sup> en la que se dispone información de los diferentes mercados junto con sus valores actuales y un histórico de los diferentes valores que toman los mercados mundiales.

Existen diferentes herramientas que permiten realizar pronósticos sobre cómo se va a comportar el mercado, entre estas herramientas se encuentran Universal Market Predictor Index<sup>[9]</sup> e intelicharts<sup>[10]</sup> las cuales son dos herramientas de software cerrado. Por otra parte existen herramientas que ofrecen los bancos y otras entidades para gestionar carteras en las que dependiendo del riesgo que esté dispuesto a asumir una persona, el capital de inversión inicial y mensual y el plazo planteado se crea una cartera de inversiones donde se detallan las expectativas en el caso mejor, peor y medio.

En cuanto a las herramientas de prototipado rápido existen varias, entre estas se encuentran herramientas como KNIME<sup>[11]</sup>, RStudio<sup>[12]</sup> y RapidMiner<sup>[13]</sup>. KNIME y RapidMiner son dos herramientas que ofrecen un entorno gráfico para la resolución de los problemas con nodos que se conectan entre sí y realizan diferentes operaciones mientras que en RStudio la resolución se hace mediante líneas de código. En KNIME existe un flujo de trabajo (herramienta de prototipado rápido) que permite hacer predicciones en bolsa.

## **2.4 Conclusiones**

En conclusión, existen herramientas matemáticas que permiten hacer un análisis matemático en este caso de la bolsa y existen herramientas software que ayudan a invertir a una persona pero son cerradas y no se pueden realizar grandes cambios en ellas.

Estas herramientas son interesantes pero de cara a este Trabajo Fin de Grado interesa más una herramienta que pueda ser modificada y ampliada, como se van a utilizar herramientas de prototipado rápido el flujo de trabajo de KNIME es un punto perfecto para el comienzo de este Trabajo Fin de Grado.

En cuanto a las metodologías de análisis de riesgo y mercados, se percibe el interés que tiene el perfeccionamiento de los métodos clásicos (como los utilizados en el análisis de las series temporales de datos), así como un abanico muy amplio de enfoques novedosos entre los que se encuentran aquellos basados en técnicas de Aprendizaje.

## **3. Propuesta de solución**

Para la propuesta de solución, planteamos una serie de requisitos que delimitarán el alcance de la solución y facilitarán decidir los criterios de diseño.

### **3.1 Requisitos**

Como fuentes de datos, se utilizarán series temporales de índices y acciones de bolsa. El objetivo concreto que se plantea será predecir qué valor van a tomar y seleccionar aquellos los cuales incrementen más porcentualmente. De esta forma el usuario podrá elegir qué periodo analizar y concluir por su cuenta si los resultados son suficientemente buenos para utilizar el sistema. El usuario también podrá utilizar varias medias móviles para el cálculo de las predicciones y comparar la mejor en cada caso y hacer predicciones en una ventana temporal de un número de días delimitado, de esta forma nos podremos adaptar a las necesidades de los usuarios en cuanto a tiempo de inversión se refiere.

En cuanto al sistema, se implementará usando herramientas de prototipado rápido que permitan la incorporación de múltiples fuentes de datos y que faciliten el análisis posterior de los resultados.

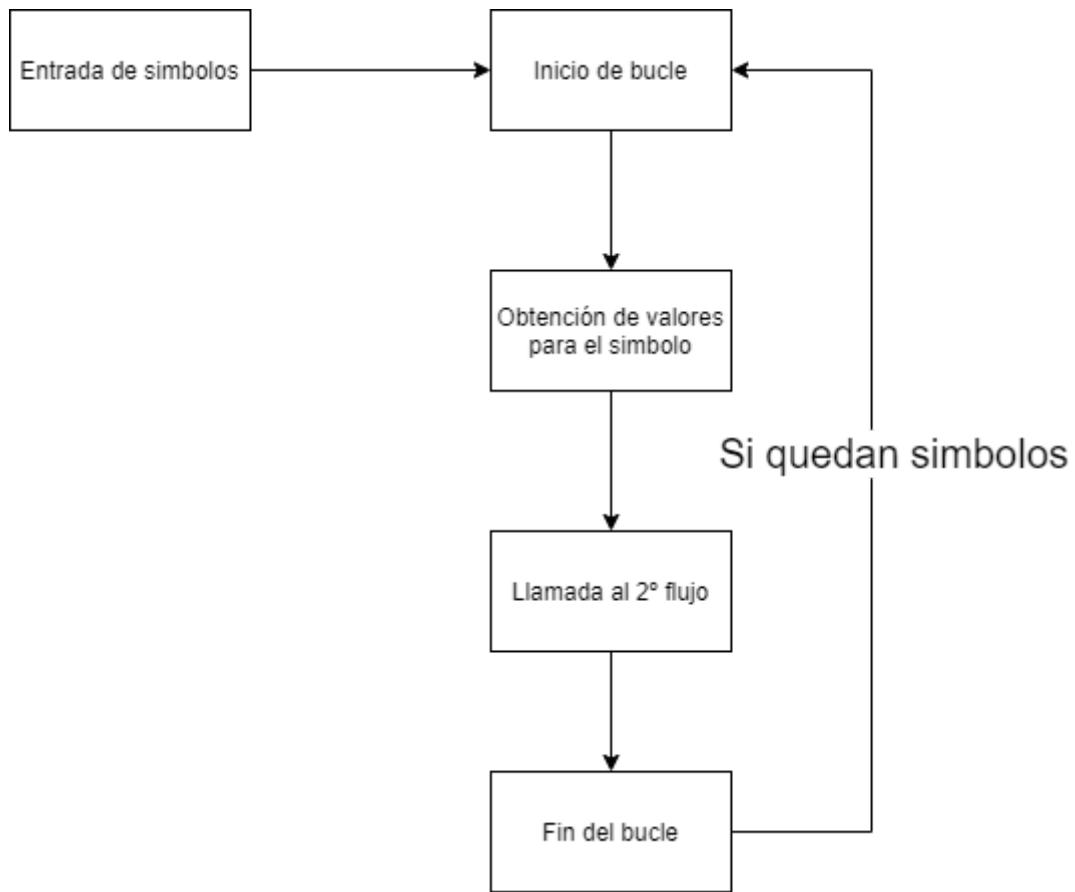
### **3.2 Diseño del Sistema**

El prototipado se realizará utilizando workflows de tareas. Este enfoque permitirá utilizar directamente KNIME, y tomar como punto de partida un workflow existente<sup>[18]</sup>. Este workflow modela un mercado y hace una predicción del valor de diferentes símbolos.

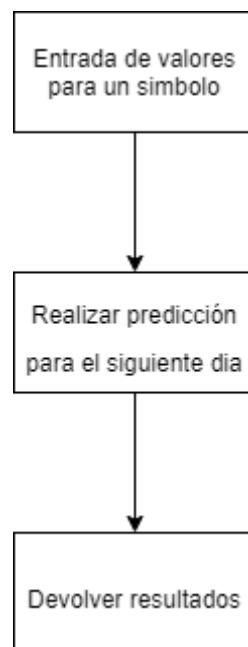
Existen diferentes formas de predecir el valor que el símbolo tendrá el siguiente día, pero esto no va a hacer que el diseño cambie, por lo que se va a explicar el diseño general que se ha utilizado para cualquier ocasión.

El diseño consiste en dos flujos de trabajo, el primer flujo de trabajo (figura 1) obtiene los valores que toman los diferentes símbolos cada día durante el periodo de tiempo indicado, una vez que se obtienen los valores de un símbolo estos se procesan en el segundo flujo (figura 2) de manera que se realiza una predicción de cuál será su valor el siguiente día. Finalmente, una vez se obtienen las predicciones de todos los

símbolos, se escriben estas predicciones en un archivo ordenadas por mayor porcentaje de cambio.



**Figura 1:** Diagrama del primer flujo.



**Figura 2:** Diagrama del segundo flujo.

### **3.2.1 Métodos de Análisis Elegidos**

Uno de los principios básicos del análisis técnico es que ciertos patrones de precios reaparecen constantemente y tienden a producir los mismos resultados. Otro principio básico del análisis técnico dice que "los precios se mueven en tendencias". Los analistas creen firmemente que estas tendencias pueden identificarse de manera oportuna y utilizarse para generar ganancias y limitar las pérdidas. En consecuencia, el seguimiento de tendencias es la estrategia de sincronización de mercado más extendida. Específicamente, cuando los precios de las acciones tienen una tendencia al alza o a la baja, es hora de comprar o vender la acción. Aunque el seguimiento de tendencias es muy simple en concepto, su realización práctica es complicada. Las medias móviles se utilizan para "suavizar" el precio de las existencias para resaltar la tendencia subyacente. Esta metodología de detección de la tendencia mediante la filtración de los ruidos resulta del análisis de series de tiempo donde se utilizan medias móviles centradas (o bilaterales). La misma metodología se aplica para predecir el movimiento futuro del precio de las acciones.

#### **Regresión Lineal<sup>[19]</sup>**

En estadística, la regresión lineal o el ajuste lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente  $Y$ , las variables independientes  $X_i$  y un término aleatorio  $\epsilon$ . El modelo que expresa esto es el representado en la ecuación 1.

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

**Ecuación 1:** Modelo que expresa la regresión lineal.

donde:

$Y_t$ : variable dependiente.

$X_1, X_2, \dots, X_p$ : variables independientes.

$\beta_0, \beta_1, \beta_2, \dots, \beta_p$ : parámetros, miden la influencia que las variables independientes tienen sobre la variable dependiente.

$\beta_0$  es la intersección o término "constante", las  $\beta_i$  ( $i > 0$ ) son los parámetros predeterminados a cada variable independiente, y  $p$  es el número de parámetros independientes a tener en cuenta en la regresión.

## Exponential Moving Average (EMA)<sup>[17, pág. 30]</sup>

Antes de explicar las medias móviles utilizadas en este trabajo, se va a explicar la media móvil en la que se basan las tres medias móviles utilizadas para construir sus fórmulas.

La **EMA** es una media móvil ponderada exponencialmente. Se trata de la media aritmética de los n valores anteriores con factores de ponderación que decrecen exponencialmente. La ponderación para cada punto de datos más antiguo decrece exponencialmente, nunca llegando a cero. La fórmula para calcular esta media móvil es la que aparece en la ecuación 2, donde  $0 < \lambda \leq 1$  es un factor de descomposición. Cuando  $\lambda < 1$ , el promedio móvil ponderado exponencialmente asigna mayores pesos a los precios más recientes. Al variar el valor de  $\lambda$ , se puede ajustar la ponderación para dar mayor o menor peso al precio más reciente.

$$EMA_t(\lambda, n) = \frac{P_t + \lambda P_{t-1} + \lambda^2 P_{t-2} + \cdots + \lambda^{n-1} P_{t-n+1}}{1 + \lambda + \lambda^2 + \cdots + \lambda^{n-1}} = \frac{\sum_{i=0}^{n-1} \lambda^i P_{t-i}}{\sum_{i=0}^{n-1} \lambda^i}$$

**Ecuación 2:** Fórmula para calcular EMA.

## Double Exponential Moving Average (DEMA)<sup>[17, pág 43]</sup>

Una media móvil exponencial doble (**DEMA**) es una media móvil mixta propuesta por Mulloy. La idea original de Mulloy era reducir el tiempo de retraso de la EMA normal colocando más peso (que en la EMA normal) en los precios más recientes. El valor de DEMA se calcula de acuerdo con la fórmula de la ecuación 3.

$$DEMA = 2 \times EMA_n(P) - EMA_n(EMA_n(P)).$$

**Ecuación 3:** Fórmula para calcular DEMA.

Para reducir el tiempo de retraso promedio, **DEMA** empuja el valor del promedio móvil más cerca del valor del último precio de cierre.

## Triple Exponential Moving Average<sup>[17, pág 43]</sup>

Posteriormente, Mulloy propuso una Media Móvil Exponencial Triple (**TEMA**) con un tiempo de retraso incluso menor que el de DEMA. El valor de TEMA se calcula de acuerdo con la fórmula de la ecuación 4.

$$TEMA = 3 \times EMA_n(P) - 3 \times EMA_n(EMA_n(P)) + EMA_n(EMA_n(EMA_n(P))).$$

**Ecuación 4:** Fórmula para calcular TEMA.

### Zero Lag Exponential Moving Average<sup>[17, pág 39]</sup>

La idea detrás de la construcción de **ZLEMA** es la siguiente: la  $EMA_t(n, P)$  normal tiene un tiempo de retraso promedio de  $\frac{n-1}{2}$  y su valor difiere del valor del último precio de cierre  $P_t$  debido a la naturaleza retardada del promedio móvil. Esta discrepancia entre el último precio de cierre y el valor de  $EMA_t(n, P)$  puede ser estimada como  $P_t - \text{Lag}_{\frac{n-1}{2}}(P_t)$  donde  $\text{Lag}_j(P_t) = P_{t-j}$ . En otras palabras,  $\text{Lag}_j(P_t)$  es el valor del precio de la serie temporal en el tiempo  $t-j$ .

La solución propuesta por Ehlers y Way es suavizar tanto el precio como la discrepancia estimada, quedando la fórmula para calcular esta media móvil la observable en la ecuación 5.

$$ZLEMA = EMA_n\left(2P - \text{Lag}_{\frac{n-1}{2}}(P)\right) = 2 \times EMA_n(P) - EMA_n\left(\text{Lag}_{\frac{n-1}{2}}(P)\right).$$

**Ecuación 5:** Fórmula para calcular ZLEMA.

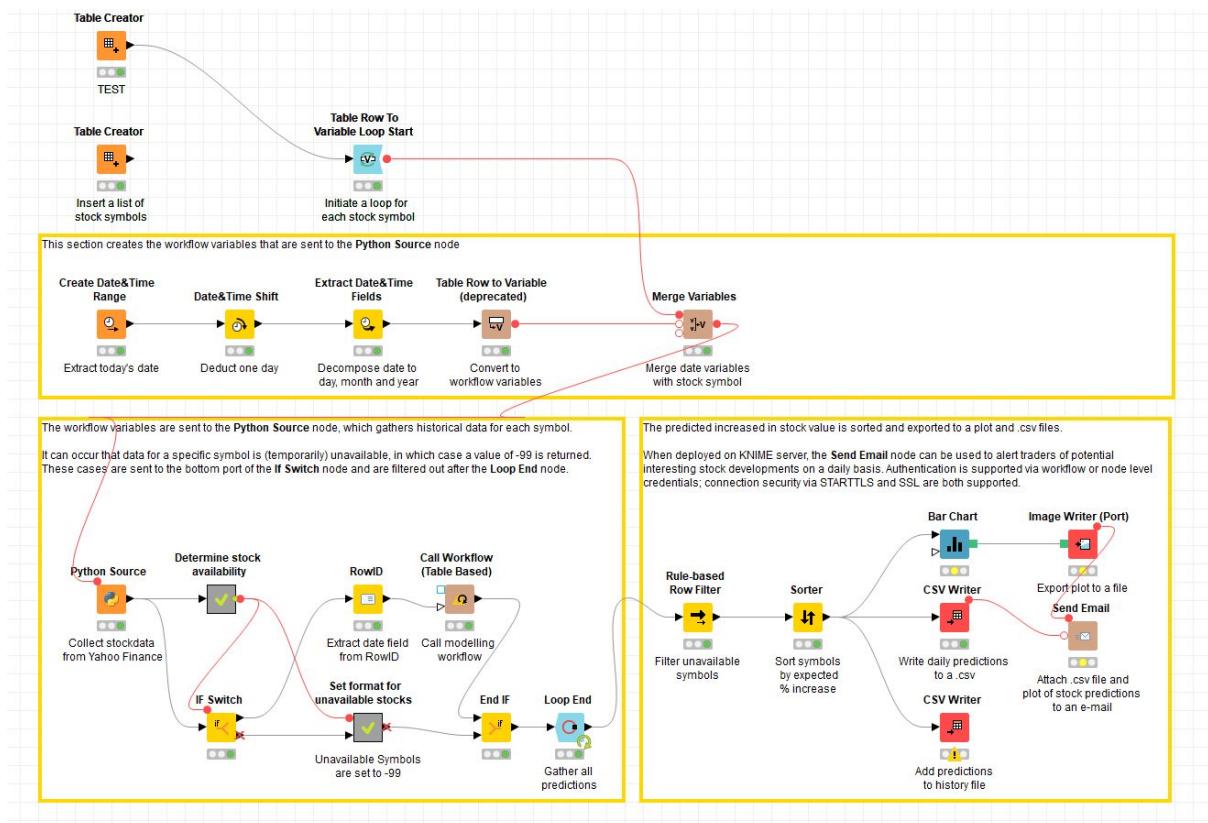
## 3.3 Implementación

En este apartado se va a explicar la implementación de ambos flujos de trabajo para las diferentes formas de predecir los valores.

### General

En primer lugar se utiliza el nodo *Table Creator* para construir una tabla que solo tiene la columna *Symbols*. A continuación, se utiliza el nodo *Table Row To Variable Loop Start* mediante el cual se inicia un bucle que tiene tantas ejecuciones como símbolos tenga la tabla. Al comienzo del bucle se utiliza el node *Python Source* el cual utiliza el módulo **pandas\_datareader** para obtener los valores que dicho símbolo tiene cada día durante un determinado periodo de tiempo, una vez se tienen estos valores se llama a un segundo flujo dependiendo de la forma en la que se calcule la predicción (Regresión Lineal, Double Exponential Moving Average, Triple Exponential Moving Average, Zero Lag Exponential Moving Average).

En la figura 1 se puede observar el flujo del workflow general.



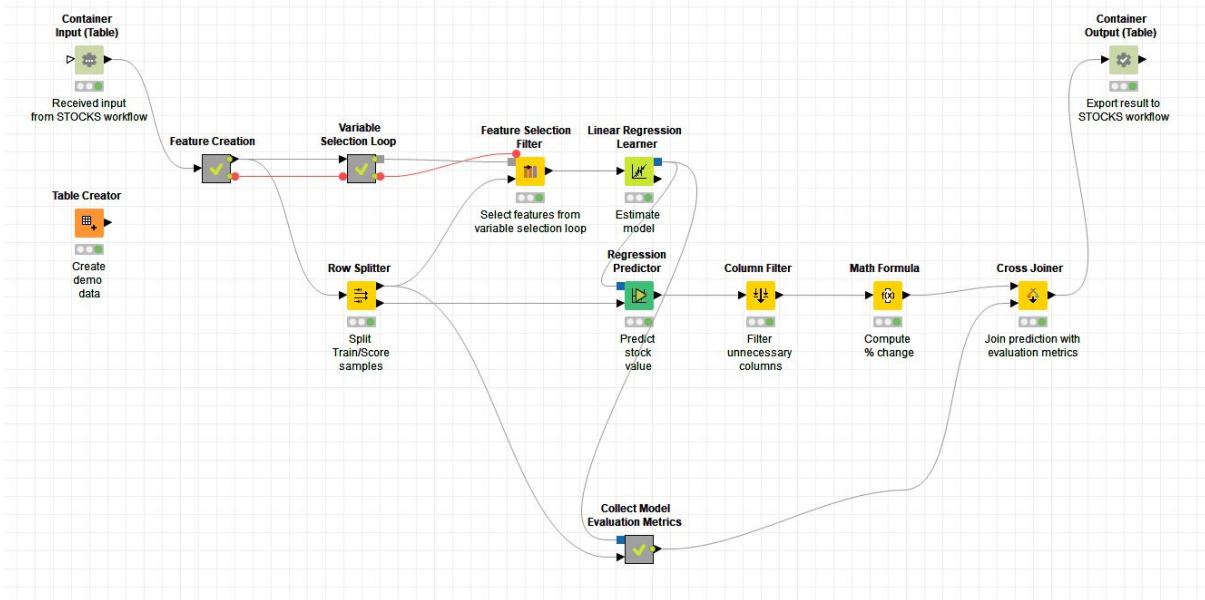
**Figura 1:** Esquema del workflow general.

## Regresión Lineal

Una de las formas mediante las que se ha tratado de predecir el valor del día siguiente de un símbolo es mediante una regresión lineal, para esto se han marcado todos los valores del símbolo como **TRAIN**, de esta forma pertenecerán al conjunto de entrenamiento y se ha añadido una nueva fila que se ha marcado como **SCORE**, esta fila pertenece al siguiente día y el valor del símbolo al cierre es el valor a predecir.

Antes de realizar la regresión lineal se realiza otra regresión lineal para determinar que variables son las mejores para construir el modelo de predicción del valor. Una vez se obtienen estas variables se utiliza el nodo *Linear Regression Learner* para que aprenda de los datos en el conjunto de entrenamiento y posteriormente se utiliza el nodo *Regression Predictor* para intentar predecir el valor que tomará el símbolo el próximo día.

En la figura 2 se puede observar el flujo del workflow para realizar la regresión lineal.

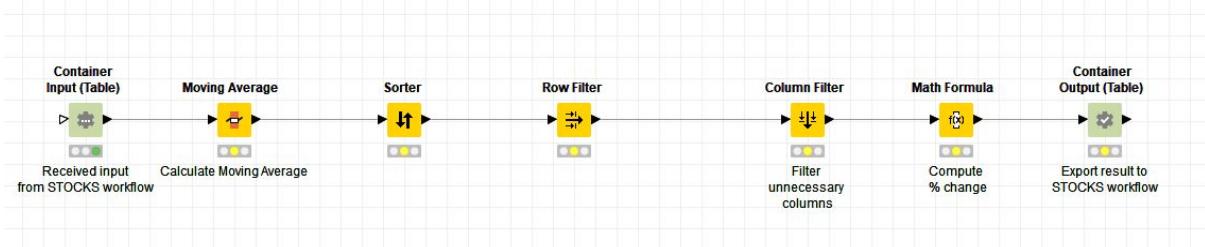


**Figura 2:** Esquema del workflow de la regresión lineal.

### Double Exponential Moving Average

Para el cálculo de Double Exponential Moving Average se utiliza el nodo *Moving Average* el cual tiene integrada la fórmula para realizar esta media móvil, a continuación se ordena utilizando el nodo *Sorter* para ordenar por fecha descendente y así quedarnos con la primera fila, que es la que contendrá el valor de la media móvil para el próximo día. Se ha implementado esta media móvil con una ventana de 15 días para poder comparar con el modelo inicial.

En la figura 3 se puede observar el flujo del workflow para realizar el cálculo de DEMA.



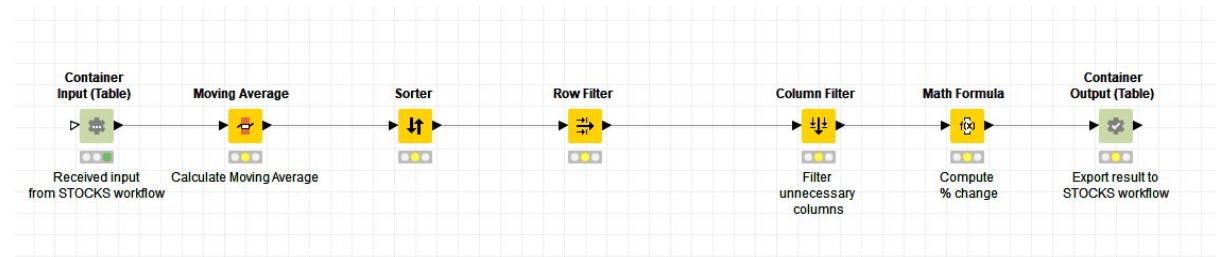
**Figura 3:** Esquema del workflow para calcular DEMA.

### Triple Exponential Moving Average

Para el cálculo de Triple Exponential Moving Average se utiliza el nodo *Moving Average* el cual tiene integrada la fórmula para realizar esta media móvil, a continuación se ordena utilizando el nodo *Sorter* para ordenar por fecha descendente y así quedarnos con la primera fila, que es

la que contendrá el valor de la media móvil para el próximo día. Se ha implementado esta media móvil con una ventana de 15 días para poder comparar con el modelo inicial.

En la figura 4 se puede observar el flujo del workflow para realizar el cálculo de TEMA. Se puede observar que es el mismo flujo que el de la figura 3 pero con la peculiaridad de que el nodo *Moving Average* está configurado para calcular TEMA en lugar de DEMA.

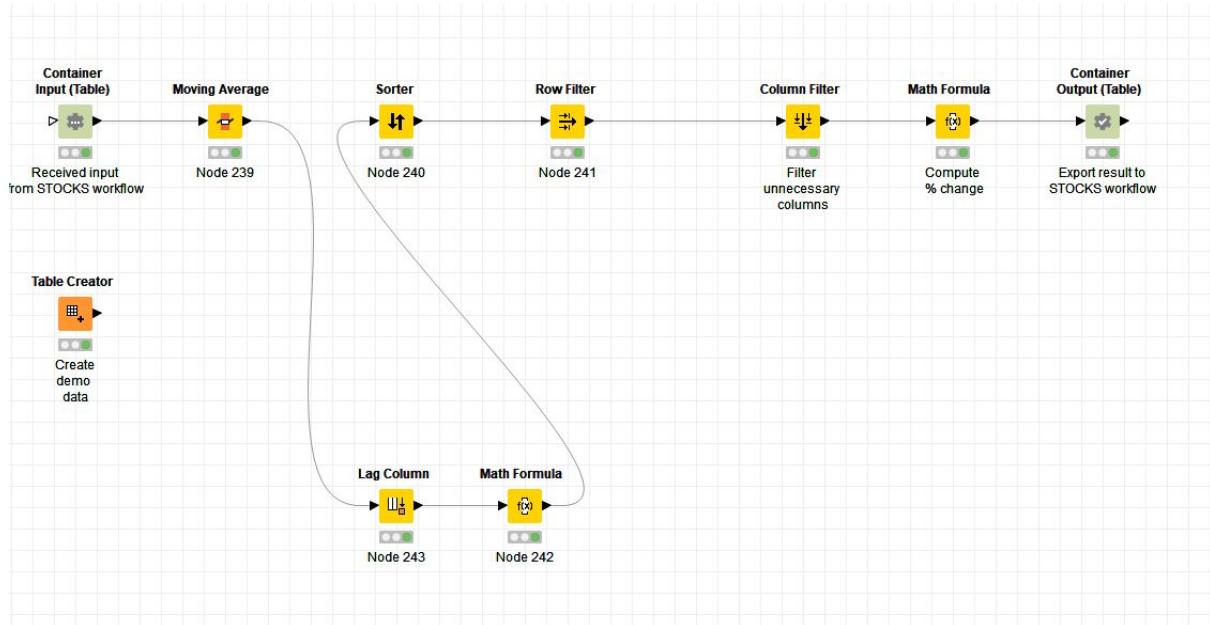


**Figura 4:** Esquema del workflow para calcular TEMA.

### Zero Lag Exponential Moving Average

Para el cálculo de Zero Lag Exponential Moving Average se utiliza el nodo *Moving Average* el cual tiene integrada la fórmula para calcular la Simple Exponential que es necesaria para el cálculo de esta media móvil. A continuación se utiliza el nodo *Lag Column* para poder utilizar el valor de la media móvil en  $(n - 1)/2$  donde n es 15, necesario para realizar el cálculo de esta media móvil. Por último se utiliza el nodo *Math Formula* para calcular esta media móvil ( $2 * \text{Simple Moving Average} - \text{Simple Moving Average} (-7)$ ) y se ha ordenado utilizando el nodo *Sorter* para ordenar por fecha descendente y así quedarnos con la primera fila, que es la que contendrá el valor de la media móvil para el próximo día.

En la figura 5 se puede observar el flujo del workflow para realizar el cálculo de TEMA.



**Figura 5:** Esquema del workflow para calcular ZLEMA.

## 4. Experimentos

Los experimentos han consistido en la obtención de datos de diferentes símbolos entre el 01/01/2016 y el 31/12/2017 e intentar predecir que va a ocurrir el siguiente día, en una semana, mes, tres meses y año utilizando las diferentes herramientas matemáticas (Regresión lineal, DEMA, TEMA y ZLEMA).

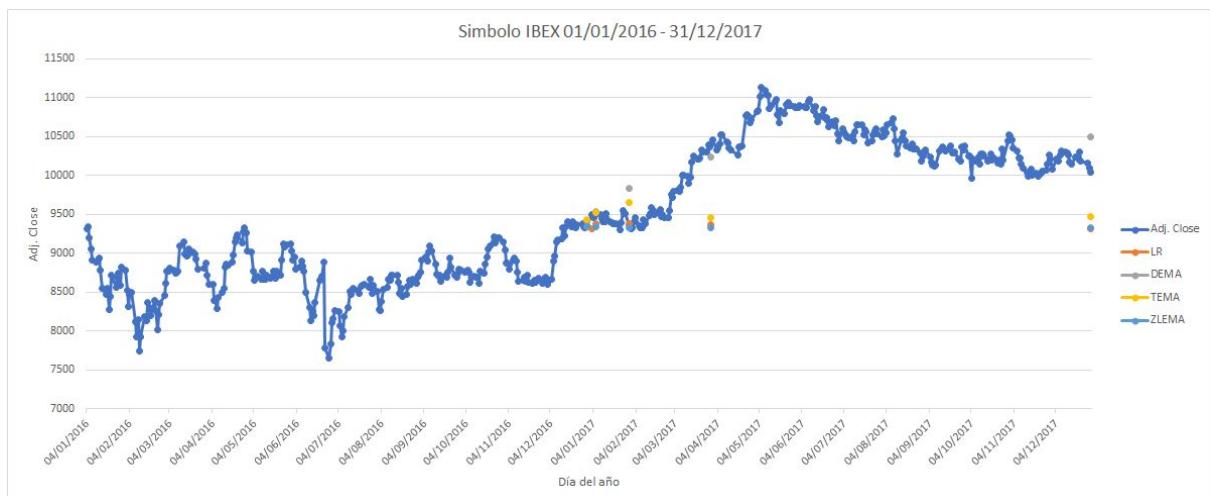
Se ha hecho uso de este año ya que fue un año estable en la economía mundial, es decir, no existió ningún tipo de crisis ni pandemia mundial que alterara los mercados debido a factores externos.

Después de haber predecido los datos se han descargado los datos de los símbolos utilizados entre el 01/01/2016 y 31/12/2017 para poder calcular el error que han tenido los diferentes métodos probados para las diferentes predicciones.

Los errores en este caso se presentan en las gráficas como tanto por uno.

### IBEX

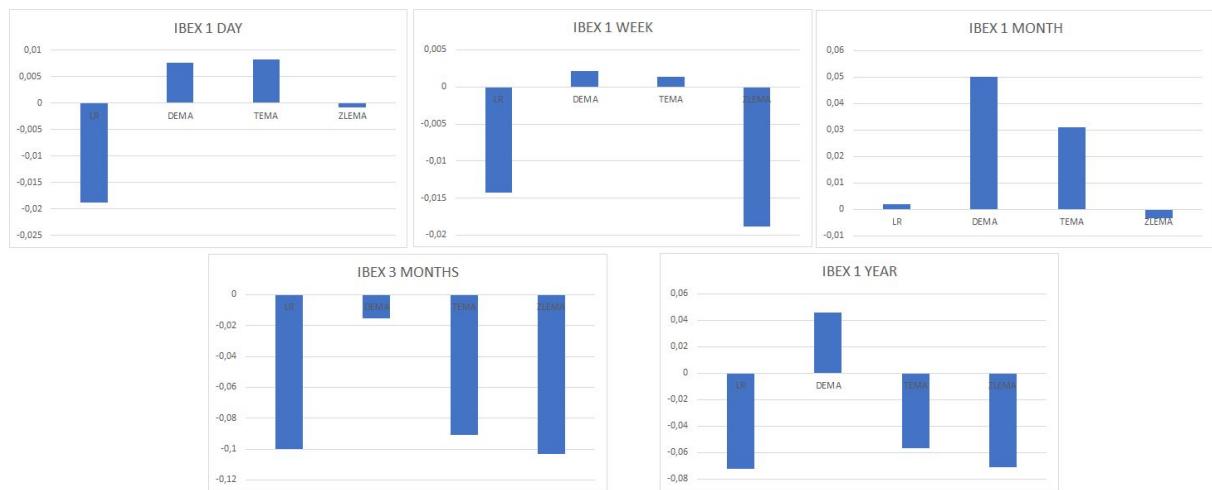
Se ha creado una gráfica (figura 6) en la que se muestran los valores del cierre del mercado entre las fechas utilizadas para realizar los experimentos, en esta gráfica se han insertado también puntos para marcar las predicciones de las diferentes herramientas matemáticas a un día, una semana, un mes, tres meses y un año.



**Figura 6:** Valores de cierre para el símbolo IBEX junto con los valores predichos (01/01/2016 - 31/12/2017)

El porcentaje de error para cada herramienta matemática en cada predicción se ha desglosado en las gráficas de la figura 7. De las gráficas de esta figura se puede ver que ZLEMA es la herramienta que mejor funciona hasta un mes, dando errores inferiores al 2%.

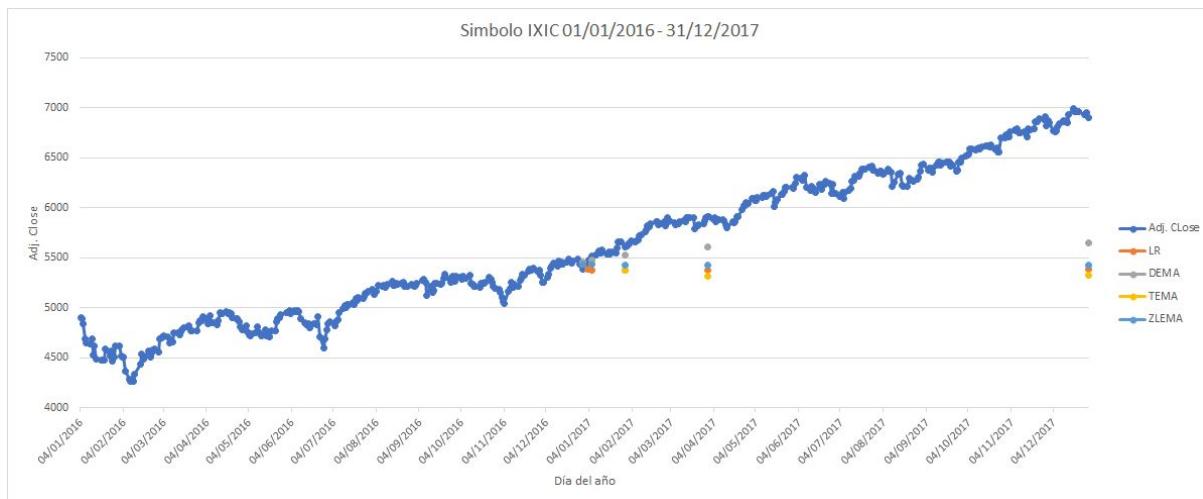
Por otro lado DEMA es la herramienta que mejor predice en global siendo el caso de mes en el que realiza una estimación peor (error del 5%), pero dando unas predicciones a largo plazo bastante interesantes aunque con un error considerable.



**Figura 7:** Porcentajes de error para las diferentes herramientas matemáticas en las diferentes predicciones de IBEX.

## IXIC

Se ha creado una gráfica (figura 8) en la que se muestran los valores del cierre del mercado entre las fechas utilizadas para realizar los experimentos, en esta gráfica se han insertado también puntos para marcar las predicciones de las diferentes herramientas matemáticas a un día, una semana, un mes, tres meses y un año.

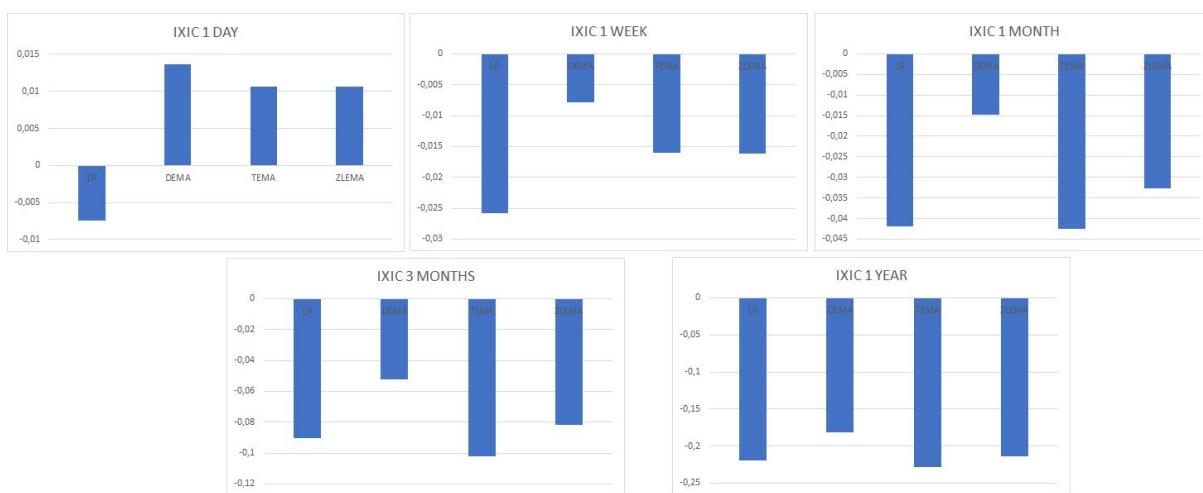


**Figura 8:** Valores de cierre para el símbolo IXIC junto con los valores predichos (01/01/2016 - 31/12/2017)

El porcentaje de error para cada herramienta matemática en cada predicción se ha desglosado en las gráficas de la figura 9. De las gráficas de esta figura se puede observar que la regresión lineal es la herramienta que mejor predice a un día.

Se puede que DEMA es la herramienta que mejor funciona a nivel global, dando en este caso porcentajes de error muy buenos hasta un mes (inferiores a 2% de error), pero con errores mucho más significativos a partir de 3 meses donde se llega hasta el 5% de error.

En este caso es destacable que las predicciones a un año para cualquier herramienta se alejan mucho de la realidad con predicciones rondando el 20% de error.



**Figura 9:** Porcentajes de error para las diferentes herramientas matemáticas en las diferentes predicciones de IXIC.

## GSPC

Se ha creado una gráfica (figura 10) en la que se muestran los valores del cierre del mercado entre las fechas utilizadas para realizar los experimentos, en esta gráfica se han insertado también puntos para marcar las predicciones de las diferentes herramientas matemáticas a un día, una semana, un mes, tres meses y un año.



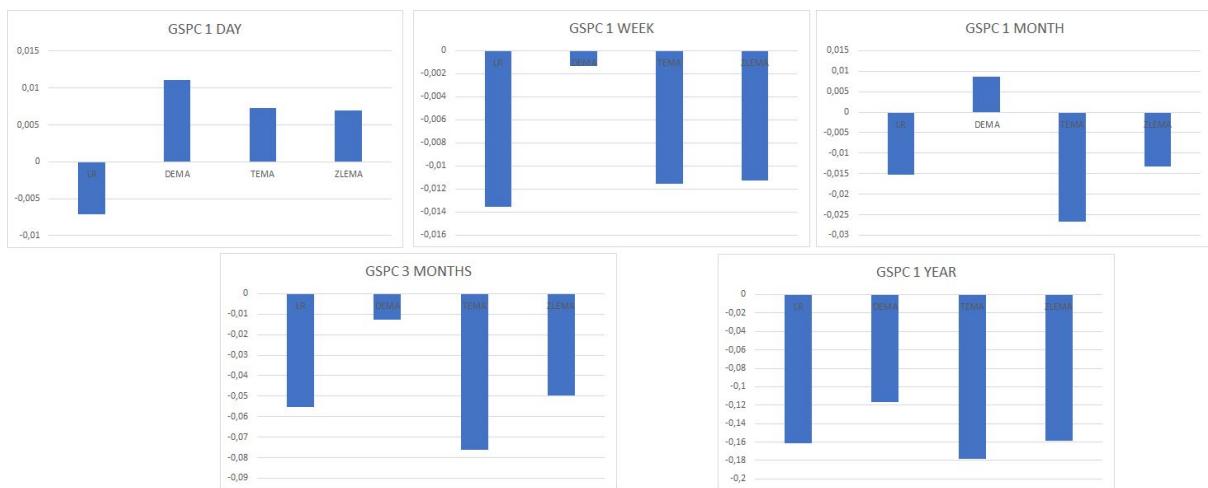
**Figura 10:** Valores de cierre para el símbolo GSPC junto con los valores predichos (01/01/2016 - 31/12/2017)

El porcentaje de error para cada herramienta matemática en cada predicción se ha desglosado en las gráficas de la figura 11. De las gráficas de esta figura se puede observar que la regresión lineal es la herramienta que mejor predice a un solo día.

Se puede observar que DEMA es la herramienta que mejor predice a nivel global dando resultados muy buenos hasta los tres meses, donde nunca llega a superar el 1% de error.

En este caso ZLEMA y la regresión lineal también proporcionan resultado bastante buenos hasta un mes, con predicciones que no superan el 1,5% de error.

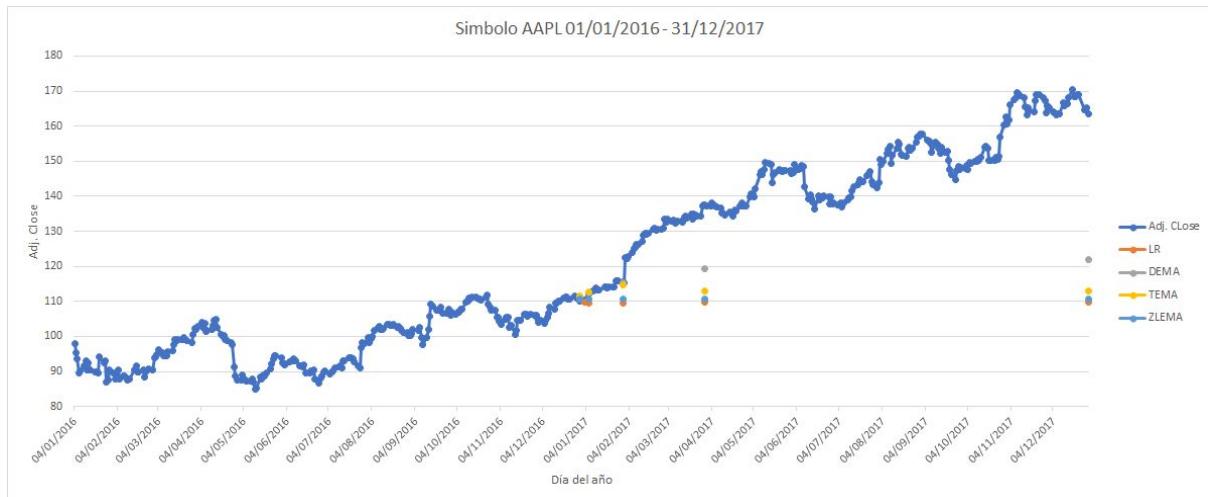
En este caso es destacable que las predicciones a un año para cualquier herramienta se alejan mucho de la realidad.



**Figura 11:** Porcentajes de error para las diferentes herramientas matemáticas en las diferentes predicciones de GSPC.

## AAPL

Se ha creado una gráfica (figura 12) en la que se muestran los valores del cierre del mercado entre las fechas utilizadas para realizar los experimentos, en esta gráfica se han insertado también puntos para marcar las predicciones de las diferentes herramientas matemáticas a un día, una semana, un mes, tres meses y un año.

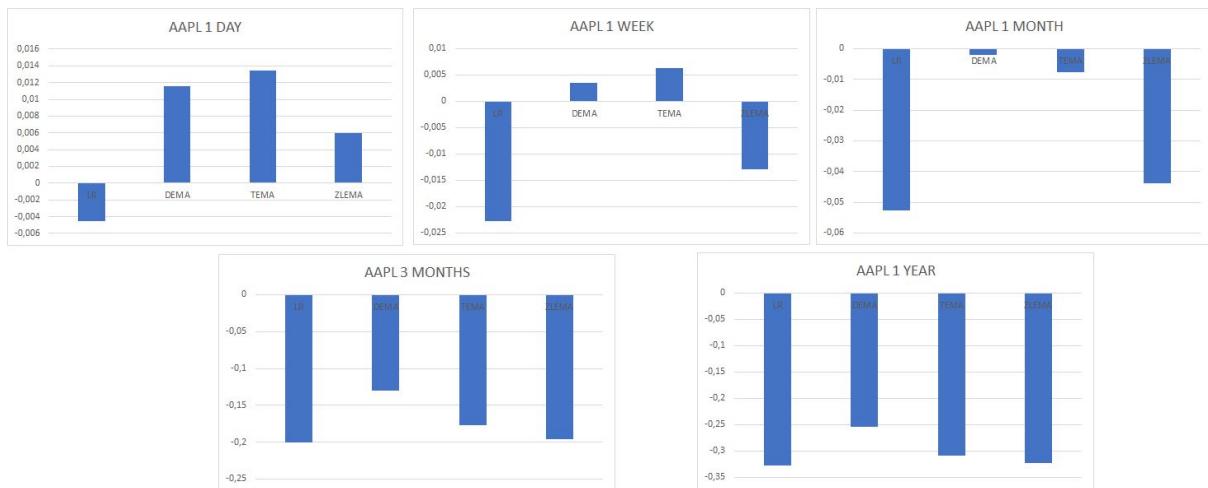


**Figura 12:** Valores de cierre para el símbolo AAPL junto con los valores predichos (01/01/2016 - 31/12/2017)

El porcentaje de error para cada herramienta matemática en cada predicción se ha desglosado en las gráficas de la figura 13. De las gráficas de esta figura se puede extraer que la regresión lineal es la herramienta que mejor predice a un día.

Para predicciones en ventanas temporales más grandes nos encontramos con que DEMA y TEMA son las mejores herramientas que predicen hasta un mes, con errores en el peor caso próximos al 1%.

Las predicciones para este símbolo a tres meses y un año distan mucho de la realidad, con errores superiores al 10% para todas las herramientas.



**Figura 13:** Porcentajes de error para las diferentes herramientas matemáticas en las diferentes predicciones de AAPL.

## AMZN

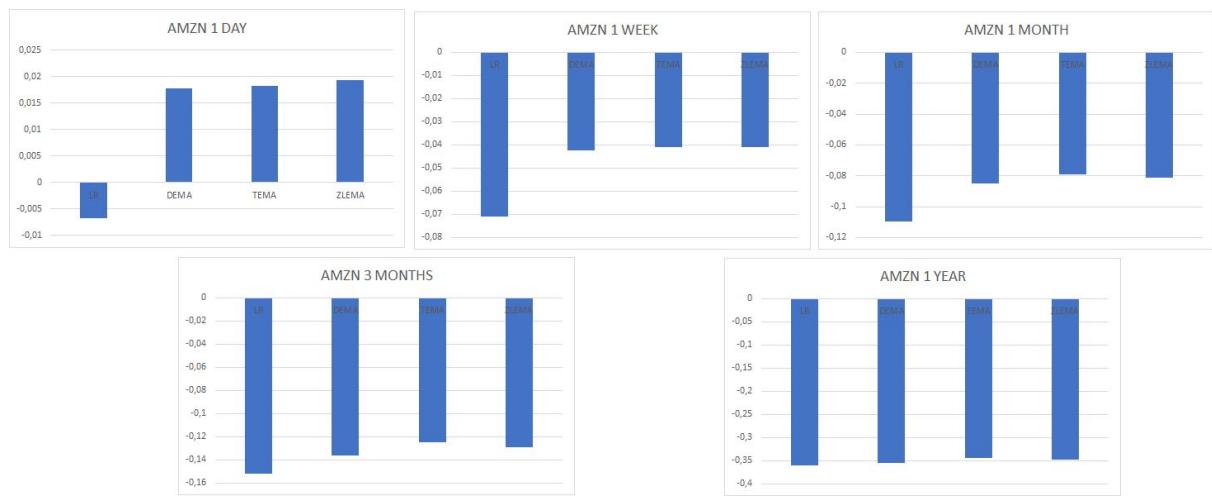
Se ha creado una gráfica (figura 14) en la que se muestran los valores del cierre del mercado entre las fechas utilizadas para realizar los experimentos, en esta gráfica se han insertado también puntos para marcar las predicciones de las diferentes herramientas matemáticas a un día, una semana, un mes, tres meses y un año.



**Figura 14:** Valores de cierre para el símbolo AMZN junto con los valores predichos (01/01/2016 - 31/12/2017)

El porcentaje de error para cada herramienta matemática en cada predicción se ha desglosado en las gráficas de la figura 15. De las gráficas de esta figura se puede ver que la herramienta que mejor predice a un día es la regresión lineal.

En este caso el resto de predicciones son malas ya que para una semana el error de la predicción para cualquier herramienta es como mínimo del 4%, incrementándose este error conforme las predicciones son más lejanas en el tiempo.



**Figura 15:** Porcentajes de error para las diferentes herramientas matemáticas en las diferentes predicciones de AMZN.

## GOOG

Se ha creado una gráfica (figura 16) en la que se muestran los valores del cierre del mercado entre las fechas utilizadas para realizar los experimentos, en esta gráfica se han insertado también puntos para marcar las predicciones de las diferentes herramientas matemáticas a un dia, una semana, un mes, tres meses y un año.



**Figura 16:** Valores de cierre para el símbolo GOOG junto con los valores predichos (01/01/2016 - 31/12/2017)

El porcentaje de error para cada herramienta matemática en cada predicción se ha desglosado en las gráficas de la figura 17. De las gráficas de esta figura se puede observar que todas las herramientas tienen un error similar para un día.

En este caso DEMA realiza las mejores predicciones hasta 3 meses, con un error máximo del 3%. Cabe destacar también que TEMA y ZLEMA tienen buenos resultados también a la hora de predecir hasta un mes, con errores que no superan el 3.5%, pero a partir de este punto los predicciones comienzan a alejarse mucho de la realidad.

Las predicciones para este símbolo a un año distan mucho de la realidad, con errores superiores al 20% para todas las herramientas.



**Figura 17:** Porcentajes de error para las diferentes herramientas matemáticas en las diferentes predicciones de GOOG.

## Conclusiones de los experimentos

Se observa que la regresión lineal es mejor a un dia para todos los símbolos excepto para el IBEX, por lo que se puede deducir que para operaciones a corto plazo se utilizará esta herramienta.

DEMA tiene un error máximo para GOOG del 3% hasta los 3 meses, para AAPL del 1% hasta 1 mes, para GSPC del 1% hasta 3 meses, para IXIC del 2% hasta 1 mes y es la herramienta que mejor predice a largo tiempo para el IBEX. ZLEMA tiene un error máximo para IBEX del 2% hasta 1 mes, para GSPC del 1.5% hasta 1 mes y para GOOG del 3.5% hasta 1 mes. TEMA tiene un error máximo para AAPL de 1% hasta 1 mes, para GOOG del 3.5% hasta 1 mes.

Las predicciones a un año de todas las herramientas para cualquier símbolo son malas, por lo que se puede concluir que no se debería usar ninguna de estas herramientas para predecir valores más allá de un año.

ZLEMA puede funcionar bien en algún caso pero DEMA es la que mejor predice a lo largo del tiempo y para todos los símbolos analizados menos AMZN, en el que ninguna herramienta ha sido capaz de hacer una predicción acertada. Por esto nos quedamos con DEMA como herramienta a utilizar a la hora de realizar predicciones que no sean a 1 día, aunque hay que tener en cuenta que a partir de los 3 meses empieza a dar resultados peores, por lo que se recomienda a cualquier persona que utilice esta herramienta realizar predicciones a 1 mes y advirtiendo de que las predicciones a 3 meses no son tan buenas.

## **5. Conclusiones y trabajo futuro**

En esta sección se recogen las conclusiones y trabajo futuro que se derivan de este trabajo. Adicionalmente, se recoge la información relacionada con el propio desarrollo del trabajo

### **5.1 Conclusiones sobre el trabajo**

Finalmente se ha conseguido implementar un Sistema de Ayuda a la Toma de Decisiones para la Gestión de Carteras de Bolsa que permite decidir cuales son los símbolos que mejor van a funcionar para un posible inversor utilizando diferentes herramientas matemáticas y considerando distintos horizontes de inversión.

Los mercados de valores son estudiados por un número muy elevado de expertos, y todos tratan de crear un modelo que se adapte lo más posible al comportamiento de estos mercados para poder así poder sacar rendimiento económico de estos. A día de hoy existen muchas herramientas que intentan predecir estos mercados, y cada experto utiliza las herramientas que mejor funcionen bajo su criterio.

El sistema de prototipado rápido ha resultado muy útil a la hora de implementar varias herramientas matemáticas clásicas y variantes evolucionadas. El entorno de prototipado KNIME proporciona muchas funcionalidades que habría que haber codificado si se hubiera utilizado una herramienta de programación convencional, lo que ha permitido ahorrar todo el tiempo de programación y depuración.

Los experimentos han demostrado que, considerando un contexto de inversión sin eventos extraordinarios, existen herramientas que pueden predecir de manera bastante acertada el comportamiento de los mercados de valores, lo que es interesante para cualquier tipo de inversor.

### **5.2 Dedicación**

En esta sección se va a incluir una tabla (tabla 1) en el cual se indican las diferentes tareas que se han hecho para la realización del Trabajo de Fin de Grado así como las horas dedicadas a cada tarea. Las horas de redacción de la memoria se han incluido en las diferentes tareas de la tabla.

Descripción del hito	Horas	Inicio	Fin
<b>Introducción</b>			
Motivación y Objetivos	5	12/02/2020	14/02/2020
Alcance	3	17/06/2020	17/06/2020
Contenido del documento	3	17/06/2020	17/06/2020
<b>Estado del arte</b>			
Lectura de libros sobre mercados de valores	19	20/02/2020	26/02/2020
Desarrollo funcionamiento mercados de valores	3,5	27/02/2020	28/02/2020
Lectura libros sobre metodologías y herramientas de análisis	26	02/03/2020	07/03/2020
Desarrollo metodologías y Herramientas de Análisis	4,5	08/03/2020	11/03/2020
Desarrollo fuentes de datos y entornos software de prototipado rápido	3	14/03/2020	16/03/2020
<b>Propuesta de solución</b>			
Requisitos	2	20/03/2020	21/03/2020
Estudio del diseño inicial y diseño del sistema	25	26/03/2020	02/04/2020
Obtener datos de diversas fuentes	47	10/04/2020	13/04/2020

Implementación de los métodos de análisis	22	16/04/2020	30/04/2020
Pruebas del funcionamiento de la implementación	5	02/05/2020	03/05/2020
<b>Experimentos</b>			
Diseño de los experimentos para cada herramienta	35	20/05/2020	30/05/2020
Diseño de workflows que calculen errores para cada herramienta	23	02/06/2020	04/06/2020
Diseño de gráficas comparativas	11	07/06/2020	08/06/2020
Conclusiones de los experimentos	6	10/06/2020	11/06/2020
<b>Conclusiones y trabajo futuro</b>			
Conclusiones	3	14/06/2020	15/06/2020
Trabajo futuro	2	14/06/2020	15/06/2020
<b>Reuniones</b>			
Reunión	0,5	18/10/2019	18/10/2019
Reunión	0,5	13/02/2020	13/02/2020
Reunión	0,5	11/03/2020	11/03/2020
Reunión	0,5	26/03/2020	26/03/2020

Reunión	0,5	10/04/2020	10/04/2020
Reunión	0,5	16/04/2020	16/04/2020
Reunión	0,5	24/04/2020	24/04/2020
Reunión	0,5	08/05/2020	08/05/2020
Reunión	0,5	29/05/2020	29/05/2020
Reunión	0,5	11/06/2020	11/06/2020
Reunión	0,5	16/06/2020	16/06/2020
Reunión	0,5	22/06/2020	22/06/2020
<b>Memoria</b>			
Revisión y repaso	5	14/06/2020	24/06/2020

**Tabla 1:** Tareas realizadas junto con las horas y fechas de realización.

### 5.3 Trabajo futuro

Utilizar un algoritmo de aprendizaje, como por ejemplo un árbol de decisión, para tratar de introducir algún criterio externo extra que refuerce las predicciones. De esta forma, podría utilizarse la predicción inicial aportada por el prototipo ya desarrollado basada en la regresión lineal o alguna de las medias móviles implementadas, para luego añadir progresivamente otros que refuercen las hipótesis que se planteen y así robustecer los modelos iniciales.

# Bibliografía

- [1] "Mercado De Valores." *Wikipedia*, Wikimedia Foundation, 13 Jan. 2020, [es.wikipedia.org/wiki/Mercado\\_de\\_valores](https://es.wikipedia.org/wiki/Mercado_de_valores).
- [2] Granel, María. "¿Cuál Es El Funcionamiento De La Bolsa De Valores?" *Rankia*, Rankia, 21 Jan. 2020, [www.rankia.cl/blog/analisis-ipsa/4031929-cual-funcionamiento-bolsa-valores](https://www.rankia.cl/blog/analisis-ipsa/4031929-cual-funcionamiento-bolsa-valores).
- [3] "SPSS." *Wikipedia*, Wikimedia Foundation, 21 Nov. 2019, [es.wikipedia.org/wiki/SPSS](https://es.wikipedia.org/wiki/SPSS).
- [4] "R (Lenguaje De Programación)." *Wikipedia*, Wikimedia Foundation, 29 Feb. 2020, [es.wikipedia.org/wiki/R\\_\(lenguaje\\_de\\_programaci%C3%B3n\)](https://es.wikipedia.org/wiki/R_(lenguaje_de_programaci%C3%B3n)).
- [5] "Yahoo Finanzas: Bolsa De Valores En Directo, Cotizaciones, Noticias Empresariales y Financieras." *Yahoo! Finance*, Yahoo!, [es.finance.yahoo.com](https://es.finance.yahoo.com).
- [6] "COTIZACIONES BOLSA ESPAÑOLA TIEMPO REAL" *COTIZACIONES BOLSA ESPAÑOLA TIEMPO*, [www.iberbolsa.com](https://www.iberbolsa.com).
- [7] "Bolsa,Ibex,Prima Riesgo, Euribor, Euro Dólar Hoy,Bono." *Infobolsa*, [www.infobolsa.es](https://www.infobolsa.es).
- [8] "Finanzas, Noticias y Bolsa De Valores." *Investing.com Español*, [es.investing.com](https://es.investing.com).
- [9] Universal Market Predictor, LLC. "The Universal Market Predictor Index (UMPI): The First Reliable Market Predicting Tool." *UMPI*, [www.umpindex.com](https://www.umpindex.com).

- [10] *InteliCharts*, [intelicharts.com](http://intelicharts.com).
- [11] Berthold, Michael. "Open for Innovation." *KNIME*, [www.knime.com](http://www.knime.com).
- [12] "RStudio: Open Source & Professional Software for Data Science Teams." *RStudio, PBC*, 7 Jan. 2020, [rstudio.com](http://rstudio.com).
- [13] "Best Data Science & Machine Learning Platform." *RapidMiner*, [rapidminer.com](http://rapidminer.com).
- [14] "Anaconda Python/R Distribution - Free Download." *Anaconda*, [www.anaconda.com/distribution](http://www.anaconda.com/distribution).
- [15] "Quandl." Quandl.com, [www.quandl.com](http://www.quandl.com).
- [16] "Stock Market Tools." Tiingo, [www.tiingo.com](http://www.tiingo.com).
- [17] Zakamulin, V.: "Market Timing with Moving Averages: The Anatomy and Performance of Trading Rules", Ed. Springer (2017).
- [18] "Repository of Vincenzo – 58\_Stock-Modeling." KNIME Hub, [hub.knime.com/vincenzo/spaces/Public/latest/58\\_Stock-Modeling/](http://hub.knime.com/vincenzo/spaces/Public/latest/58_Stock-Modeling/).
- [19] "Regresión Lineal." Wikipedia, Wikimedia Foundation, 13 May 2020, [es.wikipedia.org/wiki/Regresión\\_lineal](http://es.wikipedia.org/wiki/Regresión_lineal).
- [20] Dowd, K.: "Measuring Market Risk", (2002)
- [21] Lachowicz, P: "Applied Portfolio VaR Decomposition. Impact vs Moving Elements, Quant At Risk", (2015)
- [22] Charles Chalermkraivuth, K., Lynn Kiaer, C., Bollapragada, S., Chakraborty, A.: "Methods and systems for Analytical-Based Multifactor Multiobjective Portfolio Risk Optimization", US Patent 7593880 B2 (2009)
- [23] C. Thomsett, M.: "Stock market math essential formulas for selecting and managing stock and risk", Boston Walter de Gruyter Inc (2017)

- [24] Watada, J.: "Risk Analysis of Portfolio Selection Based on Kernel Density Estimation", (2020)
- [25] Xu, Z., Zhao, H., Hao, Z.: "Several Intuitionistic Fuzzy Multi-Attribute Decision Making Methods and their Applications", (2020)
- [26] Kreinovich, V., Thach, N.N., Trung, N.D., Van Thanh, D.: "Beyond Traditional Probabilistic Methods in Economics", (2018)
- [27] Connors, L. , Alvarez, C., Research, C.: "How Markets Really Work: Quantitative Guide to Stock Market Behavior", (2012)

# Anexo I

## Guia de instalación

Lo primero a realizar es descargar la herramienta KNIME<sup>[11]</sup> de su página web, una vez realizada la descarga se instala KNIME y se selecciona un directorio de trabajo.

Se crea un nuevo flujo de trabajo y se importan los ficheros que realizan las predicciones, para que los ficheros funcionen se debe de instalar anaconda<sup>[14]</sup>, en concreto la versión de Python 3.7.

Por último hay acceder a File->Preferences->KNIME->Python y establecer la versión de Python por defecto a la 3, además se tiene que crear un nuevo entorno para lo que hay que clicar en "*new environment*".

Una vez realizada la configuración en KNIME hay que anotar el nombre del entorno y abrir la terminal de anaconda llamada **Anaconda Prompt**. En la terminal hay que escribir los siguientes comandos:

1. activate nombre\_entorno
2. pip install pandas-datareader

Si la instalación se ha realizado de forma adecuada el programa se ejecutará sin problemas.

## Entrada y manipulación de datos

En este apartado se va a explicar cómo recopila información de diferentes fuentes de datos.

### Google

**Google finance** no proporciona una descarga de datos común, en su lugar se encuentra **stooq** que tiene estos datos para descargar. Se utiliza el módulo para python **pandas-datareader** que ya utiliza el flujo de trabajo del cual partimos. Google no proporciona el parámetro Adj. Close necesario para realizar la predicción en el flujo, por lo que se ha duplicado la columna del parámetro Close y se ha utilizado esta como Adj. Close.

## Quandl

Para poder obtener los datos de esta fuente primero se ha tenido que registrar una cuenta en la página web de Quandl<sup>[15]</sup> y obtener una clave de api. Se ha instalado el módulo de python de quandl en el entorno de python que utiliza knime:

1. activate nombre\_entorno
2. pip install quandl

Por último se han recopilado los datos de la api y se han ajustado al modelo.

## Tiingo

Para poder obtener los datos de esta fuente primero se ha tenido que registrar una cuenta en la página web de Tiingo<sup>[16]</sup> y obtener una clave de api. Se ha instalado el módulo de python de tiingo en el entorno de python que utiliza knime:

1. activate nombre\_entorno
2. pip install tiingo

Por último se han recopilado los datos de la api y se han ajustado al modelo.