

Máster Universitario en Tecnologías del sector financiero  
2019-2020

*Trabajo Fin de Máster*

“Predicción de volatilidad de la cotización de empresas del IBEX35 basada en Google Trends.”

D. Álvaro Andrés Suárez Alfonso

Tutor

Dr. Ricardo Martín Manso

Leganés, 2020

*[Incluir en el caso del interés de su publicación en el archivo abierto]*



Esta obra se encuentra sujeta a la licencia Creative Commons **Reconocimiento – No Comercial – Sin Obra Derivada**

# Resumen

Este trabajo fin de máster aplica técnicas de extracción y minado de datos desde Google trends y Yahoo finance para predecir la volatilidad de los principales índices bursátiles de los mercados financieros.

Este proyecto hace un análisis de comparación entre tendencias de búsqueda en diferentes fuentes de Google y la convergencia de precios de las cotizaciones de un índice bursátil para demostrar una posible correlación entre dichas fuentes de información.

# Abstract

# Palabras clave

Índice

[Resumen 2](#_Toc47023101)

[Abstract 3](#_Toc47023102)

[Palabras clave 4](#_Toc47023103)

[1. Introducción 6](#_Toc47023104)

[1.1 Definiciones y acrónimos 6](#_Toc47023105)

[1.1.1 Renta variable 6](#_Toc47023106)

[1.1.1.1 Volatilidad de mercado 6](#_Toc47023107)

[1.1.2 Índice bursátil 7](#_Toc47023108)

[1.1.2.1 Índices de importancia a nivel mundial 7](#_Toc47023109)

[1.1.2.1.1 Ibex 35 7](#_Toc47023110)

[1.1.2.1.2 DAX 30 8](#_Toc47023111)

[1.1.2.1.3 S&P 500 8](#_Toc47023112)

[1.1.2.1.4 Nasdaq 100 8](#_Toc47023113)

[1.1.3 Análisis de datos 8](#_Toc47023114)

[1.1.3.1 Predicción de mercado 8](#_Toc47023115)

[1.2 Objetivo de la tesis 8](#_Toc47023116)

[2. Estado del arte 8](#_Toc47023117)

[2.1 Predicción de precios de mercado con series temporales binarias 10](#_Toc47023118)

[2.2 Predicción mediante correlación noticias y movimiento de precio de índices 10](#_Toc47023119)

[2.3 Motivación 11](#_Toc47023120)

[3. Metodología y ejecución del proyecto 11](#_Toc47023121)

[3.1 Preprocesado de los datos 11](#_Toc47023124)

[3.2 Exploración de datos 16](#_Toc47023125)

[3.3 Modificación y transformación de los datos 18](#_Toc47023126)

[3.4 Clasificación y predicción 19](#_Toc47023127)

[4. Arquitectura TI 20](#_Toc47023128)

[4.1 Arquitectura física 20](#_Toc47023129)

[4.2 Diseño 20](#_Toc47023132)

[4.2.1 Backend 20](#_Toc47023133)

[4.2.2 Frontend 20](#_Toc47023134)

[4.3 Planificación del proyecto 21](#_Toc47023135)

[4.4 Estructuras de datos 21](#_Toc47023136)

[5. Análisis de resultados 21](#_Toc47023137)

[6. Conclusiones 21](#_Toc47023138)

[7. Bibliografía 22](#_Toc47023139)

# Introducción

# Definiciones y acrónimos

# Renta variable

En el mercado bursátil se define la renta variable como un tipo de inversión en el que existe un instrumento como parte de un capital en representación de una empresa. Se define como variable a aquel instrumento del cual no se tiene conocimiento sobre su rentabilidad, es decir, a un plazo futuro se pueden reportar ganancias o pérdidas, por lo cual se considera la renta variable como una renta con riesgo alto.

# Volatilidad de mercado

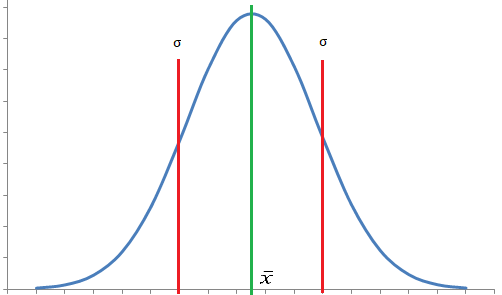
Conocido también como la desviación estándar del precio de un activo, en matemáticas financieras corresponde a la frecuencia en que el precio de un activo se mueve, ya sea positiva o negativamente respecto de su valor inicial.

Los agentes de mercado especulan sobre la volatilidad de mercado con el fin de hacer inversiones de riesgo en el que puedan obtener rentabilidad al vender el activo en los picos de precio más altos, u otra forma es comprando a su precio más bajo para posteriormente tomar una posición corta.

En estadística la desviación estándar es conocida por la letra griega σ (sigma) y se calcula de la siguiente forma:

Donde n es el número de muestras, el valor de la muestra i-esima y la media de las muestras.

La siguiente figura es una representación de la desviación estándar en una curva de Gauss:



Una desviación estándar alta indica que las muestras se encuentran alejadas de su media, mientras que una desviación estándar baja señala que la mayoría de los datos están agrupados cerca de su media, acentuando de esta forma la curva de distribución normal como se muestra a continuación.

# Índice bursátil

El índice bursátil concierne a un grupo de acciones clasificadas por una o varias características similares. Por ejemplo, un grupo de acciones puede ser clasificado por su similitud en el índice de capitalización o por que se encuentran en la misma bolsa de mercado. Este es de gran importancia para las gestoras de fondos ya que con estos permiten buscar una cartera que replique dicho índice

# Índices de importancia a nivel mundial

# Ibex 35

Es el índice bursátil propio de la bolsa española y creado por Bolsas y mercados españoles (BME) en el como su numero lo indica es un grupo formado por 35 empresas con mayor liquidez que cotizan en la bolsa española. Empresas se calcula en base a la ponderación sobre la capitalización de las sus empresas componentes, estas empresas pertenecen a diversos sectores económicos como energía, telecomunicaciones, finanzas, entre otras. En las empresas principales que lo componen se encuentran Banco Santander, Inditex e Iberdrola.

# DAX 30

# S&P 500

# Nasdaq 100

# Análisis de datos

# Predicción de mercado

# Objetivo de la tesis

# Estado del arte

La capitalización total de mercado de todas las acciones en el mundo era aproximadamente 70.75 Trillones de Dólares [1]. El IBEX 35 por si solo alcanza un Trillón de Euros[2]. En 2013, el volumen medio diario de ejecuciones en el New York Stock Exchange fue de 169 Billones de Dólares [3]. Estas cifras nos hacen conscientes de las magnitudes que se manejan en el mercado de acciones a nivel mundial y nacional en términos de riqueza, así como la importancia de gestionar los riesgos en estos mercados para la preservación del capital en la economía mundial.

Para una correcta gestión del riesgo se necesitan monitorizar adecuadamente las principales métricas de riesgo, para poder distribuir el capital de forma que el riesgo quede minimizado. Esta práctica tiene origen en la Teoría de Gestión de Carteras del premio Nobel H. Markowitz[4], y sigue siendo la base en gestión eficiente de riesgo en carteras financieras de activos. La repercusión de este trabajo reside en la presentación de una forma de monitorizar cambios en la medida de riesgo más importante, la volatilidad, de forma anticipada y precisa gracias a la tecnología. Esto tiene implicaciones prácticas tanto en la gestión del riesgo, como medida del sentimiento de mercado y para la toma de decisiones de compra o venta de productos derivados de volatilidad como opciones. El marco de este trabajo es el de predicción de series temporales de activos financieros por medio de análisis de sentimiento de mercado. Más específicamente, el trabajo se enfoca en la predicción de grandes cambios (más de una desviación estándar) en la volatilidad de los activos financieros.

En la literatura del análisis de sentimiento de mercado encontramos que la gran mayoría de investigaciones se enfocan en identificar la polaridad entre opiniones positivas y negativas para inferir subidas o bajadas del precio de los activos, como en S. Nausheen et al. [5]. Pero también, dentro de la literatura encontramos trabajos enfocados en la medición de atención que se da a un término en específico como en F. Audrino et al.[6]. Este enfoque dentro de la literatura de análisis de sentimiento de mercado, pero dedicado a la medición de atención de ciertos términos es el que se adopta en este trabajo. Ambos enfoques son muy similares en técnicas de procesado de datos y en los posteriores modelos de clasificación o predicción, generalmente con técnicas supervisadas de Machine Learning. En [6], los autores realizan un estudio de regresión para identificar las palabras que mayor impacto tienen en la predicción de la volatilidad realizada (la desviación estándar del precio, para este documento volatilidad). Utilizan como medida de atención el número de veces que aparece al día la búsqueda de la palabra referente al stock en Google trends. Este trabajo difiere de este en que nosotros nos enfocamos en la predicción de grandes cambios en la volatilidad cuando se producen grandes cambios en la medición de atención de ciertos términos que nosotros ya conocemos de antemano. Esto hace que nos fijemos en los extremos de esta relación y no en el continuo como hacen estos autores con la regresión y podamos así probar nuestra hipótesis de partida. Otro estudio similar es el de M. Y. Huang et al. [7], en donde los autores analizan las relaciones entre los volúmenes de búsqueda con Google Trends y los movimientos direccionales del SP500. Encuentran que la relación entre cambios en el volumen de búsquedas y cambios direccionales es condicional al sentimiento de mercado implícito en el término buscado. Además, identifican términos consistentes con esta regla anterior y utilizan un modelo para predecir los movimientos en base a estos términos y construyen una estrategia de trading con muy buenos resultados. Nuestro estudio se diferencia de este en que, al igual que en [6], los autores analizan cambios y no eventos extremos de las variables.

En la literatura de sentimiento de mercado encontramos dos ejes principales, el de minería de datos de texto en la web con el que se obtiene la variable independiente y el de modelos de predicción (y en algunos casos de clasificación) de la variable dependiente en cuestión. Para el primer eje, encontramos técnicas como Procesamiento del Lenguaje Natural (NLP), como en WS Chan [8] donde las aplican a diferentes textos de internet o las aplican a blogs, noticias y redes sociales como en I. Vaishali y S. Deshmukh [9], C. Vega y R.A. Albuquerque [10], y C. Jiang et al. [11]. También encontramos casos como el nuestro en los que se mide la atención del inversor de ciertos términos en los que NLP no es necesario. Tradicionalmente para medir el interés del inversor se utilizaban movimientos extremos del activo, cambios en el volumen de trading o límites de precio. Con el avance de la tecnología se han añadido noticias y titulares, así como ciertas fuentes de internet [7]. En H.S. Moat et al. [12], se investiga el impacto que tiene el número de visitas que en páginas de Wikipedia en el movimiento del precio de la acción. En Z. Da et al. [13], los autores analizan el número de búsquedas en Google de compañías cerca a un evento IPO para predecir el precio de una acción. En N. Vlastakis y R.N. Markellos [14], los autores aproximan la demanda de acciones e índices con las búsquedas semanales en Google Trends y lo usan para analizar la volatilidad. M. Bank et al. [15], muestran la relación entre el volumen de búsquedas en Google Trends y el retorno de las acciones. Se enfocan en los nombres de las compañías para las búsquedas y demuestran una relativamente alta correlación con el retorno positivo del activo. En C. Curme et al. [16], los autores encuentran que la búsqueda en Google trends de términos políticos y financieros esta relacionado con una caída del precio del activo en el siguiente día. También observan una caída de retornos en estrategias de trading con compras o ventas basadas en señales de atención de términos de Google Trends. Por otra parte, en T. Preis et al. [17], encuentran que se acumula mucho ruido cuando se utiliza un solo término de atención como señal. Como bien hemos dicho antes, nuestro estudio difiere de los anteriores en que nosotros nos enfocamos en la relación entre cambios extremos de variables y no en relaciones en un continuo de su valor.

Para llevar a cabo nuestro trabajo debemos convertir las series continuas de volatilidades y de volumen de búsquedas en series binarias. Este es un método que entra dentro de la literatura de extracción de características, según la cual extraemos de un proceso continuo una característica como es su comportamiento en valores extremos en función del tiempo. Esto nos permite evaluar nuestra hipótesis con técnicas similares a las antes mencionadas sobre dichas nuevas características. En cuanto a los modelos de predicción utilizados en la literatura de sentimiento de mercado y atención encontramos tanto modelos estadísticos y econométricos tradicionales, como modelos de aprendizaje automático tanto de clasificación como de predicción. Para una detallada literatura de modelos de predicción referir a G. Mahalakshmi et al. [18]. No incidimos en los modelos de predicción porque a diferencia de la mayoría de estudios en la literatura nosotros no utilizamos un modelo para predecir o tomar decisiones ajustando unos parámetros, o una ecuación a los datos, sin embargo probamos la hipótesis que el volumen de búsquedas de ciertos términos conocidos de antemano categorizados en valores extremos tienen una beta de 1 con cambios extremos del mismo orden estadístico de la volatilidad del activo.

# Predicción de precios de mercado con series temporales binarias

* Identificación: Evolving Hypernetwork Models of Binary Time Series for Forecasting Price Movements on Stock Markets
* Autores: Elena Bautu, Sun Kim, Andrei Bautu, Henri Luchian and Byoung-Tak Zhang.
* Objetivo general: Predecir subida o bajada de precios de mercado utilizando mediante análisis de series temporales binarias.
* Categoría: Análisis cuantitativo
* Instrumentos de Desarrollo de la investigación: Se utiliza un algoritmo de hiperred el cual es un grafo de relaciones probabilísticas de orden superior para encontrar patrones en datos históricos utilizando el índice Dow Jones industrial y el índice compuesto de precios de las acciones del mercado koreano
* Resultados:

, Objetivo general, categoría, instrumentos de desarrollo de la investigación, resultados

# Predicción mediante correlación noticias y movimiento de precio de índices

* Identificación:
* Autores:
* Objetivo general:
* Categoria:
* Instrumentos de Desarrollo de la investigación
* Resultados:

# Motivación

# Metodología y ejecución del proyecto

La evaluación del modelo consiste en extraer datos de cierre de posiciones de los últimos 10 años de unos de los índices bursátiles mas importantes existentes en los mercados financieros, el IBEX35 del mercado español, el S&P 500 índice estadounidense y posiblemente el índice mas característico del clima mercantil actual, y el NASDAQ 100, también de origen estadounidense que aglomera los 100 valores de las compañías de tecnología mas relevantes en este mismo mercado. La fuente de los datos proviene de [Yahoo Finance](https://es.finance.yahoo.com/) y son extraídos los datos históricos como series temporales con frecuencia diaria.

Adicionalmente se agrega al modelo una serie temporal de tendencias de búsqueda en Google el cual se extrae a través del servicio de [Google Trends](https://trends.google.es/) y que muestra el conteo de búsquedas que se han hecho diariamente a una palabra determinada.

Teniendo estas dos series temporales se procede a ver un patrón entre la tendencia de búsquedas del índice o una empresa específica del mismo, y la variación atípica en su precio de cierre.



# Preprocesado de los datos

La data de las dos series es extraída mediante un servicio hecho con el framework NodeJS el cual hace uso del paquete de npm google-trends-api y yahoo-finance:

const googleTrends = require('google-trends-api');

const yahooFinance = require('yahoo-finance');

Se desarrolla un api que hace uso de estas dos api’s de npm:

##### Servicio de Google trends:

app.get('/volatilitypred/extractTrends/:trendWord/:startDate/:endDate', (req, res, next) => {

  const keyword = req.params.trendWord;

  const startDate = new Date(req.params.startDate);

  const endDate = new Date(req.params.endDate);

  const optionsObject = {

    keyword: keyword,

    startTime: startDate,

    endTime: endDate

  }

  googleTrends.interestOverTime(optionsObject)

  .then(function(results){

     results = JSON.parse(results).default.timelineData;

     results.forEach(function(result) {

      result.time = result.time

      result.formattedTime = result.formattedTime;

      result.formattedAxisTime = result.formattedAxisTime;

      result.value = result.value[0];

      result.formattedValue = result.formattedValue;

      result.symbol= keyword;

    });

    res.status(201).json({

      gtrendsdata: results

    });

  })

  .catch(function(err){

    console.error('Oh no there was an error', err);

  });

});

El cual responde con datos con la siguiente estructura:

Time: Fecha en formato numerico

Formattedtime: Fecha de consulta de la palabra

FormattedAxisTime: Fecha del registro

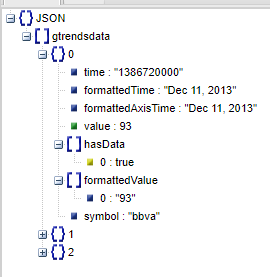
Value: Conteo total de la tendencia (palabra buscada) en registro.

hasData: Si value arroja un resultado mayor a 0

symbol: Palabra buscada

Aquí en la imagen 1 un ejemplo de la respuesta del api rest:

Imagen 2



##### Servicio de Yahoo finance:

app.get('/volatilitypred/extractFinance/:stock/:startDate/:endDate', (req, res, next) => {

  const startDate = new Date(req.params.startDate);

  const endDate = new Date(req.params.endDate);

  const optionsObject = {

    symbol: req.params.stock,

    from: startDate,

    to: endDate

  }

  yahooFinance.historical(optionsObject)

    .then(function(results){

      res.status(201).json({

        stockdata: results.reverse()

      });

  });

});

Los datos provenientes de yahoo cuentan con la siguiente estructura:

Date: Fecha en que se hace el registro.

Open: Precio de apertura en el mercado del activo, bono o valor.

High: Precio más alto alcanzado entre la hora de apertura y la de cierre.

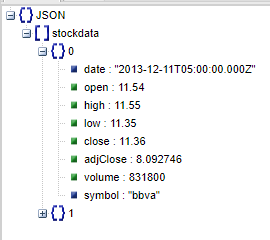
Low: Precio más bajo alcanzado entre la hora de apertura y la de cierre.

Close: Precio de cierre del día.

Adj Close: Precio de cierre del día después de pagar dividendos o stock splits. Este en particular va a ser el que incluiremos en la muestra de preprocesado de datos posterior, debido a que es mas preciso a la hora de valorar la subida o la caída de una acción gracias al análisis en los cálculos de los diferentes factores contemplados.

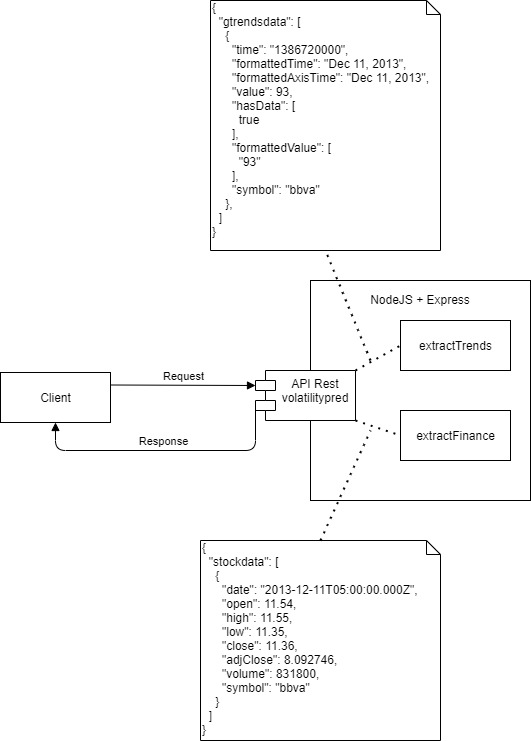
Aquí en la imagen 2 un ejemplo de la respuesta del api rest:

Imagen 3



Como se muestra en el siguiente diagrama el api rest está disponible para ser consumida por cualquier cliente a través de una petición http:

Imagen 4

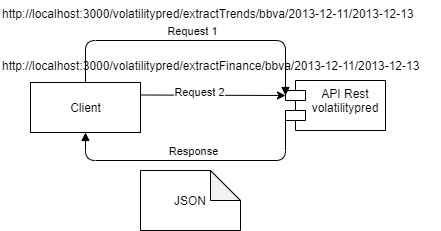


# Exploración de datos

En este apartado y correspondiente a la imagen 3, nos centraremos en el cliente el cual está encargado de hacer la extracción de los datos y pre transformarlos de tal forma que podamos manipularlos fácilmente para verificar la hipótesis.

Se hacen dos llamados al api, uno para extraer Google trends y el otro para yahoo finance. La respuesta se recibe en formato JSON:

Imagen 5



Y desde el cliente efectuamos una transformación por cada serie recibida:

def extract\_data(getDataFromAPI: GetDataFromAPI, start\_date, end\_date):

    response\_trend\_df = getDataFromAPI.request\_transform\_trend\_data(start\_date, end\_date)

    response\_finance\_df = getDataFromAPI.request\_transform\_finance\_data(start\_date, end\_date)

    return response\_trend\_df, response\_finance\_df

def request\_transform\_trend\_data(self, start\_date, end\_date) -> pd.core.frame.DataFrame:

        response\_json = self.request\_service(CONSTANTS.googleTrendsAPIMethodURL + self.trend\_word + '/' + start\_date + '/' + end\_date)

        if len(response\_json['gtrendsdata']) > 0:

            response\_df = pd.DataFrame(response\_json['gtrendsdata'])\

                .drop(['formattedTime', 'formattedAxisTime', 'formattedValue', 'hasData', 'symbol'], axis=1)

            response\_df['time'] = response\_df['time'].apply(self.format\_date\_fromtimestamp)

            response\_df = response\_df.rename(columns={"time": "date"}, errors="raise")

        else:

            response\_df = None

        return response\_df

def request\_transform\_finance\_data(self, start\_date, end\_date):

        response\_json = self.request\_service(CONSTANTS.yahooFinanceAPIMethodURL + self.stock + '/' + start\_date + '/' + end\_date)

        if len(response\_json['stockdata']) > 0:

            response\_df = pd.DataFrame(response\_json['stockdata'])\

                .drop(['symbol', 'open', 'close', 'volume', 'low', 'high'], axis=1)

            response\_df['date'] = response\_df['date'].apply(self.format\_date\_string)

        else:

            response\_df = None

        return response\_df

De esta forma se cargan los datos ya transformados en dataframe con las estructuras siguientes. En la tabla se muestra un ejemplo de la serie temporal extraída de Yahoo Finance:

Tabla 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date | Open | High | Low | Close | Adj Close | Volume |
| 28/07/2011 | 9544.299805 | 9707.400391 | 9517.099609 | 9657.000000 | 9656.990234 | 247610000 |
| 29/07/2011 | 9540.299805 | 9679.000000 | 9499.799805 | 9630.700195 | 9630.690430 | 254006000 |
| 01/08/2011 | 9713.000000 | 9752.299805 | 9316.400391 | 9318.200195 | 9318.190430 | 290079000 |
| 02/08/2011 | 9236.099609 | 9342.599609 | 9114.900391 | 9114.900391 | 9114.891602 | 387974000 |
| 03/08/2011 | 9033.700195 | 9297.299805 | 8958.599609 | 9037.700195 | 9037.691406 | 306543000 |
| 04/08/2011 | 9217.700195 | 9219.700195 | 8686.500000 | 8686.500000 | 8686.491211 | 385372000 |

Date: Fecha de registro de la tendencia (palabra buscada).

Value: Conteo total de la tendencia (palabra buscada) en registro.

La siguiente tabla muestra un ejemplo de los datos provenientes de Google trends:

Tabla 2

|  |  |
| --- | --- |
| **date** | **value** |
| 2011-01-02 00:00:00 | 50 |
| 2011-01-03 00:00:00 | 36 |
| 2011-01-04 00:00:00 | 45 |
| 2011-01-05 00:00:00 | 44 |
| 2011-01-09 00:00:00 | 57 |
| 2011-01-10 00:00:00 | 42 |
| 2011-01-11 00:00:00 | 20 |

# Modificación y transformación de los datos

En el marco del desarrollo de la hipótesis se hace un procesado de los datos extraídos de yahoo finance y Google trends, y se hace una poda inicial de las series temporales eliminando los sábados de las dos series, domingos de la serie finance y los viernes de la serie trends quedando con los días lunes a viernes para los datos yahoo finance y domingo a jueves para los de Google trends:

response\_trend\_df = prd.prune\_day\_from\_dataframe(response\_trend\_df, [4, 5])  # Viernes y Sabado

response\_finance\_df = prd.prune\_day\_from\_dataframe(response\_finance\_df, [5, 6])  # Sabado y Domingo

def prune\_day\_from\_dataframe(df, days\_to\_exclude):

    # Quita un día determinado de un time series data frame

    df["weekday"] = pd.to\_datetime(df.date).dt.dayofweek

    for day\_to\_exclude in days\_to\_exclude:

        df = df.loc[df["weekday"] != day\_to\_exclude]

    df = df.reset\_index(drop=True)

    return df

El objetivo de purgar estos datos es que el mercado de stocks no se mueve en días no hábiles pero las tendencias de búsqueda contienen información de todos los días calendario del año, por lo cual se deben ajustar las dos series entre si eliminando el ruido que genera la serie de Google trends. A tener en cuenta y es que con el siguiente algoritmo se ajustan las dos series también contemplando los días feriados:

def prune\_row\_correspondance\_by\_value(left\_df, right\_df,

                                      column\_name, value, right\_correspondance, date\_column\_name):

    # Selecciono primero los valores que se van a podar en el primer dataframe

    df = left\_df.loc[left\_df[column\_name] == value].copy()

    # Se guarda un array con las fechas que van a servir para podar los datos en el otro array

    df[date\_column\_name] = df[date\_column\_name] + timedelta(days=right\_correspondance)

    date\_np = np.array(df[date\_column\_name])

    # Borro en el df derecho

    for date\_ in date\_np:

        right\_df = right\_df[right\_df[date\_column\_name] != date\_]

    right\_df = right\_df.reset\_index(drop=True)

    # Borro en el df izquierdo

    left\_df = left\_df.loc[left\_df[column\_name] != value]

    left\_df = left\_df.reset\_index(drop=True)

    return left\_df, right\_df

Finalmente, y en consiguiente a la transformación y purga de los datos de las dos series con la ayuda de la librería numpy se obtienen los diferentes datos estadísticos que vamos a usar para binarizar las series, la media y la variación estándar (sigma). D este modo tenemos dos muestras con su correspondiente distribución.

# Clasificación y predicción

En este apartado se hablará acerca del método de clasificación y predicción posterior a la generación de la distribución de las dos muestras, la de google trends y la de yahoo finance.

Como se menciona previamente para la serie de yahoo finance se tomará como referencia el precio de cierre ajustado debido a que es más preciso a la hora de valorar la subida o la caída de una acción gracias al análisis en los cálculos de los diferentes factores contemplados. Con respecto al precio de cierre se hará un cálculo de los retornos en t con respecto a t-1:

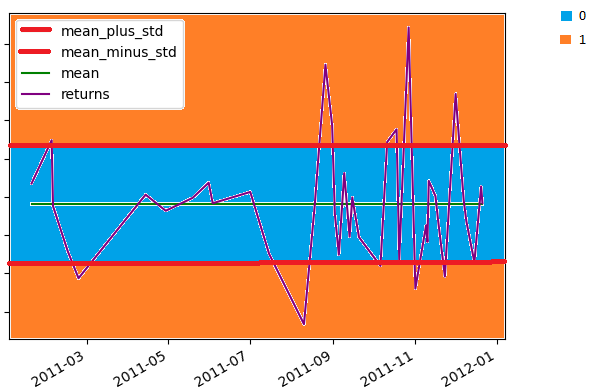
De esta forma se calculan iterativamente los retornos para cada instante t:

Tabla 3

|  |  |  |
| --- | --- | --- |
| **date** | **adjClose** | **returns** |
| 2011-01-03 00:00:00 | 5,881024 | 0 |
| 2011-01-04 00:00:00 | 5,869412 | -0,00197449 |
| 2011-01-05 00:00:00 | 5,74485 | -0,02122223 |
| 2011-01-06 00:00:00 | 5,504013 | -0,04192224 |
| 2011-01-10 00:00:00 | 5,304294 | -0,03628607 |
| 2011-01-11 00:00:00 | 5,410028 | 0,019933661 |

Y con respecto a este retorno se obtiene la media y variación estándar de la muestra. La media será la constante más cercana a cada punto de la distribución y sigma+media y sigma-media serán los limites en los cuales se considera que la mayoría de los datos de retorno se encuentran aglomerados, los puntos de la serie que se encuentren por fuera de estos límites (límite superior y límite inferior) son retornos atípicos por lo cual nos servirá para poder hacer una clasificación binaria en función del tiempo t:

Imagen 6



Por lo anterior lo que se encuentre por fuera de los limites sigma+ y sigma- en la hipótesis se toma con un cambio brusco en el retorno y por tanto una volatilidad por fuera de la normal.

En el caso de la serie de tendencias se calcula la media y variación estándar con respecto del valor del conteo de búsqueda del stock y se obtiene su distribución. Lo que se encuentre por fuera de los limites de su sigma+ y sigma- se considera una señal para el posible cambio atípico en la volatilidad del stock en un instante futuro, es decir, un valor binario positivo en la serie de tendencias en el instante t, implicaría un aumento o decremento considerable en la volatilidad del activo analizado en t+1:

Imagen 7



#TODO

// técnicas de análisis explicado a alto nivel (explicación a usuario final)

// técnicas de análisis—métricas utilizadas

# Arquitectura TI

# Arquitectura física



# Diseño

# Backend

# Frontend

# Planificación del proyecto

# Estructuras de datos

//5. diseño experimiental: rango de datos, de donde salen, como son los datos, close ajustado o sin ajustar.

# Análisis de resultados

//matrices de confusión, roc, etc

# Conclusiones

//opiniones

//Trabajo futuro

#TODO

- modelo de prediccion con la beta para ajustar los parametros del modelo (la beta es la correlacion entre las series)

- analisis de sentimiento en google trends (negativo - positivo) para prededcir la sigma atípica

# Bibliografía

[1] Stock market, \Stock market| Wikipedia, the free encyclopedia," 2004, [Online; accessed 28-July-2020]. [Online]. Available: https://en.wikipedia.org/wiki/Stock market

[2] New York Stock Exchange, \ New York Stock Exchange, | Wikipedia, the free encyclopedia," 2004, [Online; accessed 28-July-2020]. [Online]. Available: https://en.wikipedia.org/wiki/ New\_York\_Stock\_Exchange

[3] IBEX 35, \ IBEX 35, | Wikipedia, the free encyclopedia," 2011, [Online; accessed 28-July-2020]. [Online]. Available: https://en.wikipedia.org/wiki/IBEX\_35

[4] H. Markowitz, \PORTFOLIO SELECTION," The Journal of Finance, vol. 7, no. 1,

pp. 77{91, mar 1952.

[5] Nausheen S, Anil Kumar M, and Amrutha K K. (2017). “SURVEY ON SENTIMENT ANALYSIS OF STOCK MARKET.” International Journal of Research - Granthaalayah, 5(4) RACSIT, 69-75. https://doi.org/10.5281/zenodo.572298.

[6]Audrino, F.; Sigrist, F. & Ballinari, D.The impact of sentiment and attention measures on stock market volatility International Journal of Forecasting, Elsevier, **2020***, 36*, 334-357

[7] Huang, M.Y., Rojas, R.R. & Convery, P.D. Forecasting stock market movements using Google Trend searches. *Empir Econ* (2019). <https://doi.org/10.1007/s00181-019-01725-1>

[8] Chan WS. Stock Price reaction to news and no-news: drift and reversal after headlines Journal of Financial Economics, Elsevier, **2003***, 70*, 223-260

[9] Ingle V., Deshmukh S. (2017) Live News Streams Extraction for Visualization of Stock Market Trends. In: Lobiyal D., Mohapatra D., Nagar A., Sahoo M. (eds) *Proceedings of the International Conference on Signal, Networks, Computing, and Systems.* Lecture Notes in Electrical Engineering, vol 395. Springer, New Delhi. <https://doi.org/10.1007/978-81-322-3592-7_30>

[10] Vega, Clara and Albuquerque, Rui A., Economic News and International Stock Market Co-Movement (January 19, 2012). Review of Finance Vol. 13, 2009, EFA 2006 Zurich Meetings, Available at SSRN: <https://ssrn.com/abstract=1139927>

[11] Jiang, C., Liang, K., Chen, H. et al. Analyzing market performance via social media: a case study of a banking industry crisis. Sci. China Inf. Sci. 57, 1–18 (2014). https://doi.org/10.1007/s11432-013-4860-3

[12] Moat, H., Curme, C., Avakian, A. et al. Quantifying Wikipedia Usage Patterns Before Stock Market Moves. Sci Rep 3, 1801 (2013). <https://doi.org/10.1038/srep01801>

[13] Da Z, Engelberg J, Gao P (2011) In search of attention. J Finance 66:1461–1499

[14] Vlastiakis N, Markellos RN (2010) Information demand and stock market volatility. SSRN, eLibrary

[15] Bank M, Larch M, Peter G (2011) Google search volume and its influence on liquidity and returns of German stocks. Financ Mark Portf Manag 25:239

[16] Curme, Chester and Preis, Tobias and Stanley, H. Eugene and Moat, Helen Susannah, Quantifying the Semantics of Search Behavior Before Stock Market Moves (August 12, 2014). Proceedings of the National Academy of Sciences 111, 11600-11605; DOI:10.1073/pnas.1324054111 (2014) , Available at SSRN: <https://ssrn.com/abstract=2480274>

[17] Preis, Tobias and Moat, Helen Susannah and Stanley, H. Eugene, Quantifying Trading Behavior in Financial Markets Using Google Trends (April 25, 2013). Scientific Reports, Vol. 3, pp. 1684; DOI:10.1038/srep01684 (2013), Available at SSRN: <https://ssrn.com/abstract=2260189>

[18] Mahalakshmi, G.; Sridevi, S. & Rajaram, S. A survey on forecasting of time series data 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), **2016**, 1-8