

## Estimación del mercado de valores con base en las visitas a Wikipedia

Stock market estimation based on Wikipedia visits

Mercado de valores oñemongu'eháicha oñekundahávo Vikipetã

**Swarnava Mitra y José Nicanor Franco-Riquelme**

*Universidad Politécnica de Madrid*

### **Nota de los autores**

*Departamento de Lenguajes y Sistemas Informáticos e Ingeniería de Software.*

*Departamento de Ingeniería de Organización, Administración de Empresas y Estadística.*

*j.franco.riquelme@upm.es*

### **Resumen**

Este documento propone una metodología para estimar el movimiento de los principales índices de mercado y con base en las visitas a Wikipedia utilizando el algoritmo FuzzyCoCo. Estudios anteriores han demostrado que el aumento del número de visitas a las páginas de Wikipedia en temas relacionados con la economía y las finanzas tiene un efecto en los mercados financieros. Se eligieron tres categorías de temas, una relacionada con deudas económicas, hechos sociopolíticos y otra relacionada con información específica de la empresa. Se utilizó un período de datos históricos de 5 años, desde enero de 2010 hasta diciembre de 2014. Los datos financieros consistieron en los principales índices bursátiles de la UE y EE. UU., como el Promedio Industrial Dow Jones (DJIA) y el S&P500 para los mercados de EE. UU., y FTSE100 y DAX30 para los mercados de la UE, además de los precios de las acciones de Facebook, Apple Inc. y Citigroup. La serie temporal de visitas a la página se probó primero para una prueba de causalidad de Granger y luego se utilizó una variable exógena para predecir los movimientos del mercado junto con los indicadores técnicos de uso común. La principal contribución del trabajo radica en el uso de las páginas vistas de Wikipedia como un indicador basado en el sentimiento social para la predicción de los movimientos del mercado. Los niveles de precisión direccional logrados hacen que la metodología sea atractiva para ser utilizada por los inversores para incorporar el sentimiento general del mercado con respecto a la recesión económica y el malestar social.

**Palabras clave:** Visitas a páginas de Wikipedia, mercado de valores, Fuzzy CoCo (método Cooperativo y Coevolutivo para la regla de Lógica Difusa).

### **Abstract**

This paper proposes a methodology for estimating the movement of the major market indexes (and equity stocks) based on Wikipedia page views using the FuzzyCoCo algorithm. Previous studies have shown that an increasing number of Wikipedia page

views of article-related topics concerning economics and finance have had an effect on financial markets. Three categories of articles were chosen, one relating to economic debts, another to socio-political occurrences, and the third relating to company-specific information. A 5-year period of historical data from January 2010 to December 2014 was used. The financial data consisted of major EU and US stock indexes, as the average, DJIA and S&P500 for US markets, and FTSE100 and DAX30 for EU markets, plus the share prices of Facebook, Apple Inc., and Citigroup. The page view time series was first tested for a Granger causality test and then an exogenous variable was used for predicting market movements along with commonly used technical indicators. The main contribution of the work lies in the use of Wikipedia page views as a social sentiment-based indicator for the prediction of market movements. The directional accuracy levels achieved make the use of this methodology attractive for investors in order to incorporate the general market sentiment with regards to the economic downturn and social unrest.

**Keywords:** Wikipedia page views, Stock market, Fuzzy CoCo (Co-operative Co-evolutionary method for Fuzzy rule-based system).

### Mombykypyre

Ko kuatia rupive ojekuaauka tembiapokue ojehechauka rekávo mba'éichapa mercado de valroes oñemongu'e ojeikévo oñekundaha Vikipetãme, upevarã ojeporu algoritmo FuzzyCoCo. Tembiapokue ymave guare ohechauka amáke ojeike oñeikundaha Vikepetã rogue umi mba'e oñe'éva economía ha finanza rehe, katuetei opokoha mercado financiero rehe. Mbohapy pojoapy ojeporu oñemba'apo rekávo, pete'iva ohóva deuda económica rehe, ambue hecho sociopolítico ha ambueve marandu umi empresa reheguáva. Ojeporu marandu 5 ary pukukue javeve guaréva, jasyte'í 2010 guive jasyapakõi 2014 peve. Finanza rehegua marandu oñeguenohẽ viru oñemongu'éva UE ha EE.UU-pegui, ha upevarã ojeporu Promedio Industrial Dow Jones (DJIA) ha S&P500 mercado EE. UU peguápe; FTSE100 ha DAX30 umi mercado UE-peguápe gũarã, ha avei umi Facebook, Apple Inc. ha Citigroup mba'erepy. Ojehakuaa rekávo mba'éichapa oñeikundaha umi tenda rogue, tenonderã ojeporu prueba de causalidad Granger, ha upe rire katu ojegueroike variable okaguivogua ojecha ha'ña mba'éichapa oñemongu'e mercado ojeporúvo pojoapy upevarã oñembosako'iva. Ko tembiapo tuicha oporoipytyvõ ojeporu ha'ña Vikepetã rogue ojekuarãvo tapichakuéra rembihecha, jehehápa ojechakuaa avei mba'éichapa oñemongu'ehína mercado. Pojoapy ojeporúva niko oporoisãmbhyhy porã oñemongo'õ rekávo marandu ojehekáva, upéicha rupi ikatuporãnte oiporu umi inversor ohechakuaa ha'ña mercado jeguata, economía oguata'i ha tapichakuéra ndahetia'epái javéramo upevakuére.

**Mba'e mba'érepa oñe'ne'ẽ:** Vikipetã rogue ñeikundaha, mercado de valores, Fuzzy CoCo (método Cooperativo ha Coevolutivo regla de Lógica Difusa-pe gũarã).

Fecha de recepción: 30/04/2021

Fecha de aprobación: 28/05/2021

## Introducción

La asimetría de la información en los mercados financieros se ha reducido considerablemente con la introducción de las tecnologías de la información, ya que cada vez más personas pueden acceder a información que antes estaba disponible solamente para unos pocos. La información oportuna sobre los diferentes factores macroeconómicos que influyen en los mercados financieros es importante para que los intermediarios tomen decisiones acertadas. Esta información macroeconómica está disponible en redes sociales y en sitios en línea de libre acceso. Por otro lado, la complejidad de los mercados mundiales con la creciente interconexión a nivel mundial ha aumentado la probabilidad de que una crisis sociopolítica o económica los afecte de manera global. Los inversores e intermediarios recopilan información de sitios relacionados en Internet visitándolos con regularidad. Existen dos categorías de sitios: las webs de noticias y las redes sociales. Los sitios web de noticias expresan opiniones e información proporcionada por periodistas financieros, mientras que los sitios de redes sociales expresan la opinión de las masas. Debido a la existencia de la “mentalidad de manada” y otros sesgos de comportamiento, la información agregada disponible en estos sitios web y de redes sociales, pueden proporcionar información inteligente sobre el comercio.

La Hipótesis de Mercado Eficiente (HME) establece que el valor de un activo financiero viene dado por toda la información disponible en ese preciso momento (Fama, 1970). Por el contrario, Hipótesis de Mercado Adaptativo (HMA) establece que la eficiencia de un mercado depende en gran medida del tipo de participantes del mercado, como administradores de fondos de pensiones, inversores minoristas, creadores de mercado, administradores de fondos de cobertura, entre otros (Lo, 2004). Para un analista financiero tener acceso a la información que poseen todos los actores del mercado, en un preciso momento, es casi imposible; pero, al utilizar esta información presente en los sitios web y de redes sociales, combinada con el análisis técnico convencional, puede brindar al inversionista financiero, cierta información valiosa.

Con respecto a las direcciones de investigación, el uso del análisis de sentimientos para el análisis financiero utilizando información disponible en Internet, se puede clasificar en los siguientes grupos según las fuentes utilizadas. El primer grupo pertenece al uso de Twitter, basándonos en el Procesamiento del Lenguaje Natural (PLN) para comprender el sentimiento general expresado por las personas en la red, la otra metodología se refiere al uso de frecuencias, para calcular el uso de una determinada palabra a lo largo del tiempo y utilizarla como variable exógena en un modelo de predicción. El segundo grupo pertenece al uso de analíticas proporcionadas por Google llamadas Google Trends. Google Trends es una aplicación donde se puede encontrar cualquier término de búsqueda consultado mediante el motor de búsqueda de Google. La frecuencia de uso de las consultas de búsqueda utilizadas para encontrar cierta información macroeconómica se ha utilizado como una variable exógena para el modelado matemático de los movimientos del mercado. El tercer grupo es el uso de sitios web que proporcionan noticias financieras e información sobre el panorama

general de la industria. El cuarto grupo es el uso de visitas a páginas de artículos de Wikipedia. Algunas investigaciones han demostrado que las visitas a las páginas de los artículos de Wikipedia relacionados con el fenómeno macroeconómico están correlacionadas con los movimientos del mercado. La principal contribución de este artículo es ampliar la investigación realizada en este último grupo.

Este artículo amplía el trabajo de la investigación de Moat et al. (2013) y Bollen et al. (2011). Moat et al. (2013) utilizaron artículos de Wikipedia, vista examinada y datos editados de 285 páginas. Los artículos de Wikipedia son temas enumerados en la subsección "conceptos económicos generales" del artículo de Wikipedia "esquema de la economía". Bollen et al. (2011) utilizaron estados de ánimo específicos de masas, relacionados con las elecciones presidenciales de los Estados Unidos de América (EE. UU.) y el día de Acción de Gracias como una variable exógena para estimar la precisión del Promedio Industrial Dow Jones (DJIA). Este trabajo presenta las preguntas de investigación que se indican a continuación:

1) ¿Podrían las páginas vistas de artículos de Wikipedia relacionados con temas como la "deuda nacional de EE. UU." y "crisis de la deuda de la UE", además de tópicos como la "primavera árabe" y páginas de empresas como "Facebook", "Apple Inc.", "Citigroup" ser tomadas como variables exógenas para predecir el mercado, así como los movimientos de las acciones de las empresas?

2) ¿Se podrían utilizar las visitas a las páginas de los artículos de Wikipedia mencionados anteriormente junto con los indicadores técnicos de uso común para lograr una mayor precisión que investigaciones previas?

Este artículo está organizado de la siguiente manera. En la siguiente sección, se presenta una revisión de la literatura que destaca el uso de la web 2.0 y el análisis de sentimiento en el comercio financiero. En la tercera sección se presenta una breve descripción general del algoritmo FuzzyCoCo. La cuarta sección ofrece una descripción breve de las diferentes variables de vistas de página de Wikipedia, los índices de mercado, los indicadores técnicos utilizados y la prueba de causalidad de Granger entre las vistas de página de Wikipedia y las cotizaciones de mercado. La quinta sección comprende el uso del conjunto de datos preparado en la sección anterior para entrenamiento y prueba, usando el algoritmo FuzzyCoCo. La última sección comprende los resultados, observaciones finales y la dirección futura de la investigación.

### **Revisión de la literatura**

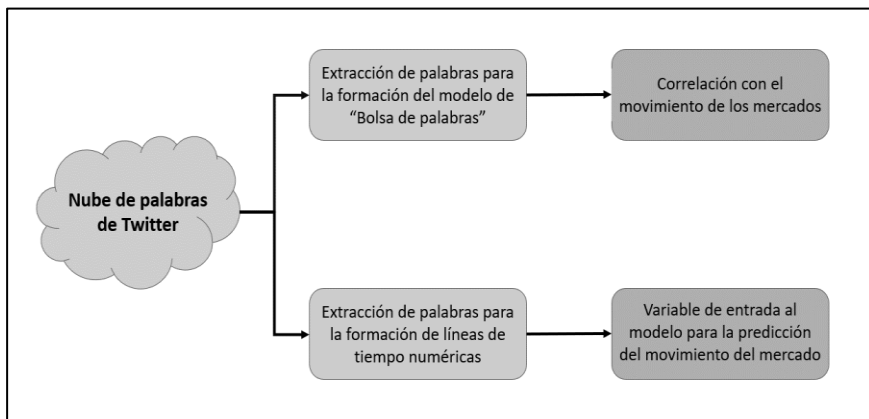
El análisis de los mercados financieros utilizando la información disponible sobre el comportamiento de las masas en Internet se puede dividir en cinco grupos según las fuentes utilizadas. A continuación, se ofrece una breve revisión de la literatura de los cinco grupos analizados en el presente trabajo.

### **El uso de Twitter**

La popularidad de Twitter ha sido una de las principales fuentes de recopilación y análisis de la opinión pública sobre los mercados, las tendencias y el desempeño financiero de las empresas. Millones de mensajes, llamados "tweets" son enviados todos

los días, y de ellos se pueden extraer palabras y sentimientos asociados a temas diversos. Una de las formas más comunes de analizar los tweets es buscando palabras que expresen el sentimiento humano y comprender las emociones tales como “esperanza”, “felicidad”, “miedo”, etc. (Maks y Vossen, 2012) entre los tweets y correlacionar las palabras con el alza y la caída de los mercados en particular. Por otro lado, una metodología desarrollada por Bollen et al. (2011) implica utilizar la frecuencia de los sentimientos de las palabras publicadas en Twitter y usarlo como una variable de entrada para el modelo de aprendizaje automático “Self-Organizing Fuzzy Neural Network” para predecir el movimiento del Promedio Industrial Dow Jones (DJIA). Siguiendo el trabajo anterior, Vu et al. (2012) utilizaron un enfoque metodológico similar para capturar el estado de ánimo del público relacionado con cuatro empresas de tecnología: Microsoft, Apple Inc., Amazon y Google, para predecir el movimiento de precios al alza o a la baja mediante un árbol de decisiones. Bar-Haim et al. (2011) usaron tweets para clasificar a quienes brindan asesoría financiera experta en Twitter de los no expertos, usando “Support Vector Machines” (SVM) y usando varios modelos de combinación.

**Figura 1.** Metodologías de uso de tweets para el análisis del mercado de valores



Aparte de la metodología descrita en la Figura 1, el análisis de sentimiento relacionado con Twitter se ha utilizado principalmente para analizar tweets basados en PLN y observando el movimiento conjunto de los precios de las acciones desde los siguientes ángulos diferentes. Zhang et al. (2012) utilizaron “Dólar”, “Oro” y otras palabras similares extraídas de los tweets y lo asociaron con la coevolución de los mercados de valores. Sprenger y Welpel (2011) utilizaron tweets para encontrar cómo las discusiones sobre empresas y sus pares de la industria en Twitter se relacionaron con el movimiento de acciones. Kontopoulos et al. (2013) introdujeron un método original basado en ontologías para una clasificación más eficaz de palabras, donde en lugar de utilizar una puntuación de sentimiento, utilizaron una calificación de sentimiento. Liu et al. (2015) utilizaron métricas de redes sociales para encontrar que las empresas con cuentas oficiales de Twitter tienen más movimientos simultáneos que la que no tienen

ninguna. Kim et al. (2014) utilizaron los sentimientos de Twitter expresados sobre las empresas que figuran en el índice KRX 100 y lo correlacionaron con los movimientos de precios. Arias et al. (2013) utilizaron datos de opinión de Twitter en forma de series de tiempo numéricas para pronosticar los movimientos del precio de las acciones y los ingresos de taquilla de las películas recién estrenadas, y Sprenger et al. (2014), utilizaron una metodología similar para usar el sentimiento basado en Twitter para relacionarlo con el movimiento de las acciones.

### El uso de Google Trends

Google Trends es una plataforma analítica que da el resultado de una palabra clave utilizando el motor de búsqueda de Google durante un período de tiempo. Luego, los datos se descargan en forma de series de tiempo. Estos datos de series de tiempo se utilizan para comprender la influencia de los términos de búsqueda durante un período de tiempo y su relación con el movimiento de los precios de las acciones. Preis et al. (2013) utilizaron los volúmenes de consultas de Google para comprender el comportamiento colectivo humano de los términos de búsqueda relacionados con las finanzas y la deuda. Los volúmenes de búsqueda se utilizaron luego para implementar una estrategia de negociación, si el volumen de búsqueda en  $t - 1$  es mayor que cero, entonces venden acciones al precio de cierre en  $t$ , y si el volumen de búsqueda es menor que cero, compran las acciones a precio de cierre en  $t$ . Esta estrategia se implementa utilizando la noción de que grandes volúmenes de búsquedas relacionadas con términos financieros son seguidos por caídas del mercado. Vlastakis y Markellos (2012) utilizaron los resultados de búsqueda de Google Trends para hacer una aproximación de la demanda de acciones, Dzielinski (2012) utilizó el volumen de búsqueda en Internet de la palabra clave “economía” para estimar la volatilidad, Bordino et al. (2012) utilizaron la búsqueda de palabras clave de la misma empresa para correlacionar con sus movimientos de precios, y Dimpfl y Jank (2012) utilizaron la búsqueda en Internet de la palabra “Dow Jones” para estimar la volatilidad.

### El uso de Wikipedia

Dos contribuciones importantes con respecto al uso de las visitas a Wikipedia se pueden citar, para comprender el comportamiento de las personas. En la primera, Moat et al. (2013) utilizaron las visitas a páginas de Wikipedia para comprender los primeros signos de los movimientos del mercado de valores y que fueron utilizadas como estrategia comercial. Según sus hallazgos, un gran número de visitas a páginas de Wikipedia fueron seguidas por un mercado bajista, similar al trabajo mencionado en la sección anterior, donde hicieron uso de la estrategia para operar. Los resultados mostraron que la incorporación de visitas a las páginas de Wikipedia para la formulación de la estrategia tiene más efecto que las ediciones de páginas de Wikipedia. Además, su investigación también exploró que sólo las visitas a las páginas de Wikipedia relacionadas con las finanzas ofrecían rendimientos rentables. Xu y Zhang (2013) basaron su investigación en el contexto de agregación de información. Este trabajo tuvo como objetivo descubrir cómo la agregación de información sobre empresas públicas en Wikipedia influye en la divulgación voluntaria de información de los administradores con

respecto a la reacción de los inversores. La principal conclusión del trabajo fue la incorporación de Wikipedia para la mejora del entorno de información en el mercado financiero.

### **El uso de Facebook**

Facebook ha creado una métrica llamada índice de felicidad nacional bruta (FNB) basado en la categorización de palabras positivas o negativas que mide la felicidad de un país en función de las actualizaciones de estado de la población en Facebook. Karabulut (2013) implementó el índice FNB para comprender los cambios en VIX (índice de volatilidad) usando VAR (regresión automática de vectores). El índice FNB muestra una significancia estadística considerable con el movimiento de precios del índice al día siguiente. Una desviación estándar de 1 en el FNB da como resultado un cambio de 11,23 puntos básicos. Este autor amplió el trabajo para encontrar el efecto del FNB en el mercado de valores de otros países. La principal conclusión del trabajo es que el índice FNB sirve como indicador del sentimiento de los inversores.

### **El uso de otros sitios web**

El uso de datos textuales para predecir los movimientos del mercado se inició antes de la era de las redes sociales. Wuthrich et al. (1998) utilizaron información de sitios financieros como Financial Times, The Wall Street Journal, etc. junto con datos financieros para aumentar la riqueza de los datos de entrada. Una combinación de palabras clave como "Bond Strong", "Dollar Flater", "Property Weak" para predecir el precio de cierre de los índices de mercado como el DJIA, Nikkei 225 y FTSE 100. Peramunetilleke y Wong (2002) utilizaron titulares de noticias relacionados con el mercado monetario para hacer una predicción sobre si el mercado se moverá al revés o hacia abajo o si se mantendrá estable. Antweiler y Frank (2004) utilizaron 1,5 millones de mensajes publicados en Yahoo Finance y Raging Bull sobre 45 empresas que figuran en el Promedio Industrial Dow Jones y el Dow Jones Internet Index (DJIA). Adicionalmente, el clasificador bayesiano ingenuo "Naive Bayes" fue utilizado para averiguar si los mensajes son alcistas, bajistas o ninguno y, en consecuencia, se desarrolló una estrategia comercial para comprar, mantener o vender. Fung et al. (2003) utilizaron la fuente de noticias Reuters Market para encontrar noticias que influyan en el movimiento de las acciones correlacionadas. Mittermayer (2004) propusieron "NewsCATS", un sistema de comercio inteligente basado en noticias financieras. Utilizaron la alimentación financiera de PR Newswire y emplearon la categorización de artículos de noticias para el comercio. Soni et al. (2007) utilizaron el sitio web de noticias de Financial Times y un enfoque de visualización de mapas conceptuales para analizar la clasificación de texto utilizando el algoritmo Naive y SVM-bolsa de palabras como clasificador. Das y Chen (2007) utilizaron el sentimiento de los pequeños inversores en los foros de mensajes de la empresa. Rachlin et al. (2007) construyeron un sistema de comercio basado en noticias financieras llamado ADMIRAL utilizando fuentes de datos de Forbes y Reuters a través del algoritmo C4.5 Decision Tree. Lupiani-Ruiz et al. (2011) construyeron un motor semántico de noticias financieras a partir de fuentes de datos heterogéneas en idioma español y lo han aplicado para el IBEX35. Hong y Han (2002) desarrollaron el Knowledge-Based News

Miner (KBNMiner), que es una herramienta de análisis semántico construida utilizando el conocimiento previo de expertos en tasas de interés. Este conocimiento se utiliza para extraer información de noticias escritas en coreano siendo enviada a una red neuronal para la predicción de tasas de interés. Tetlock et al. (2008) utilizaron fuentes de noticias de Wall Street Journal y Dow Jones News Service para predecir las ganancias contables y los rendimientos de las acciones de las empresas individuales, los hallazgos se basaron en que las palabras negativas tienen más impacto. Esteban-Gil et al. (2012) construyeron el Social BROKER que utiliza métodos semánticos para comprender la interacción social entre grupos de inversores. Nizer y Nievola (2012) utilizaron artículos de noticias escritos en portugués para predecir la volatilidad del mercado de valores utilizando un modelo GARCH. Wang et al. (2008) utilizaron métodos basados en ontología para clasificar las noticias y luego usaron mapas causales para encontrar cómo las clases de noticias se relacionan con los instrumentos financieros. Hogenbooma et al. (2013) incorporaron información basada en noticias en un modelo de VaR para mejorar su desempeño para predecir eventos raros. Por último, Nassirtoussi et al. (2015) utilizaron titulares de noticias de Forex, Mahajan et al. (2008) noticias de capitalmarket.com, Schumaker y Chen (2009) Yahoo Finance, Huang et al. (2010) los principales diarios nacionales de Taiwán, y Groth y Muntermann (2011) anuncios ad-hoc, donde Yu et al. (2013) implementaron una mezcla de fuentes.

### **Fuzzy CoCo (Cooperativo y Coevolutivo)**

Durante la construcción del modelo “Fuzzy Rule Based System” (FRBS), el objetivo principal es identificar los parámetros para el sistema de inferencia difusa. El modelado evolutivo difuso apareció por primera vez en los trabajos de Karr (1991), después de lo cual se ha aplicado en muchas áreas, incluidas las finanzas (Kuo et al., 2001, Hadavandi et al. 2010, Kim et al., 2004). En un modelo FRBS, hay tres tipos diferentes de parámetros: ajuste del conocimiento (parámetros operacionales), aprendizaje del comportamiento (parámetros conectivos), aprendizaje de la estructura (parámetros estructurales) y para implementar el algoritmo evolutivo se aplican tres métodos diferentes, el enfoque de Michigan, el enfoque de Pittsburgh y el enfoque iterativo. El método Fuzzy CoCo aplica un método cooperativo coevolutivo al modelado difuso. Este enfoque se basa en la metodología propuesta por Potter y De Jong (1994). En un enfoque cooperativo coevolutivo para el cálculo evolutivo, dos elementos cooperan entre sí para obtener un recurso con aptitud acoplada. En un enfoque de Fuzzy CoCo, dos elementos cooperativos coevolutivos son, la base de datos (función de membresía) y la base de reglas. Así, un enfoque coevolutivo cooperativo para el modelado difuso es un proceso interconectado para la búsqueda de funciones de pertenencia que definen las variables difusas y la búsqueda de las reglas que realiza la inferencia difusa (Pena-Reyes y Sipper, 2001).

### **Descripción de los datos**

Los datos utilizados en esta investigación son de dos tipos. La primera categoría es la Apertura-Máxima-Mínima-Cierre diario de índices de mercado como el S&P 500, Promedio Industrial Dow Jones (DJIA), FTSE 100 y DAX30 y empresas como Facebook,



Apple Inc. y Citigroup. El segundo conjunto de datos consiste en las visitas a las páginas de Wikipedia de temas económicos como la “crisis de la deuda europea”, la deuda nacional de los EE. UU, temas como la Primavera Árabe y los artículos de Wikipedia que describen una empresa. Los datos elegidos fueron del período del 1 de enero de 2010 al 31 de diciembre de 2014, excepto el caso de Facebook, cuya salida a bolsa se lanzó a mediados de 2012. La razón por la que se eligió este período fue el volumen considerable de datos, marcado por una constante agitación financiera y social.

### Datos financieros

El primer conjunto de datos, es decir, los precios diarios de Apertura-Máxima-Mínima-Cierre de los índices de mercado y acciones de las empresas de tecnología, se utilizó para calcular los siguientes indicadores técnicos financieros. El uso de indicadores técnicos es una metodología común, y ha sido empleada en trabajos previos (Kim, 2003, Tsaih et al., 1998, Kara et al., 2011, Kim y Han, 2000, Patel et al., 2015). En esta investigación, se utilizó un conjunto de siete indicadores técnicos basados en tendencias y basados en impulso de uso común.

La media móvil exponencial (MME) es una técnica de análisis técnico que utiliza un enfoque de media móvil (MM) simple para calcular un precio actualizado que otorga mayores pesos a la MM calculada recientemente. En esta investigación se calculó un promedio móvil de 5 días, como regla general, si el precio está por encima de la MME, la tendencia es alcista y si no, la tendencia es a la baja. A continuación, se restó la MME del período de 5 días del precio de apertura del índice o precio de las acciones y se construyó el indicador de tendencia.

La media móvil exponencial doble (MMED), como sugiere su nombre, se calcula restando la MME de la MME de dos veces la MME. Como indicador direccional de tendencia, es más reactivo a los cambios de dirección que la MME. Se calculó una MMED de 10 días y de igual manera se restó del precio de apertura para encontrar tendencias alcistas o bajistas.

El índice de fuerza relativa (IFR) es un indicador basado en el impulso. Por lo general, varía entre 0 y 100. Un valor por encima de 70 representa una condición de sobrecompra por lo que en un futuro cercano los precios bajarán indicando un impulso a la baja, mientras que los valores por debajo de 30 representan una condición de sobreventa que indica un impulso alcista.

El indicador de convergencia y divergencia promedio (ICDP) se calcula restando la MME de 26 días de la MME de 12 días y luego se traza la MME de 9 días del ICDP como una línea de señal. En esta investigación, solo se tomó la línea de señal como indicador direccional de tendencia.

El índice de impulso estocástico (IIE) es una categoría de oscilador estocástico. Suelen oscilar entre -100 y +100. El IIE se utiliza generalmente para representar condiciones de sobrecompra y sobreventa, un valor superior a +40 indica tendencias alcistas, mientras que uno por debajo de -40 indica tendencias bajistas.

El indicador de banda de Bollinger consta de una línea central de la MME habitual y dos canales, uno hacia arriba y otro hacia abajo. Cuando el precio de las

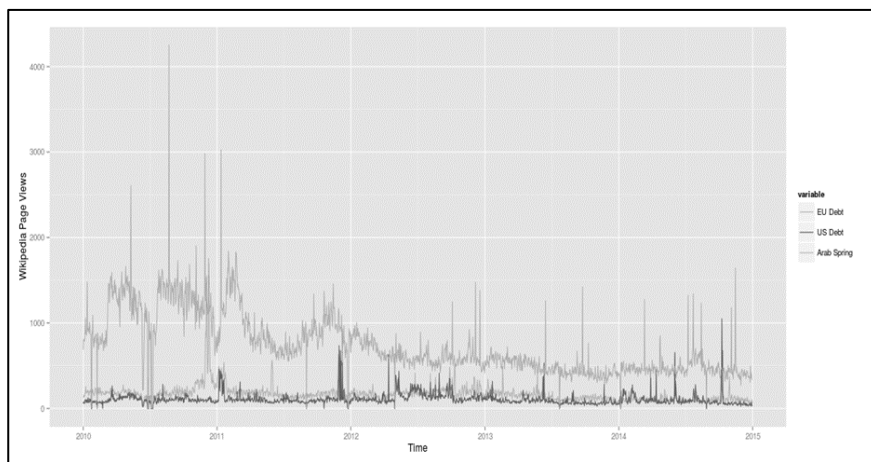
acciones sigue tocando la banda de Bollinger alcista, se cree que las acciones están sobrecompradas y viceversa. En esta investigación calculamos un porcentaje de banda de Bollinger, una medida adicional. Un porcentaje de Bollinger calcula el precio de un valor en relación con la banda de Bollinger hacia arriba y hacia abajo. El porcentaje de Bollinger se utilizó como medida en esta investigación para comprender las direcciones de las tendencias.

El índice de canal de commodity (ICC) calcula la relación entre la media móvil de las acciones y las desviaciones normales. Suele variar entre -100 y +100. Cuando el ICC supera los 100 indica una tendencia alcista y por debajo de -100 una tendencia bajista.

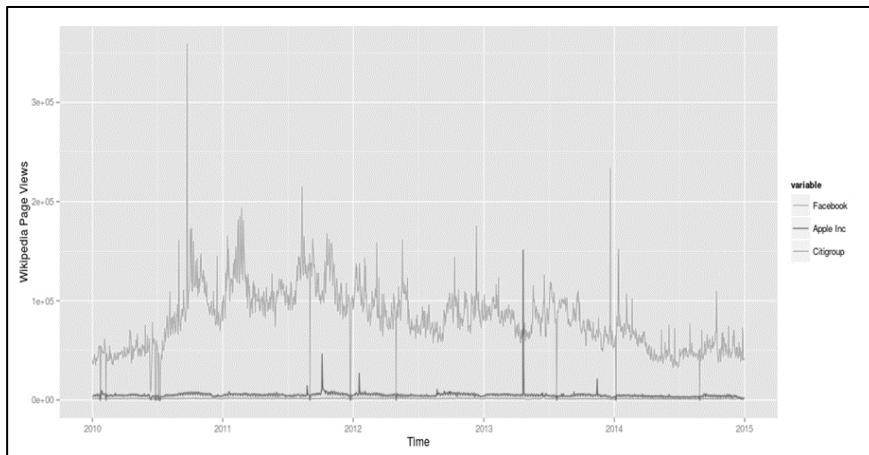
### Vistas de las páginas de Wikipedia

Los datos de visitas a la página de Wikipedia se descargaron del sitio web <http://stats.grok.se/>. Este consiste en las vistas diarias de los artículos de Wikipedia seleccionados. Como las visitas a las páginas de estos artículos abarcaban fines de semana y días festivos, se extrajeron para su análisis los datos que correspondían únicamente a los días de negociación. Estos dos tipos de datos se utilizaron para construir las variables predictoras que se utilizarán en la siguiente sección para entrenar y probar el modelo Fuzzy CoCo.

**Figura 2.** Vistas de las páginas de Wikipedia sobre artículos relacionados con la situación económica y política



**Figura 3.** Vistas de las páginas de Wikipedia sobre artículos relacionados con empresas



**Prueba de causalidad de Granger**

Se realizó un conjunto de prueba de causalidad de Granger (Granger, 1969) sobre las variables exógenas y sus correspondientes variables predictoras para estimar si son causadas por las variables exógenas de Granger o no. Según el concepto de causalidad de Granger "X Granger causa Y" si y solo si los valores pasados de X ayudan a predecir los cambios en Y, si no, entonces "X no Granger causa Y". Las hipótesis relacionadas de acuerdo con los tres tipos de vistas de página de Wikipedia seleccionadas se formulan y presentan en la Tabla 1. Luego, se realiza una prueba de causalidad de Granger por pares para encontrar la significancia estadística.

**Tabla 1.** Formulación de hipótesis de causalidad de Granger

Hipótesis Nula H0	Hipótesis Alternativa H1
Las visitas a páginas de Wikipedia de artículos relacionados con empresas (Facebook, Apple, Citigroup) no tienen causalidad de Granger con respecto al precio de las acciones de las respectivas empresas.	Visitas a páginas de Wikipedia de artículos relacionados con empresas individuales (Facebook, Apple, Citigroup) Granger provoca el movimiento del precio de las acciones de las respectivas empresas.
Las visitas a páginas de Wikipedia de artículos relacionados con la situación económica en los países (Deuda nacional de EE. UU., Crisis de deuda europea) no tienen causalidad de Granger con respecto a precios de los principales índices de mercado de la UE (DAX, FTSE 100) y EE. UU. (S & P500, DJIA).	Visitas a páginas de Wikipedia de artículos relacionados con la situación económica en los países (Deuda nacional de EE. UU., Crisis de deuda europea) Granger provoca movimientos de precios de los principales países de la UE (DAX, FTSE 100) y EE. UU. (S & P500, DJIA)

Las visitas a páginas de Wikipedia de artículos relacionados con eventos sociopolíticos (Primavera Árabe) no tienen causalidad de Granger con respecto a los precios de los principales índices de mercado de la UE (DAX, FTSE 100) y EE. UU. (S & P500, DJIA)	Visitas a páginas de Wikipedia de artículos relacionados con eventos sociopolíticos (Primavera Árabe) Granger provocan movimientos de precios de los principales índices de mercado de la UE (DAX, FTSE 100) y EE. UU. (S & P500, DJIA)
--	---

**Tabla 2.** Resultados de causalidad de Granger de Facebook, Google y Apple Inc con sus correspondientes vistas de página de Wikipedia

Variable independiente	Variable dependiente	Lag	F statistic	p- Valor
Vistas de la página wiki de Facebook	Precio de apertura de las acciones de Facebook	1	60.734	2.572e-14
Vistas de la página wiki de Apple Inc.	Precio de apertura de las acciones de Apple	1	8.7036	0.003235
Vistas de la página wiki de Citigroup Inc	Precio de apertura de las acciones de Citigroup	1	7.8177	0.005253
Vistas de la página de Wikipedia de la crisis de la deuda de la UE	FTSE100	1	14.434	0.0001519
Vistas de la página de Wikipedia de la crisis de la deuda de la UE	DAX30	1	18.197	2.139e-05
Vistas de la página de Wikipedia de la deuda nacional de EE. UU.	DJIA	1	8.0174	0.004707
Vistas de la página de Wikipedia de la deuda nacional de EE. UU.	S&P500	1	9.0423	0.00269
Wikipedia Vistas de la página del artículo Primavera Árabe	FTSE100	1	57.455	6.569e-14
Wikipedia Vistas de la página del artículo Primavera Árabe	DAX30	1	67.479	5.177e-16
Wikipedia Vistas de la página del artículo Primavera Árabe	DJIA	1	120.17	2.2e-16
Wikipedia Vistas de la página del artículo Primavera Árabe	S&P500	1	99.363	2.2e-16

### Configuración experimental y resultados

Como el conjunto de datos formado consta de atributos de páginas vistas de Wikipedia y tendencias de varios indicadores técnicos, fue necesario normalizarlos para tener una media cero y una desviación estándar de 1, este procedimiento de normalización lleva todos los valores a la misma escala. El entrenamiento y las pruebas de los modelos de aprendizaje automático se realizaron dividiendo la muestra en un

conjunto de entrenamiento de muestra y un conjunto de prueba fuera de muestra. El conjunto de datos hasta el período que comienza en enero de 2014 se usa para el entrenamiento de muestra y el resto de los datos no vistos se usa para pruebas fuera de muestra. El objetivo principal es predecir el movimiento del índice de mercado o de las acciones en un día de negociación. Entonces, el problema da como resultado un problema de clasificación binaria de si el precio subirá o bajará al día siguiente.

Como se indica en la sección de descripción de los datos, se crearon 4 conjuntos de datos diferentes para todos los conjuntos de hipótesis posibles. Algunas de las métricas de rendimiento comunes que se utilizan para medir los resultados de clasificación para la predicción de los movimientos del mercado son exactitud, sensibilidad o tasa de aciertos, especificidad, precisión o valor positivo predicho y valor negativo predicho según lo utilizado por Patel et al. (2015) y Schumaker y Chen (2009). Las siguientes ecuaciones definen las métricas.

$$\textit{Sensibilidad} = \frac{TP}{TP+FN'} \quad (1)$$

$$\textit{Especificidad} = \frac{TN}{FP+TN'} \quad (2)$$

$$\textit{Valor Neg. Pred.} = \frac{TN}{TN+FN'} \quad (3)$$

$$\textit{Valor Pos. Pred.} = \frac{TP}{TP+FP'} \quad (4)$$

$$\textit{Exactitud} = \frac{TP+TN}{TP+FP+FN+TN'} \quad (5)$$

Aquí TP es verdadero positivo, FP es falso negativo, TN y FN es verdadero negativo y falso negativo respectivamente.

A continuación, se enumeran las tablas con los valores de la precisión de la clasificación para la predicción de las visitas a Wikipedia, de acuerdo.

**Tabla 3.** Precisión de clasificación para la predicción de las visitas a Wikipedia relacionadas con la deuda nacional de los EE. UU. en datos de la muestra

Indice	Exactitud	Sensibilidad	Especificidad	Valor Pos. Pred.	Valor Neg. Pred.
DJIA	80.89	70.71	88.89	83.33	79.43
S&P500	80.44	74.23	85.16	79.12	81.34

**Tabla 4.** Precisión de la clasificación para la predicción de las visitas a Wikipedia relacionadas con la crisis de la deuda de la UE en datos de la muestra

Indice	Exactitud	Sensibilidad	Especificidad	Valor Pos. Pred.	Valor Neg. Pred.
FTSE100	86,36	72,63	84,80	87,77	86,18
DAX30	76,83	76,56	77,12	78,40	75,21

**Tabla 5.** Precisión de clasificación para la predicción de las visitas a Wikipedia relacionadas con acontecimientos sociales y políticos como la Primavera Árabe en datos de la muestra

Indice	Exactitud	Sensibilidad	Especificidad	Valor Pos. Pred.	Valor Neg. Pred.
DJIA	80,44	74,75	84,92	79,57	81,06
S&P500	82,22	76,29	86,72	81,32	82,84
FTSE100	55,84	49,54	61,48	53,47	57,69
DAX30	78,46	75,78	81,36	81,51	75,59

**Tabla 6.** Precisión de clasificación para la predicción de visitas a Wikipedia relacionadas con empresas específicas en datos de la muestra

Indice	Exactitud	Sensibilidad	Especificidad	Valor Pos. Pred.	Valor Neg. Pred.
Facebook	83,33	82,14	84,29	80,70	85,51
Apple Inc	68,89	60,91	76,52	71,28	74,34
Citigroup	77,33	88,89	66,67	71,11	86,67

### Conclusión y trabajos futuros

El objetivo que se propuso alcanzar en este trabajo fue doble. El primer objetivo era ver si las visitas a las páginas de Wikipedia de artículos relacionados con la deuda económica como la “crisis de la deuda europea”, las condiciones sociopolíticas de la “Deuda nacional de EE. UU.” La “Primavera Árabe” y las páginas empresas como “Facebook”, “Apple Inc.” y “Citigroup” se pueden utilizar como un indicador de sentimiento social junto con los indicadores técnicos de uso común para la predicción de los movimientos del mercado. Al principio se llevó a cabo una serie de pruebas de causalidad de Granger. Esta intenta encontrar si el valor rezagado de una serie de tiempo puede predecir otra serie de tiempo. Las pruebas de causalidad realizadas en la sección 4 muestran una significancia estadística alta y media. En el trabajo de Moat et al. (2013) y Preis et al. (2013), los temas relacionados con la economía fueron de naturaleza muy general, el aporte agregado que se presenta en este artículo con respecto a los temas,

son específicos y han tratado de encontrar si la crisis económica nacional o regional ha afectado el principal índice bursátil de un país. Con respecto a la crisis regional, se utilizó la página de Wikipedia de la "crisis de la deuda europea" junto con los siete indicadores técnicos para comprender su uso como indicador de mercado para el FTSE 100 y el DAX 30, el índice más importante de Reino Unido y Alemania, de manera similar, la "Deuda nacional de EE. UU." con S&P500 y DJIA. Los resultados obtenidos para predecir la precisión direccional para los mercados demuestran que las visitas a la página se pueden utilizar junto con los indicadores técnicos para la predicción de los mercados. Los préstamos grandes y considerables proporcionados por los bancos británicos (Popov y Van Horen, 2013) y las propuestas de rescate de Alemania son probablemente las principales razones por las que los mercados se vieron afectados; se puede aplicar una explicación similar para los mercados estadounidenses.

La segunda categoría de artículos seleccionados pertenecía a acontecimientos sociopolíticos. En este caso se eligió la Primavera Árabe, que fue una sucesión de importantes levantamientos políticos que tuvieron lugar en Oriente Medio y países del norte de África. Anteriormente, Abdelbaki (2013) ha propuesto la idea del efecto que tiene la Primavera Árabe en la libra egipcia y el dólar estadounidense y en importantes índices del mercado egipcio, EGX30 y EGX70. Chau et al. (2014) han analizado el efecto de la Primavera Árabe en los mercados de los países de Medio Oriente y África del Norte y en los mercados de los países desarrollados. Los resultados muestran que, aunque tuvo efectos en los mercados regionales, apenas afectó a los mercados globales. Los resultados de precisión para índices globales como DJIA y S&P500 son prometedores en comparación con FTSE100 y DAX30. Una posible razón puede ser que DJIA y S&P500 albergan pocas compañías importantes de petróleo y gas en comparación con los índices europeos como FTSE100. Este resultado apoya el trabajo de Filis et al. (2011).

La tercera categoría de artículos que se eligió pertenecía a la descripción de empresas. Se eligió un conjunto de empresas de alta tecnología como Facebook y Apple Inc. junto con Citigroup en banca y finanzas. Los resultados de Facebook fueron mucho más prometedores que los de Apple. La OPI (Oferta Pública Inicial) de Facebook se lanzó en mayo de 2012 y había sido objeto de muchas controversias, Facebook también fue objeto de demandas más tarde. Apple Inc, por otro lado, tuvo un crecimiento estable. Los resultados de precisión para Citigroup estaban entre Facebook y Apple, y un vistazo rápido a la prueba de causalidad de Granger revelaron que los resultados están en orden de importancia. Se ha demostrado en investigaciones anteriores que el sentimiento de los inversores puede influir en los precios de las acciones. El objetivo de la investigación realizada fue ver si se puede utilizar en forma de visitas a la página de Wikipedia para predecir los movimientos del precio de las acciones.

El segundo objetivo que se planteó fue determinar si los resultados de predicción logrados pueden ser mejores que los resultados informados anteriormente. Para los resultados informados anteriormente, nuestro punto de referencia para la comparación se realizó en Arias et al. (2013) y Bollen et al. (2011). En el caso de Bollen et al. (2011) el nivel de precisión alcanzado fue del 86%, igual que en el caso nuestro, pero el conjunto de datos utilizado fue mucho menor desde octubre de 2008 a diciembre de 2008, y se utilizaron menos eventos sociopolíticos como las elecciones presidenciales de EE. UU. y

el día de Acción de Gracias, donde como en nuestro caso, el conjunto de datos era de 5 años y los casos iban desde sucesos sociopolíticos hasta crisis económicas y descripciones de empresas. En el segundo artículo, Arias et al. (2013) aplicaron un enfoque similar. Los niveles de precisión para el retraso 2 en el caso de predecir VXO la volatilidad implícita de S&P100 con SVM que tiene un núcleo sigmoide fue del 90,67%, luego, nuevamente, el conjunto de datos empleado fue por un breve período de tiempo, del 22 de marzo al 18 de noviembre de 2011. Ahora comparando ambos trabajos con la investigación realizada, se puede concluir que, aunque en el segundo caso la precisión de la predicción alcanzada es mayor, el sentimiento de Twitter relacionado con dicha predicción no representa ningún evento o acontecimiento en particular y es de naturaleza mucho más general y el conjunto de datos utilizado es mucho más corto. Por lo tanto, la pregunta sigue siendo si tales niveles de precisión pueden repetirse durante períodos de tiempo mucho más largos y eventos diversos. Así, desde el punto de vista de la fiabilidad y la validez, esta investigación realizada es mucho más robusta.

La implicación gerencial de esta investigación tiene dos contribuciones significativas. La primera contribución es la integración de las vistas de páginas de Wikipedia en R (software R para computación estadística). No se requiere más procesamiento y creación de series de tiempo como se hizo en investigaciones anteriores. La segunda contribución es la incorporación de un indicador social (crisis económica, levantamiento político social, información de la empresa) que refleja el sentimiento del mercado en la metodología de modelado, lo que brinda a los gerentes financieros una herramienta para tomar decisiones comerciales bien informadas.

La limitación asociada con esta investigación es el uso de un solo modelo de aprendizaje automático. La generalización de la investigación se restringe un poco debido al uso de solo el modelo FuzzyCoCo. Por lo tanto, la investigación futura se dirigirá a la comparación entre otros modelos de aprendizaje automático de uso habitual como redes neuronales, "Support Vector Machines" o "Random Forest". La otra posible dirección de investigación futura puede ser la comparación de diferentes índices basados en redes sociales como las tendencias de Google, las visitas a páginas de Wikipedia y las series de tiempo de Twitter para predecir los movimientos del mercado.



## Referencias

- Abdelbaki, H. H., 2013. The impact of arab spring on stock market performance. *British Journal of Management & Economics* 3 (3).
- Antweiler, W., Frank, M. Z., 2004. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance* 59 (3), 1259–1294.
- Arias, M., Arratia, A., Xuriguera, R., 2013. Forecasting with twitter data. *ACM Transactions on Intel ligent Systems and Technology (TIST)* 5 (1), 8.
- Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., Goldstein, G., 2011. Identifying and following expert investors in stock microblogs. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 1310–1319.
- Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2 (1), 1–8.
- Bollen, J., Pepe, A., Mao, H., 2009. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *arXiv preprint arXiv:0911.1583*.
- Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., Weber, I., 2012. Web search queries can predict stock market volumes. *PloS one* 7 (7), e40014.
- Chau, F., Deesomsak, R., Wang, J., 2014. Political uncertainty and stock market volatility in the middle east and north african (mena) countries. *Journal of International Financial Markets, Institutions and Money* 28, 1–19.
- Das, S. R., Chen, M. Y., 2007. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science* 53 (9), 1375–1388.
- Dimpfl, T., Jank, S., 2012. Can internet search queries help to predict stock market volatility? In: *Paris December 2012 Finance Meeting EUROFIDAI- AFFI Paper*.
- Dzielinski, M., 2012. Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters* 9 (3), 167–175.
- Esteban-Gil, A., Garcia-Sanchez, F., Valencia-Garcia, R., Fernandez-Breis, J. T., 2012. Socialbroker: A collaborative social space for gathering semantically-enhanced financial information. *Expert Systems with Applications* 39 (10), 9715–9722.
- Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work\*. *The journal of Finance* 25 (2), 383–417.
- Filis, G., Degiannakis, S., Floros, C., 2011. Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. *International Review of Financial Analysis* 20 (3), 152–164.
- Fung, G. P. C., Yu, J. X., Lam, W., 2003. Stock prediction: Integrating text mining approach using real-time news. In: *Computational Intel ligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference on*. IEEE, pp. 395–402.
- Granger, C. W., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.
- Groth, S. S., Muntermann, J., 2011. An intra day market risk management approach based on textual analysis. *Decision Support Systems* 50 (4), 680– 691.

- Hadavandi, E., Shavandi, H., Ghanbari, A., 2010. Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based Systems* 23 (8), 800–808.
- Hogenboom, F., de Wintera, M., Frasinca, F., Kaymakb, U., 2013. A news event-driven approach for the historical value at risk method. *Expert Systems With Applications*.
- Hong, T., Han, I., 2002. Knowledge-based data mining of news information on the internet using cognitive maps and neural networks. *Expert systems with applications* 23 (1), 1–8.
- Huang, S.-C., Chuang, P.-J., Wu, C.-F., Lai, H.-J., 2010. Chaos-based support vector regressions for exchange rate forecasting. *Expert Systems with Applications* 37 (12), 8590–8598.
- Kara, Y., Boyacioglu, M. A., Baykan, O. K., 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange. *Expert systems with Applications* 38 (5), 5311–5319.
- Karabulut, Y., 2013. Can Facebook predict stock market activity? In: *AFA 2013 San Diego Meetings Paper*.
- Karr, C., 1991. Genetic algorithms for fuzzy controllers. *Ai Expert* 6 (2), 26–33.
- Kim, K.-j., 2003. Financial time series forecasting using support vector machines. *Neurocomputing* 55 (1), 307–319.
- Kim, K.-j., Han, I., 2000. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert systems with Applications* 19 (2), 125–132.
- Kim, M. J., Han, I., Lee, K. C., 2004. Hybrid knowledge integration using the fuzzy genetic algorithm: Prediction of the korea stock price index. *Intelligent Systems in Accounting, Finance and Management* 12 (1), 43–60.
- Kim, T., Jung, W.-J., Lee, S.-Y. T., 2014. The analysis on the relationship between firms exposures to sns and stock prices in Korea. *Asia Pacific Journal of Information Systems* 24 (2).
- Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N., 2013. Ontology- based sentiment analysis of twitter posts. *Expert systems with applications* 40 (10), 4065–4074.
- Kuo, R. J., Chen, C., Hwang, Y., 2001. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy sets and systems* 118 (1), 21–45.
- Liu, L., Wu, J., Li, P., Li, Q., 2015. A social-media-based approach to predicting stock comovement. *Expert Systems with Applications*.
- Lo, A. W., 2004. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management* 30, 15–29.
- Lupiani-Ruiz, E., Garcia-Manotas, I., Valencia-Garcia, R., Garcia-Sanchez, F., Castellanos-Nieves, D., Fernandez-Breis, J. T., Camón-Herrero, J. B., 2011. Financial news semantic search engine. *Expert systems with applications* 38 (12), 15565–15572.
- Mahajan, A., Dey, L., Haque, S. M., 2008. Mining financial news for major events and their impacts on the market. In: *Web Intel ligence and Intel ligent Agent*

- Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, pp. 423–426.
- Maks, I., Vossen, P., 2012. A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems* 53 (4), 680–688.
- Mittermayer, M.-A., 2004. Forecasting intraday stock price trends with text mining techniques. In: *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on. IEEE*, pp. 10–pp.
- Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., Preis, T., 2013. Quantifying wikipedia usage patterns before stock market moves. *Scientific reports* 3.
- Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., Ngo, D. C. L., 2014. Text mining for market prediction: A systematic review. *Expert Systems with Applications* 41 (16), 7653–7670.
- Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., Ngo, D. C. L., 2015. Text mining of news-headlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications* 42 (1), 306–324.
- Nizer, P., Nievola, J. C., 2012. Predicting published news effect in the brazilian stock market. *Expert Systems with Applications* 39 (12), 10674–10680.
- Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015. Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications* 42 (4), 2162–2172.
- Pena-Reyes, C. A., Sipper, M., 2001. Fuzzy CoCo: A cooperative-coevolutionary approach to fuzzy modeling. *Fuzzy Systems, IEEE Transactions on* 9 (5), 727–737.
- Peramunetil leke, D., Wong, R. K., 2002. Currency exchange rate forecasting from news headlines. *Australian Computer Science Communications* 24 (2), 131–139.
- Popov, A. A., Van Horen, N., 2013. The impact of sovereign debt exposure on bank lending: Evidence from the european debt crisis.
- Potter, M. A., De Jong, K. A., 1994. A cooperative coevolutionary approach to function optimization. In: *Parallel problem solving from nature PPSN III. Springer*, pp. 249–257.
- Preis, T., Moat, H. S., Stanley, H. E., 2013. Quantifying trading behavior in financial markets using google trends. *Scientific reports* 3.
- Rachlin, G., Last, M., Alberg, D., Kandel, A., 2007. Admiral: A data mining based financial trading system. In: *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on. IEEE*, pp. 720–725.
- Schumaker, R. P., Chen, H., 2009. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)* 27 (2), 12.
- Soni, A., van Eck, N. J., Kaymak, U., 2007. Prediction of stock price movements based on concept map information. In: *Computational Intelligence in Multicriteria Decision Making, IEEE Symposium on. IEEE*, pp. 205–211.
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., Welpe, I. M., 2014. Tweets and trades: The information content of stock microblogs. *European Financial Management* 20 (5), 926–957.

- Sprenger, T. O., Welpe, I. M., 2011. Tweets and peers: defining industry groups and strategic peers based on investor perceptions of stocks on twitter. *Algorithmic Finance* 1 (1), 57–76.
- Tetlock, P. C., SAAR-TSECHANSKY, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63 (3), 1437–1467.
- Tsaih, R., Hsu, Y., Lai, C. C., 1998. Forecasting S&P 500 stock index futures with a hybrid ai system. *Decision Support Systems* 23 (2), 161–174.
- Vlastakis, N., Markellos, R. N., 2012. Information demand and stock market volatility. *Journal of Banking & Finance* 36 (6), 1808–1821.
- Vu, T.-T., Chang, S., Ha, Q. T., Collier, N., 2012. An experiment in integrating sentiment features for tech stock prediction in twitter.
- Wang, S., Zhe, Z., Kang, Y., Wang, H., Chen, X., 2008. An ontology for causal relationships between news and financial instruments. *Expert Systems with Applications* 35 (3), 569–580.
- Wuthrich, B., Cho, V., Leung, S., Permunetille, D., Sankaran, K., Zhang, J., 1998. Daily stock market forecast from textual web data. In: *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*. Vol. 3. IEEE, pp. 2720–2725.
- Xu, S. X., Zhang, X. M., 2013. Impact of wikipedia on market information environment: Evidence on management disclosure and investor reaction. *Mis Quarterly* 37 (4), 1043–1068.
- Yu, Y., Duan, W., Cao, Q., 2013. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems* 55 (4), 919–926.
- Zhang, X., Fuehrer, H., Gloor, P. A., 2012. Predicting asset value through Twitter buzz. In: *Advances in Collective Intelligence 2011*. Springer, pp. 23–34.