

No.29

AGOSTO DE 2019

Documentos CEDE

ISSN 1657-7191 Edición electrónica.

Análisis de sentimientos de noticias e
inversionistas en el mercado bursátil

Germán Eduardo González



Serie Documentos Cede, 2019-29
ISSN 1657-7191 Edición electrónica.
Agosto de 2019

© 2019, Universidad de los Andes, Facultad de Economía,
CEDE. Calle 19A No. 1 – 37 Este, Bloque W.
Bogotá, D. C., Colombia Teléfonos: 3394949- 3394999,
extensiones 2400, 2049, 2467
infocede@uniandes.edu.co
<http://economia.uniandes.edu.co>

Impreso en Colombia – Printed in Colombia

La serie de Documentos de Trabajo CEDE se circula con propósitos de discusión y divulgación. Los artículos no han sido evaluados por pares ni sujetos a ningún tipo de evaluación formal por parte del equipo de trabajo del CEDE.

El contenido de la presente publicación se encuentra protegido por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por tanto su utilización, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso, digital o en cualquier formato conocido o por conocer, se encuentran prohibidos, y sólo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito del autor o titular. Las limitaciones y excepciones al Derecho de Autor, sólo serán aplicables en la medida en que se den dentro de los denominados Usos Honrados (Fair use), estén previa y expresamente establecidas, no causen un grave e injustificado perjuicio a los intereses legítimos del autor o titular, y no atenten contra la normal explotación de la obra.

Universidad de los Andes | Vigilada Mineducación
Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

Análisis de sentimientos de noticias e inversionistas en el mercado bursátil *

Germán Eduardo González[†]

Resumen

Este trabajo estudia en la relación estadística entre el sentimiento de inversionistas y de noticias con la dirección del movimiento del S&P 500 en ventanas de 30 minutos. Específicamente, se utilizan algoritmos de Machine Learning para estimar una medida representativa del sentimiento de los inversionistas y de noticias a partir de mensajes de usuarios de StockTwits, que luego se utilizaron de manera conjunta e individual para pronosticar la dirección del SP 500. Se encontró que con la inclusión del sentimiento de los inversionistas se alcanzó una exactitud del 58%, por encima del 50% obtenido por un modelo autorregresivo. No obstante, se determinó que la integración del análisis de sentimiento de noticias no aumentó la predictibilidad del índice en cuestión.

Palabras clave: Predicción de la dirección de los retornos intradía, S&P 500, análisis de sentimiento de los inversionistas, y análisis de sentimiento de noticias.

Clasificación JEL: G11, G14, G15, G17 y G41.

*Tesis para la Maestría en Economía, Universidad de los Andes.

[†]Estoy profundamente agradecido con Diego Jara por el tiempo dedicado a leer este documento y a sus comentarios. También me gustaría agradecer a los participantes del seminario de matemáticas financieras, y a los investigadores de Quantil por las discusiones constructivas y retadoras.

News and investors sentiments analysis in the stock market ^{*}

Germán Eduardo González [†]

Abstract

This article studies the statistical relationship between investors' sentiments and news with the direction of the S&P 500 in a 30-minute time frame. Specifically, this paper applies Machine Learning algorithms to estimate a representative measure of investors sentiment and news using messages from StockTwits users. Later, we used that information to forecast the direction of the S&P 500. We found that by using investor' sentiments, the accuracy of the S&P prediction improved to 58% from 50% obtained by an autoregressive model. However, we determined that using StockTwits' news did not enhance the model's predictive power for the S&P 500.

Keywords: Forecast of the directions of the intraday yields, S&P 500, sentiment analysis, invertors, news, tweets, and StockTwets.

JEL Classification: G11, G14, G15, G17 y G41.

^{*}Thesis for the Masters in Economics, Universidad de los Andes.

[†]Universidad de los Andes Quantil | Matemáticas financieras, ge.gonzalez10@uniandes.edu.co. I am deeply grateful with Diego Jara for the time reading this document and his invaluable comments. I would also like to thank to the financial mathematics seminar, and Quantil researchers for their productive and challenging discussions.

1 Introducción

En las décadas más recientes, las firmas y los hogares que conforman los mercados bursátiles están empleando nuevas tecnologías de información que van desde el consumo de artículos financieros hasta el uso de redes sociales para expresar sus sentimientos hacia un determinado mercado. Particularmente, desde la economía financiera se puede incorporar y agregar esta nueva información con el objetivo de identificar patrones que permitan modelar la dirección de los retornos de acciones, e incluso alertar sobre posibles burbujas en un mercado determinado.

Dado lo anterior, la pregunta que se responde en este documento es ¿cuál es el grado de la predictibilidad de la dirección de los movimientos de corto plazo del S&P 500 a partir de los sentimientos de los inversionistas publicados en StockTwits? Esta investigación busca responder esta pregunta por medio de una medida representativa del sentimiento de los inversionistas, que luego es complementada con la inclusión de un canal constituido a partir de noticias que permiten construir juicios a los inversionistas respecto al desempeño futuro del índice.

Esta investigación surge en un contexto de modelos de valoración (CAPM, CBM, HJM, entre otros), medidas de riesgo y pronósticos que utilizan el supuesto de no arbitraje para la estimación del rendimiento esperado de un activo. Si bien estos modelos son válidos en diferentes contextos, surge un grupo de autores que se centran en estudiar los escenarios en donde el supuesto de no arbitraje entra en conflicto con asimetrías de información (Grossman & Stiglitz, 1980). Es por lo anterior que surge la necesidad de profundizar en aproximaciones que permitan la inclusión de asimetrías de información generadas por el comportamiento humano dentro de la economía financiera.

Uno de los primeros autores en cuestionarse sobre el efecto del comportamiento humano en los mercados bursátiles fue Keynes en 1936. Particularmente, Keynes le dio una importancia al sentimiento de los inversionistas en el mercado, hasta el punto en el que creía que existían situaciones en que el mercado era guiado por un *“espíritu animal”* y estaba sujeto a *“oleadas de sentimientos optimistas o pesimistas, que son irracionales y sin embargo legítimos”* (Keynes, 2014, p. 141). Si bien Keynes consideraba que la psicología humana tenía que tenerse en cuenta a la hora de analizar los mercados financieros, esta aproximación quedó en el olvido. Solo fue hasta la década de los años 70 en donde el concepto del *“espíritu animal”* fue

1. INTRODUCCIÓN

retomado para explicar dinámicas económicas como el aumento de los precios del petróleo generado por la OPEP (Akerlof & Shiller, 2010), entre otros fenómenos inexplicables para la doctrina económica dominante.

Posteriormente, las ideas de Keynes conquistaron otras ramas de la economía que adaptaron este concepto a otros contextos. Una muestra de lo anterior son los planteamientos de De Long, Shleifer, Summers & Waldmann (1990) que, desde las finanzas conductuales, argumentan que el precio de un activo puede desviarse temporalmente de su valor fundamental en presencia de inversionistas guiados por sentimientos basados en creencias estocásticas erróneas, lo cual limita la condición de no arbitraje. Adicionalmente, trabajos como Kahneman & Tversky (2013) profundizan en la racionalidad de los individuos en escenarios que involucran estrés desde la teoría prospectiva, a partir de un modelo de elección las decisiones de los agentes ante distintos niveles de aversión al riesgo de los inversionistas.

En la última década y tras la crisis del 2007, los planteamientos de Keynes han resurgido e influenciado a una generación de autores como Renault (2017) que, a partir de un modelo de análisis de sentimiento¹, evalúa la tesis propuesta por De Long et al. (1990) y utilizando tweets extraídos de StockTwits². Renault (2017) planteó un modelo continuo que utilizó como variable independiente el sentimiento promedio de los inversionistas de StockTwits, y variable dependiente los retornos del S&P 500 en ventanas de 30 minutos. El autor concluye que el primer cambio de media hora en los sentimientos del inversionista predice el retorno de la última media hora del S&P 500. Al igual que Renault (2017), existen otros autores que proponen metodologías de análisis de sentimiento para el pronóstico de los retornos de diferentes índices accionarios. Sin embargo, en la literatura reciente no se identificaron investigaciones que aborden paralelamente el análisis de sentimiento de los inversionistas con el análisis de sentimiento de noticias. Los dos análisis se abordan de manera independiente, por lo que una contribución a la literatura consiste en integrar estos análisis.

El primer aporte de esta investigación consiste en evaluar con modelos discretos si el efecto de los sentimientos embebidos de los comentarios de los inversionistas en StockTwits son predictores de la dirección del movimiento del S&P 500 en ventanas de 30 minutos. El segundo aporte de esta investigación se fundamenta en determinar si al incorporar un canal de sentimiento constituido a partir de noticias aumenta la predictibilidad de la dirección del

¹Entendiendo el análisis de sentimiento como el procesamiento de lenguaje natural para identificar y extraer valoraciones subjetivas acerca de un determinado tema en documentos digitales (Li, Xie, Chen, Wang & Deng, 2014).

²Página web (<https://stocktwits.com>) que almacena y organiza tweets relacionados con mercados bursátiles y finanzas.

1. INTRODUCCIÓN

movimiento del S&P 500. Esta investigación se diferencia de otros trabajos como el de Renault (2017) en la inclusión de un modelo de análisis de sentimiento de noticias, y en el uso de modelos discretos no lineales para estimar la forma funcional entre las variables de sentimiento y pronóstico.

En este orden de ideas, el objetivo principal de esta investigación consiste en evaluar si el efecto de los sentimientos embebidos en los comentarios de los inversionistas en StockTwits y en noticias de diferentes portales tienen poder predictivo en los retornos diarios del índice. Para lograr lo anterior, se plantearon tres objetivos secundarios que permiten aproximarse al objetivo principal.

El primer objetivo secundario consiste en la construcción de una medida de sentimiento de los inversionistas y de las noticias que permita estimar al sentimiento que tienen los inversionistas de StockTwits sobre el mercado y las noticias embebidas en sus mensajes. Debido a que no todos los mensajes de los inversionistas de StockTwits se encuentran marcados se construyeron dos modelos no supervisados que permite clasificar el sentimiento de millones de mensajes y noticias que no tienen un sentimiento. Para lo anterior, se utilizó una red neuronal convolucional que captura el sentimiento de los documentos en bigramas, trigramas y cuatrigamas. El segundo objetivo secundario profundiza en la construcción de un modelo discreto que permite la evaluar el efecto de la inclusión de los sentimientos de los inversionistas y las noticias sobre la dirección de los retornos del S&P 500. Adicionalmente, con la finalidad de filtrar los efectos del sentimiento de los inversionistas el tercer objetivo secundario es el diseño de una metodología de agentes heterogéneos que permita castigar el sentimiento de los agentes menos informados. Asimismo se exploró una metodología de agrupación por tópicos de noticias que permitió evaluar el efecto del canal de noticias de manera desagregada.

Teniendo en cuenta los objetivos esbozados anteriormente, esta investigación se puede catalogar como empírica, en la medida en la que hará uso de información pública (tweets extraídos de StockTwits y noticias tomadas de agencias informativas) para comprender los mecanismos que afectan y rigen los mercados bursátiles. Como se mencionó anteriormente, esta investigación evaluó el efecto del sentimiento de los inversionistas y las noticias sobre la dirección de los retornos del S&P 500, debido a la gran cantidad de información disponible. Sin embargo, esta investigación es generalizable a cualquier acción o índice accionario que

2. REVISIÓN DE LITERATURA

disponga de un volumen de información significativo.

Esta investigación encontró que la inclusión del sentimiento de los inversionistas homogéneos aumentó la predictibilidad de la dirección de los retornos en un 16% respecto a un modelo autorregresivo bajo la medida de exactitud. En otras palabras, si este modelo se hubiese utilizado desde el 2017-11-17 hasta el 2018-08-01, se habría predicho correctamente 858 subidas de un total de 1.119, y 455 caídas de un total de 1.053 y 0 movimientos neutrales de 78. Adicionalmente, se determinó que la inclusión del análisis de sentimiento de noticias no mejora la predicción de la dirección de los retornos del S&P 500, y no supera a un modelo autorregresivo de orden 1 bajo ninguna de las métricas en el análisis completo y segmentado de noticias. Por otra parte, se encontró que la inclusión de usuarios heterogéneos en el análisis de sentimiento de los inversionistas y la disminución de las ventanas de 30 a 15 minutos aumenta la predictibilidad de la dirección de los retornos del S&P 500.

Este documento se estructura en ocho secciones. La primera es la introducción ya expuesta. La segunda consiste en la revisión de literatura que profundiza en los avances que han realizado en materia de esta investigación y los aportes que son valiosos para esta. La tercera sección busca esbozar la metodología de esta investigación, profundizando en el modelo, los supuestos y la intuición que existe detrás de estos, así como las ventajas y desventajas que tiene el método utilizado. La cuarta sección describe el tratamiento de datos y unas breves estadísticas descriptivas de los mismos. La quinta sección presenta los resultados del modelo propuesto. La sexta sección se centra en análisis complementarios en el que se evalúan los supuestos del modelo propuesto. Finalmente, en la séptima sección se esbozan unas conclusiones y se plantea una discusión para posibles trabajos futuros.

2 Revisión de literatura

La asociación entre el movimiento del mercado y el análisis de sentimiento se ha abordado desde diferentes aproximaciones y metodologías que buscan comprender el mercado bajo un contexto de asimetrías de información. Esta revisión de literatura se divide en tres secciones: la teórica, el análisis de sentimiento de los inversionistas y el análisis fundamental. La primera sección aborda un grupo de trabajos teóricos que argumentan que es posible generar un margen de ganancia, por medio de algoritmos que aprovechan las asimetrías de información para procesar grandes cantidades de datos con el objetivo de identificar señales de mercado a altas velocidades. La segunda y la tercera sección presentan el análisis de

2. REVISIÓN DE LITERATURA

sentimiento de los inversionistas y de noticias respectivamente, para así mostrar la desconexión que existe entre estos dos análisis. Esta revisión de literatura sienta las bases teóricas para construir un algoritmo que explote las asimetrías de información en el mercado bursátil y, a su vez, presenta un estado del arte que pone en evidencia los avances y los vacíos relevantes para esta investigación.

2 Línea teórica

Con el objetivo de explorar las condiciones de mercado en las que se pueden extraer retornos, diferentes autores han profundizado en el supuesto no-arbitraje en contextos de asimetrías de información. En esta línea, se encuentra Grossman (1977) que logró incorporar en el modelo clásico del productor las asimetrías de información, con la inclusión de empresas informadas y desinformadas. Grossman (1977) concluyó que la empresa que posee más información acerca de los futuros estados del mundo puede obtener un mayor rendimiento y, además, que el precio spot de un activo no revela toda la información de las empresas informadas. Es relevante mencionar que el precio de mercado bajo este modelo se encuentra determinado por la información de los operadores informados y por múltiples factores no determinados (“*ruido*”) por el mercado.

Posteriormente, trabajos como el de Grossman & Stiglitz (1980) argumentaron que el supuesto de que los mercados están siempre en equilibrio y perfectamente arbitrados son inconsistentes cuando el arbitraje es costoso. Éste trabajo defiende que si bien los precios desempeñan una función relevante en el transporte de información entre los agentes informados y los desinformados, ésta es imperfecta y asimétrica. Los autores concluyeron que gracias a esta asimetría de información se generan pequeños retornos en compensación a los inversionistas que gastan tiempo y dinero en monitorear continuamente la amplia variedad de fuentes de información.

Ahora bien, el trabajo de Grossman (1977) provocó que diferentes autores se cuestionaran sobre los costos de información y la posibilidad de extraer retornos en mercados financieros. De esta forma, Kyle (1985) analizó la velocidad a la que se incorpora la información en los precios y, por medio de un movimiento browniano, modeló agentes que generan ruido en el mercado, argumentando que estos proporcionan un camuflaje que permite obtener ganancias a los agentes informados. Este trabajo concluyó que el ruido aumenta proporcionalmente a la profundidad del mercado y que, a su vez, se incrementan

2. REVISIÓN DE LITERATURA

proporcionalmente los beneficios de la información. En la misma línea, se encuentra De Long et al. (1990) que, haciendo uso de un modelo de generaciones traslapadas en un mercado de activos, encontró que los agentes que generan ruido en el mercado son guiados por creencias estocásticas que afectan los precios y pueden obtener retornos mayores a los esperados. De Long et al. (1990) argumentó que la aleatoriedad de las creencias que generan ruido crea un riesgo en el precio del activo que disuade a los agentes racionales de apostar enérgicamente en su contra. Es importante resaltar que esta generación de trabajos se consolidaron en un contexto en que la tecnología y el acceso a la información era escasa y costosa.

Como se puso en evidencia en la literatura expuesta, existe un margen de ganancia creado por las asimetrías de información que se puede extraer por medio de algoritmos que procesan información y examinan señales de mercado a altas velocidades. Además, estos trabajos resaltan el rol de los agentes ruidosos y estocásticos que debe tenerse en cuenta en el entendimiento del sentimiento en el mercado. Si bien Kyle (1985) introdujo la distinción entre agentes ruidosos e informados, en este trabajo de grado se profundizó en este punto y se planteó una metodología que permitió la inclusión de inversionistas heterogéneos a partir de una agrupación no supervisada de los usuarios de StockTwits. Este punto se profundizará en la metodología.

2 Análisis de sentimiento de los inversionistas

En esta sección se encuentran investigaciones que evalúan las tesis propuestas por los autores de la línea teórica. Una primera aproximación es la de Brown & Cliff (2005) quienes, utilizando encuestas sobre el sentimiento de los inversionistas, encontraron una relación significativa entre los niveles de sentimiento y las desviaciones del precio de Dow Jones, concluyendo que en el corto plazo el mercado tiende a estar sobrevalorado durante períodos de optimismo.

Posteriormente se desarrollaron investigaciones que construyeron proxys para acercarse a los sentimientos de los inversionistas a partir de información en internet. Una de estas aproximaciones consiste en la extracción y análisis de mensajes publicados en páginas como Yahoo o foros de finanzas, en las que se puede medir el sentimiento de los inversionistas. Al hacer uso de esta aproximación surgen nuevos retos para los investigadores, pues se requieren algoritmos que clasifiquen de forma masiva los mensajes y los almacene de manera eficiente. En este sentido, surgen trabajos como los de Sabherwal, Sarkar & Zhang (2011) que a partir de

2. REVISIÓN DE LITERATURA

algoritmos de aprendizaje de máquinas logran clasificar los sentimientos de los inversionistas que publicaban en los foros de Yahoo Finanzas. Los autores concluyeron que, si bien el sentimiento de los mensajes publicados en este sitio no son buenos predictores de los rendimientos del S&P 500, estos sí tienen una correlación alta con el volumen de operaciones y la volatilidad en ventanas de 15 minutos.

Asimismo, en los dos años más recientes se han llevado a cabo investigaciones que incorporaron diferentes metodologías de clasificación y se concentraron en otros indicadores de mercado, hallando altos niveles de correlación entre el sentimiento de los inversionistas y los retornos diarios. Entre estos estudios se encuentra el desarrollado por Kordonis, Symeonidis & Arampatzis (2016), quienes estimaron errores de predicción por debajo del 10% en el precio de cierre de 16 acciones de compañías tecnológicas. Para esto emplearon una Máquina de Soporte Vectorial que tomó como variables independientes el volumen diario de las acciones y un indicador de sentimiento. En esta misma línea de investigación, se encuentra la investigación de Pagolu, Reddy, Panda & Majhi (2016) quienes emplearon una representación textual vectorial (Word2vec) para encontrar que existe una fuerte correlación entre el aumento y la caída de los precios de las acciones de Microsoft con los sentimientos de las personas en tweets.

Por otra parte, Renault (2017) evaluó la tesis propuesta por De Long et al. (1990) a partir de tweets extraídos de StockTwits, concluyendo que el primer cambio de media hora en los sentimientos del inversionista predice el retorno de la última media hora del índice S&P 500. Particularmente, esta investigación abordó los retornos diarios del S&P 500 en ventanas de 30 minutos, para lograr una mayor exactitud a la hora de pronosticar los retornos futuros (Heston, Korajczyk & Sadka, 2010).

En este punto es relevante mencionar que en la literatura del análisis de sentimiento existen dos metodologías ampliamente utilizadas para la clasificación de sentimiento. La primera es la bolsa de palabras, un método que representa un documento ignorando el orden de las palabras y, a partir de diccionarios revisados, asignan un sentimiento por medio de una medida de polaridad que está en función del número de documentos que contienen el término analizado. La segunda metodología utiliza un modelo de clasificación que utiliza una muestra previamente etiquetada para entrenar el modelo que luego clasificará los demás mensajes que no tienen una etiqueta. Finalmente, las ventajas y desventajas de las dos metodologías se abordarán en la sección 4 de este documento.

2 Análisis fundamental

Con el objetivo de identificar cambios bruscos en los mercados financieros y variables macroeconómicas surgen investigaciones que se centran en analizar noticias y variables fundamentales. En este contexto se encuentra la investigación de Baker, Bloom & Davis (2016) en la cual construyen una medida de incertidumbre política y económica de la economía de EE.UU. Esta medida se construyó a partir del análisis de texto a 10 periódicos líderes de EE.UU, el código fiscal federal y el grado de desacuerdo entre los analistas económicos sobre las futuras adquisiciones gubernamentales.

Los autores hallaron una alta correlación de esta medida con el índice VIX y con la frecuencia de ciertas palabras de crisis en el libro Beige. Los autores concluyeron que la medida de incertidumbre política y económica aumenta significativamente cuando se está en un periodo de elecciones presidenciales o en el inicio de un conflicto bélico. Por otro lado, se encuentra la investigación de Li et al. (2014) quienes encontraron una correlación positiva en rendimiento del precio de las acciones del Hong Kong Exchange (HKE), a partir de un análisis de sentimiento de artículos financieros. Así mismo, los autores determinaron que al descomponer el HKE en sectores se incrementa la exactitud de la predicción. Además, los autores probaron que utilizar modelos de polaridad de sentimiento no proporciona predicciones útiles.

Ahora bien, en esta línea de trabajo también se han explorado otras metodologías que se fundamentan en el volumen de búsquedas generadas en Google. Particularmente, Da, Engelberg & Gao (2014), a partir del volumen de búsqueda de unas palabras específicas, construyeron un índice de actitudes financieras y económicas reveladas por la búsqueda. El índice desarrollado por los autores predice los retornos agregados del S&P 500 a un horizonte de un día, encontrando que, si bien el índice está correlacionado con los retornos diarios, el modelo tiene una mayor exactitud en los rendimientos del siguiente día.

Finalmente, los hallazgos presentados indican que el análisis de noticias y variables fundamentales han sido una herramienta metodológica utilizada en diferentes contextos de investigación. No obstante, esta línea pone en evidencia la desconexión que existe entre el análisis de sentimiento de los inversionistas y el análisis que involucra noticias. Es por lo anterior, que este trabajo de grado integró estos dos análisis, para así identificar un canal que articule los sentimientos de los inversionistas con el contenido que están publicando en

portales de noticias.

3 Metodología

Con el objetivo de responder cuál es el grado de predictibilidad de los sentimientos de los inversionistas publicados en StockTwits sobre el comportamiento de los retornos del S&P 500, se construyó una metodología que, a partir de mensajes etiquetados por los usuarios de StockTwits, logró estimar una medida representativa de sentimiento de los inversionistas y de noticias. Esta medida se utilizó en un modelo de pronóstico de la dirección de los retornos de S&P 500. Adicionalmente, con esta metodología se determinó si los sentimientos de los inversionistas son predictores de la dirección de los retornos del S&P 500 en el corto plazo, por medio de un canal constituido a partir de noticias que crean conjeturas en los inversionistas respecto al desempeño futuro del índice. La metodología se resume en la Figura 1 que presenta la relación entre el análisis de sentimiento de los inversionistas, el análisis de sentimiento de noticias y el modelo de pronóstico de la dirección de los retornos del S&P 500.

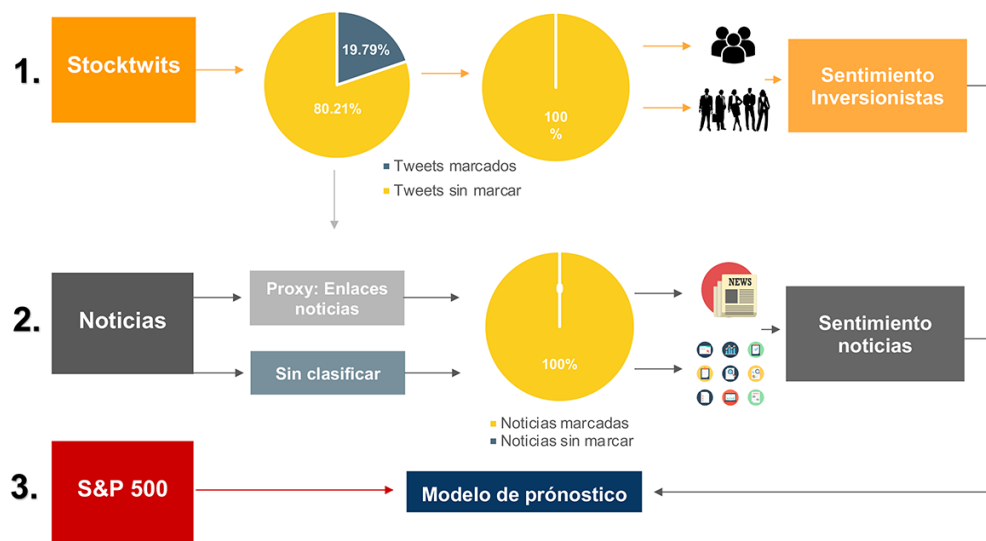


Figura 1: Esquema - Metodología.

Esta sección se encuentra dividida en cuatro partes. La primera sección profundiza en la definición de la dirección de los retornos del S&P 500 como variable dependiente. La segunda y la tercera sección hacen referencia a la estructura del análisis de sentimiento de los inversionistas y noticias respectivamente. En la cuarta sección se presenta el modelo de pronóstico que evaluó si la inclusión del análisis de sentimiento de los inversionistas y noticias son predictores de la dirección de los retornos del S&P 500.

3 Dirección de los retornos del S&P 500

La variable dependiente de esta investigación fue la dirección de los retornos del S&P 500 organizada en ventanas de 30 minutos. En primer lugar, se tomó el precio de cierre del S&P 500 minuto a minuto y, en concordancia con Renault (2017), se dividió el periodo de negociación (10:00 a.m. a 4:00 p.m) en 13 ventanas de media hora³ (Ver Tabla 1). Para cada ventana de tiempo se calcularon los retornos sobre el precio de cierre en el inicio y el final de cada ventana, luego con el objetivo de capturar la dirección se discretizaron los retornos de la siguiente manera:

Tiempo (<i>i</i>)	Inicio	Fin
1	16:00	10:00
2	10:00	10:30
3	10:30	11:00
⋮	⋮	⋮
11	14:30	15:00
12	15:00	15:30
13	15:30	16:00

Tabla 1: Intervalos ventanas de 30 minutos: S&P 500.

Por notación se definió $r_{i,t}$ como el retorno del S&P 500 del i -ésimo intervalo del día t . En este punto es importante aclarar que el primer intervalo se construyó como el retorno de media hora usando el precio de cierre en el día $t - 1$ y el precio a las 10:00 a.m. del día t . El retorno del último intervalo ($r_{13,t}$) se calculó a partir del signo del retorno de media hora entre las 3:30 p.m. y las 4:00 p.m. del día t . Si bien, el mercado inicia operaciones a las 9:30 a.m., con el objetivo de excluir las transacciones fuera de horario y los movimientos de precios de apertura y cierre durante la noche, la primera ventana abarca un horizonte más amplio. La metodología de intervalos de media hora ha sido abordada por diferentes autores como Heston et al. (2010) y Gao, Han, Li & Zhou (2015) que argumentan que existen patrones en los retornos intradía que perduran a lo largo del tiempo.

En una primera aproximación a la predicción del S&P 500 se hicieron pronósticos⁴ a los retornos intradía de cada una de las ventanas. Sin embargo, se descartó esta aproximación luego de encontrar que la inclusión del análisis de sentimiento de los inversionistas y noticias como variables independientes no mejoraba la predicción bajo el criterio MAPE, aún cuando fueron significativas a un nivel del 10% en algunas ventanas.

Con la intención de abordar este problema desde otra perspectiva que permitiera

³En la construcción de las ventanas solo se tuvo en cuenta los días hábiles del mercado de EE.UU.

⁴Se ensayaron modelos ARMA y GARCH que utilizaron como variables independientes los resultados del modelo de análisis de sentimiento de los inversionistas y las noticias para realizar un pronóstico.

3. METODOLOGÍA

identificar el efecto de los sentimientos de los inversionistas y de las noticias, se replanteó la variable dependiente en función de la dirección de los retornos. En consecuencia, la variable dependiente se definió en tres categorías que dependen del signo de los retornos (Ecuación 1):

$$d_{i,t} = \begin{cases} 1 & \text{si } r_{i,t} > 0 \\ 0 & \text{si } r_{i,t} = 0 \\ -1 & \text{si } r_{i,t} < 0 \end{cases} \quad (1)$$

Poner en un pie de página o en algún lado eso del 0 que siempre le he preguntado. Por qué no usar un epsilon? Otros autores lo hacen así?

La definición de la variable dependiente en tres categorías⁵partió del supuesto de que el mercado percibe una dirección alcista si los retornos son mayores a cero, y dirección bajista si los retornos son menores a cero. Este supuesto no es fuerte en la medida en que cualquier retorno mayor a cero o menor a cero con un nocional lo suficientemente grande puede generar beneficios o pérdidas significativas para un inversionista.

3 Análisis de sentimiento a inversionistas

Con el objetivo de construir una medida que estime el sentimiento que tienen los inversionistas sobre la dirección de los retornos del S&P 500 y, siguiendo a Renault (2017), se optó por utilizar los tweets financieros que estuvieron relacionados con el índice durante el 2010 y el 2018, a partir de la información disponible en el sitio de StockTwits. StockTwits es una plataforma que facilita la interacción entre diferentes agentes del mercado financiero en un muro que permite hacer un seguimiento al comportamiento de un índice o acción en tiempo real, y a la opinión que tienen los usuarios sobre este. En este sentido, los usuarios de StockTwits pueden escribir un mensajes de opinión en el muro de una determinada acción o índice y clasificarlo con un sentimiento alcista o bajista. En el muro de cada índice se expone información como el precio de cierre, apertura, máximo, mínimo, volumen, entre otros datos. Este sitio cuenta con una alta actividad que involucra diferentes actores del mercado, pues para

⁵Utilizando la distribución de los retornos del S&P 500, se planteó una grilla con diferentes percentiles que cortaban la muestra en tres categorías (alcista, neutral, bajista). Luego se evaluó un modelo logit en cada una de las diferentes sub-muestras, y se calculó la métrica F1 ponderado. Debido a la concentración de la distribución de los retornos del S&P 500, los resultados para umbrales extremos muy altos (P10, P90) generaban una muestra desbalanceada en donde el desempeño del clasificador era alto por la concentración de datos en una sola categoría. Por otra parte, se probó con umbrales bajos (P49,P51) y debido a la distribución de los retornos del índice, este valor era muy cercano a cero (0.0001). Dado lo anterior, y con el objetivo de capturar la dirección de los retornos del S&P 500 se optó por utilizar como umbral el cero.

3. METODOLOGÍA

el 2016 alcanzó 1.500.000 usuarios activos mensuales.

En este sentido, debido a que los tweets financieros de esta plataforma vienen acompañados de un sentimiento alcista o bajista que representa la opinión real de los inversionista respecto al comportamiento del S&P 500, se utilizó como una proxy para estimar el sentimiento del mercado. Es relevante mencionar que no todos los mensajes de esta plataforma cuentan con una etiqueta de sentimiento, debido a que algunos usuarios no clasificaron el mensaje con un sentimiento definido. Teniendo en cuenta lo anterior se construyó un modelo de análisis de sentimiento de los inversionistas que permitiera estimar la probabilidad de que el sentimiento de un tweet fuera positivo. El objetivo de este modelo fue extender la medida de sentimiento a todos los mensajes sin etiquetar de la plataforma, para así estimar una medida de sentimiento representativa del mercado.

En este orden de ideas, el modelo de análisis de sentimiento de los inversionistas es un modelo supervisado que aprovecha los mensajes etiquetados por los usuarios con un sentimiento alcista y bajista, para clasificar el sentimiento de los mensajes no etiquetados. Este modelo utilizó como variable dependiente una variable dicótoma que tomó el valor de 1 si el tweet es alcista y 0 si es bajista⁶ y como variables independientes los pesos asignados por la matriz término-documento (DTM por sus siglas en inglés) a los tweets preprocesados⁷ y clasificados por los usuarios. Es importante resaltar que la DTM se construyó con la frecuencia de término – frecuencia inversa de documento (TF-IDF por sus siglas en inglés), una medida que expresa la relevancia de una palabra de un documento inmerso en un grupo de varios documentos. La intuición de esta aproximación consiste en utilizar los mensajes previamente marcados por los usuarios, para identificar las palabras más frecuentes relativas a cada documento cuando se asigna un sentimiento. Los pesos de estas palabras luego son utilizados para clasificar los mensajes sin etiqueta.

Ahora bien, la forma funcional entre la variable dependiente y las variables independientes se estimó por medio de diferentes modelos no lineales: Logit, Random Forest (RF), Máquina de Soporte Vectorial (SVM) y Red Neuronal Convolutiva⁸(CNN por sus

⁶En la plataforma de StockTwits solo existe la posibilidad de marcar el mensaje con etiqueta alcista o bajista.

⁷El preprocesamiento es una técnica que permite la normalización y limpieza de los documentos. Esta técnica permite reducir la longitud del documento eliminando las palabras que no aportan al sentimiento de un texto como artículos, pronombres, preposiciones, entre otras.

⁸La red CNN identifica y reduce diferentes patrones de sentimiento embebido en los documentos, por medio de diversos filtros que recorren la DTM. Cada filtro que recorre la matriz recibe el nombre de convolución y tiene como objetivo capturar patrones específicos en los datos que permitan mejorar la precisión de la clasificación. La información de cada convolución es agregada con una función para realizar un mapeo causal no lineal entre el contenido del documento y el sentimiento asignado. En el caso particular del análisis de sentimiento de los inversionistas, se aplican 3 filtros diferentes: bigramas, trigramas y

3. METODOLOGÍA

siglas en inglés). Si bien estas metodologías se diferencian en la forma de estimación, todas son modelos de clasificación binaria que asignan una probabilidad de que un tweet sea clasificado como alcista o bajista. En este sentido, si un determinado tweet fue clasificado con una probabilidad cercana a 1 significa que es alcista, y si es cercana a 0 es bajista.

Si bien se puede definir un umbral de probabilidad para clasificar los mensajes de manera binaria, en esta investigación se tomó directamente la probabilidad de que un determinado tweet sea clasificado como alcista o bajista. Lo anterior con el objetivo de poder crear una medida de sentimiento agregada en ventanas de 30 minutos. Se debe resaltar que aunque no exista una categoría de clasificación neutral, el modelo asigna una probabilidad de 0,5 a los tweets que no contienen un conjunto de palabras que los diferencie entre alcista y bajista.

Por otra parte, con el objetivo de evaluar el desempeño de los diferentes modelos de clasificación binaria, se utilizó como métrica el área debajo de la curva ROC⁹ que se interpreta como la probabilidad de que ante un par de tweets, uno alcista y el otro bajista, la prueba los clasifique correctamente. El criterio AUC puede tomar valores entre 0 y 1, un valor cercano a 1 indica que el modelo de clasificación tiene un alto nivel de predictibilidad.

El segundo criterio utilizado fue la exactitud¹⁰, medida que indica el número de clasificaciones correctas dentro de la muestra. Sin embargo, es importante resaltar que debido a que los modelos anteriores arrojan una probabilidad de clasificación, ésta se discretizó utilizando un umbral genérico de 0,5 que permitió evaluar los verdaderos positivos y verdaderos negativos. Este umbral no se optimizó, debido a que se tomaron las probabilidades que arroja el modelo de análisis de sentimiento como insumo para el modelo de pronóstico de forma directa. Es por lo anterior, que si bien la exactitud da luz de los resultados del modelo, este criterio solo se utilizó como una medida informativa. En este sentido, esta investigación de grado evaluó todos los modelos de clasificación binaria bajo el criterio AUC, debido a su fácil interpretación, poder estadístico y amplio uso en el campo de Machine Learning.

Por otra parte, debido a que los modelos de clasificación elegidos exigieron la calibración de diferentes parámetros, se planteó una optimización que buscó maximizar el AUC haciendo uso de una grilla de parámetros que variaron en cada iteración. Con la finalidad de reducir la

cuatrigramas. Más información en Wu (2017).

⁹El área debajo la curva ROC (AUC) se interpreta como la probabilidad de que ante un par de tweets, uno alcista y el otro bajista, la prueba los clasifique correctamente. Para profundizar en este criterio consultar Fawcett (2006).

¹⁰La exactitud se construye a partir de los verdaderos positivos y verdaderos negativos (diagonal matriz de confusión) sobre el total de la muestra de prueba. Para profundizar en este criterio consultar Fawcett (2006).

3. METODOLOGÍA

varianza y garantizar que los estimadores del Logit, el RF y la SVM fueran independiente de la partición entre datos de entrenamiento y prueba se utilizó la metodología Bagging¹¹.

En este punto es importante resaltar que esta metodología de clasificación, a diferencia de aquellas que utilizan diccionarios¹², no requiere de un vocabulario construido de manera manual que podría estar sesgado dependiendo del objetivo con el que se creó y el momento en que éstas palabras fueron seleccionadas. En otras palabras, debido a que StockTwits brinda tweets marcados por los propios usuarios, se puede construir una muestra de entrenamiento objetiva en la que no existe manipulación externa, pues no existe un tercero que realice juicios de valor acerca del sentimiento que quiso plasmar el usuario. Adicionalmente, con la implementación de esta metodología se garantizó que el modelo de sentimiento incorporara nuevas palabras que no se encuentran definidas en los diccionarios, y cuyo significado puede variar a lo largo del tiempo.

Luego de utilizar el modelo de análisis de sentimiento de los inversionistas para clasificar todos los tweets que hablan acerca del S&P 500, se construyeron ventanas de 30 de minutos que capturaron el sentimiento agregado de los sentimientos de los inversionistas. Se debe resaltar que la agregación del sentimiento se puede realizar de diferentes maneras en las que se asignen pesos diferentes a los sentimientos de los usuarios. Es por lo anterior, y siguiendo la tesis de Kyle (1985), que es necesario diferenciar los agentes legítimos y agentes ruidosos que participan en el mercado. En este sentido, esta investigación tomó en cuenta dos aproximaciones: usuarios homogéneos y usuarios heterogéneos que asignan pesos diferentes a los sentimientos de los usuarios a la hora de agregar los sentimientos en las diferentes ventanas de tiempo.

3.2.1 Análisis de usuarios homogéneos

La aproximación de usuarios homogéneos utiliza un promedio simple para agregar el sentimiento de los inversionistas, asignándoles así el mismo peso a todos los sentimientos de los usuarios que publicaron en una ventana en particular. En otras palabras, si en una ventana cientos de usuarios publicaron mensajes, la agregación de sentimiento para esta ventana será el promedio simple de la probabilidad de sentimiento de todos los mensajes de esa ventana. A

¹¹El Bagging es una metodología que promedia N estimadores en N sub-muestras de la muestra de entrenamiento y promedia estos estimadores con pesos iguales. El número de estimadores N se seleccionó a partir de una optimización de grilla bajo el criterio AUC. Para más información consultar (James, Witten, Hastie & Tibshirani, 2013).

¹²La clasificación con diccionarios construye un vector a partir de la frecuencia de palabras en un documento determinado. Posteriormente, este vector es contrastado con un diccionario construido por un grupo de expertos para obtener una aproximación al sentimiento del documento.

3. METODOLOGÍA

modo de ejemplo, si en la ventana 2 se publicaron 2 tweets con probabilidades de sentimiento de 0,99 (alcista) y 0,11 (bajista), el sentimiento de esta ventana es de 0,55 (neutral). La intuición de esta aproximación es que todos los inversionistas tienen la misma información, y capacidad para predecir la dirección de los retornos del S&P 500.

3.2.2 Análisis de usuarios heterogéneos

La aproximación de usuarios heterogéneos utiliza un promedio ponderado que asigna diferentes pesos a los sentimientos de los inversionistas que publicaron en una determinada ventana. Con el propósito de establecer unos ponderadores objetivos, se realizó un modelo de agrupación no supervisado¹³, que clasificó a los usuarios según cuatro variables que están en función del número de aciertos históricos según el sentimiento de sus publicaciones. En otras palabras, se evaluó si cuando un inversionista publica un mensaje con un sentimiento determinado realmente la dirección de los retornos es la esperada por el inversionista.

Para lograr lo anterior se construyó la historia de las publicaciones de cada usuario y se discretizó el sentimiento de este para compararlo con el historial de la dirección de los retornos del S&P 500. Debido a que StockTwits no permite sentimientos neutrales, se tomó únicamente la dirección positiva y negativa de los retornos del índice. El resultado de este ejercicio es en una distribución Bernoulli del número de aciertos y desaciertos de cada usuario a lo largo de su historia en StockTwits. Con la distribución de cada usuario, se calculó la desviación, la entropía y la probabilidad de acertar, para utilizarlas como variables diferenciadoras en un modelo de k -medias¹⁴ que determinó los k -grupos de usuarios. El número de grupos se seleccionó a partir del criterio de la silueta que se explica más adelante.

Posterior a la definición de los grupos se ordenaron y ponderaron a partir del número de aciertos promedio en cada grupo. El ponderador de sentimiento de cada grupo se definió como la diferencia entre la probabilidad del sentimiento positivo (1) menos la neutral (0,5) sobre el número de grupos de usuarios. A modo de ejemplo, si se seleccionaron 5 grupos, el primero tiene un ponderador de 0,5, el segundo de 0,625 y así sucesivamente hasta llegar al último grupo que tiene un ponderador 1. La intuición de esta aproximación consistente en asignar un mayor peso al sentimiento de los inversionistas que han acertado históricamente al

¹³El aprendizaje no supervisado no requiere de un conocimiento a priori acerca de la muestra. Estos modelos infieren patrones a partir de los datos entrada.

¹⁴El método de agrupamiento de K-medias busca a partir unas características dividir un conjunto de datos en K grupos diferentes que no se traslapan. La intuición de este método de agrupamiento es que una buena manera de segmentar los datos es haciendo que la varianza dentro de cada grupo sea tan pequeña como sea posible. Para más información consultar Morissette & Chartier (2013).

3. METODOLOGÍA

mercado, y un menor peso a los inversionistas que no han pronosticado correctamente la dirección. Finalmente, se utilizaron los grupos para ponderar el sentimiento de los inversionistas.

En este sentido, lo que se buscó con estas dos aproximaciones fue agregar y normalizar el sentimiento percibido por cientos de usuarios únicos durante cada ventana de tiempo. El resultado final de cada una de las aproximaciones es una matriz de $N \times 13$ que contiene el sentimiento agregado de cada ventana. En la sección de resultados se presentan los hallazgos de la aproximación de usuarios homogéneos. En la sección de análisis complementarios se modificó el supuesto de usuarios homogéneos y se presentan los hallazgos del análisis de usuarios heterogéneos.

3 Análisis de sentimiento a noticias

Con el objetivo de caracterizar el canal constituido a partir de noticias que crean opiniones en los inversionistas respecto al desempeño futuro del índice, se estructuró un modelo de clasificación de sentimiento de noticias que capturó las opiniones de los inversionistas de StockTwits sobre las noticias publicadas en diferentes portales. La construcción de este modelo fue tratado de manera similar al que se utilizó en el sentimiento de los inversionistas, diferenciándose en la construcción de la muestra de entrenamiento.

En este sentido, debido a que no existe una base de datos que contenga noticias financieras y un sentimiento asignado por los usuarios, ésta se construyó haciendo uso de los tweets publicados por los inversionistas de StockTwits. Se tomaron todos los tweets que hablan acerca del S&P 500 y de las 25 acciones más importantes del índice, y se seleccionaron los mensajes que contaron con un enlace y fueron marcados con una etiqueta de sentimiento. Posteriormente, se extrajo la raíz del enlace de cada mensaje dentro de esta muestra y se construyó una tabla de frecuencias sobre los enlaces. Con esta tabla se logró identificar cuáles fueron los sitios web más referenciados por los usuarios cuando expresaron una opinión acerca del S&P 500. Finalmente, con la información de la tabla de frecuencias se filtraron los enlaces de portales de noticias que tenían una frecuencia mayor a 10, obteniendo una lista de los portales más utilizados por los inversionistas de StockTwits.

Luego de identificar las principales fuentes de noticias, se procedió a buscar los enlaces de estas fuentes dentro de los mensajes clasificados por los inversionistas. Una vez obtenida

3. METODOLOGÍA

esta información se realizó webscraping¹⁵ para la extracción de la noticia completa, la fecha de publicación y el título. Con este procedimiento se consolidó una base con las noticias marcadas como alcistas o bajistas por los usuarios.

En este sentido, utilizando la base de datos construida se entrenó un modelo de clasificación que permitió estimar el sentimiento de las noticias no clasificadas provenientes de diferentes agencias de información. Esta estimación se realizó haciendo uso de un modelo supervisado cuya variable dependiente es una variable dicótoma que toma el valor de 1 (si la noticia es positiva) y 0 (si es negativa), y las variables dependientes fueron los pesos que asigna la matriz DTM a cada noticia. En este caso, al igual que el modelo de sentimiento de los inversionista, se estimó la forma funcional por medio Logit, una SVM, un RF y una red CNN. El criterio de decisión y la calibración de los diferentes parámetros de los modelos se realizó de la misma manera que el modelo de análisis de sentimiento de los inversionistas. La intuición del análisis de sentimiento de noticias radicó en utilizar como proxy los tweets clasificados por los usuarios para clasificar las noticias, y así poder identificar las palabras más frecuentes relativas a cada noticia cuando los usuarios marcaron un sentimiento alcista o bajista en su mensaje.

Luego de entrenar el modelo, éste se utilizó para clasificar las noticias de los principales sitios de noticias. Para esto se construyó una base de datos con todo tipo de noticias de las principales agencias citadas por los inversionistas de StockTwits. En este caso, al igual que en el análisis de sentimiento, una noticia con una probabilidad cercana a 1 fue clasificada con un sentimiento alcista y 0 con un sentimiento bajista. Así mismo, el modelo asignó una probabilidad de 0,5 a las noticias que no contaron con un conjunto de palabras que las diferenciara entre positiva y negativa. Con la estimación del sentimiento de todas las noticias se construyó un índice de sentimiento para cada una de las ventanas que se calculó a partir de la media aritmética de la probabilidad.

En este punto es importante mencionar que para caracterizar el canal entre las noticias financieras y el sentimiento de los inversionistas, se supuso que el sentimiento de la noticia es el mismo otorgado por el inversionista en su tweet. Este supuesto es plausible en la medida en que el inversionista al publicar un mensaje que contiene una noticia y un sentimiento es porque primero leyó la noticia y a partir de esta consolidó una opinión acerca del mercado. Adicionalmente, se partió del supuesto de que cualquier noticia, sin importar si es financiera o

¹⁵Técnica de extracción de información masiva a partir de sitios web.

3. METODOLOGÍA

no, puede afectar al S&P 500. Este supuesto no es fuerte en la medida en que el S&P 500 está compuesto por 500 empresas de diferentes sectores que son propensas a cambios en la coyuntura económica, política y social de EE.UU y de diferentes países líderes.

Por otro lado, con el objetivo de flexibilizar el supuesto anterior, en la sección de análisis complementarios se evaluó el efecto del sentimiento de las noticias sobre la dirección de los retornos del S&P500 utilizando una agrupación de los temas de las noticias. Para esto se realizó un modelo de clasificación de tópicos latentes¹⁶ (LDA, por sus siglas en inglés -Latent Dirichlet Allocation-) no clasificado que permitió identificar los temas de las noticias en una muestra de entrenamiento, para luego clasificar la muestra de prueba. Posteriormente, se construyeron ventanas de 30 minutos con las noticias de cada tópico, y se evaluó su significancia sobre la dirección de los retornos. La intuición de la agrupación de noticias fue poder evaluar la existencia de grupos de tópicos que mejoran la predicción de los retornos del S&P 500. A modo de ejemplo, se pensaría que el efecto del sentimiento de las noticias hablan sobre las publicaciones de resultados de las diferentes compañías de EE.UU afectan en mayor medida el S&P 500, que las noticias de farándula.

3 Modelo

Con el objetivo de comprobar si el sentimiento de los mensajes de los inversionistas de StockTwits son predictores de la dirección de los retornos del S&P 500, por medio de un canal constituido a partir de noticias, se estimaron 3 especificaciones del modelo de pronóstico que incorpora de manera individual y conjunta el sentimiento de los inversionistas y de las noticias.

La primera y segunda especificación del modelo de pronóstico evaluaron el efecto individual del canal de sentimiento de los inversionistas y de noticias sobre la dirección de los retornos del S&P 500, estas dos especificaciones se escribieron de la siguiente manera (Ver Ecuación 2):

$$d_{i,t} = f(stweets_{i-1,t}) + \epsilon_{i,t} \quad d_{i,t} = f(snews_{i-1,t}) + \epsilon_{i,t} \quad (2)$$

La tercera especificación evaluó la inclusión conjunta de los modelos de sentimiento, tomando como variable dependiente la dirección de los retornos del S&P y como variables independientes los rezagos de la probabilidad de sentimiento de los inversionistas y de las noticias (Ver Ecuación

¹⁶El modelo LDA es un modelo no supervisado que asigna tópicos a una muestra de documentos a partir del cálculo probabilístico de una estructura de temas latentes que con una probabilidad p , generan la colección de documentos observados en la muestra de entrenamiento. Más información consultar (Blei, Ng & Jordan, 2003).

3. METODOLOGÍA

3).

$$d_{i,t} = f(stweets_{i-1,t}, snews_{i-1,t}, stweets_{i-1,t} * snews_{i-1,t}) + \epsilon_{i,t}, \quad (3)$$

En los tres modelos la variable dependiente que captura la dirección de los retornos del S&P 500 es $d_{i,t}$. Las variables independientes que representan la probabilidad agregada de que el sentimiento de los tweets y noticias publicadas en la ventana $i - 1$ del día t sean positivos son $stweets_{i-1,t}$ y $snews_{i-1,t}$ respectivamente. Además, la tercera especificación incluyó un tercer término que captura la relación entre el sentimiento de los tweets y de las noticias, para así caracterizar el canal constituido por noticias. Finalmente, los tres modelos tienen $\epsilon_{i,t}$ que hace referencia a un término de error que captura choques estocásticos en la ventana i del día t .

En este punto, se debe aclarar que los tres modelos utilizaron rezagos de la variable de sentimiento, ya que en la realidad al pronosticar la ventana i, t solo se tiene información disponible del sentimiento de los inversionistas y de las noticias de la ventana anterior $i - 1, t$. A modo de ejemplo, si el 20 de febrero del 2018 a las 10:30 a.m. (ventana 2) un inversionista quiere pronosticar la dirección de los retornos de las 11:00 a.m (ventana 3), solo observa la información de los tweets y noticias publicadas en la ventana.

Por otra parte, debido a que se desconoce la relación entre las variables dependientes y las independientes de los tres modelos se especificó una forma funcional general. En este sentido, se evaluaron diferentes familias de modelos discretos supervisados para estimar la forma funcional dados unos parámetros. Particularmente se exploró un modelo Logit, una SVM y un RF para aproximarse a la forma funcional de los datos.

Al tratarse de modelos con tres categorías en la variable dependiente, el AUC no se encuentra definido. Es por lo anterior que existen otras métricas para evaluar modelos que involucren más de una clase, entre las que se encuentran medidas como la exactitud, la precisión¹⁷, la exhaustividad¹⁸ y el puntaje F1¹⁹. Si bien la exactitud es una buena aproximación para problemas balanceados, en donde el castigo al equivocarse de clasificación es igual para todas las categorías, el puntaje F1 al ser el promedio armónico simple²⁰ de la precisión y la exhaustividad captura la penalización que hace la precisión y la exhaustividad en

¹⁷Medida que indica el porcentaje de predicciones realizadas en una categoría que son correctas. Se construye a partir de los verdaderos positivos sobre los verdaderos positivos y falsos positivos.

¹⁸Medida de precisión que indica la proporción de positivos reales que se identificaron correctamente. Esta se construye a partir de los verdaderos positivos sobre los verdaderos positivos y falsos negativos.

¹⁹Medida resultante del promedio ponderado entre la precisión y la recuperación. Tiene en cuenta los falsos positivos y los falsos negativos.

²⁰ $F1 = \frac{2 * (precision * exhaustividad)}{(precision + exhaustividad)}$

3. METODOLOGÍA

cuanto a los falsos positivos y falsos negativos.

Si bien estas cuatro medidas son ampliamente utilizadas, en esta investigación se utilizaron dos criterios de evaluación que penalizaron de manera diferente las categorías de la variable dependiente. La primera es la exactitud que asigna pesos iguales a los movimientos alcistas, neutrales y bajistas. La segunda es el puntaje F1 ponderado que estima este puntaje para cada una de las categorías de la variable dependiente y luego calcula el promedio ponderado por número de instancias verdaderas por cada categoría. El F1 ponderado al ser una función de la precisión y la exhaustividad de cada una de las categorías asigna pesos diferentes a las equivocaciones en las diferentes categorías.

Por otra parte, debido a que el puntaje F1 ponderado está en función de las instancias verdaderas de cada una de las categorías, para lograr interpretar el F1 de manera correcta se estimó un puntaje mínimo de esta métrica. Este puntaje se calculó a partir del promedio de 10.000 simulaciones de clasificadores estocásticos que respetan el total de observaciones de la muestra de prueba y la distribución histórica de las tres categorías. La intuición detrás de este ejercicio es sentar un valor mínimo del F1 ponderado, con el que se pueda comparar un modelo totalmente aleatorio contra los modelos de pronóstico.

Luego de fijar una métrica con la cual se pudiera evaluar los diferentes modelos, se planteó una segunda métrica de comparación para determinar si la inclusión de las variables independientes de sentimiento aumentan la predictibilidad de la dirección de los retornos del S&P 500. En este sentido, la segunda métrica se definió como el desempeño de un modelo discreto cuya única variable independiente es el rezago anterior:

$$d_{i,t} = f(d_{i-1,t}) + \epsilon_{i,t}, \quad (4)$$

Con el modelo autorregresivo de orden 1 se evaluó si con la incorporación de los sentimientos de los inversionistas y de las noticias se superó el nivel de predictibilidad de un modelo que incorpora únicamente la información del estado pasado.

A modo de resumen, esta metodología aborda la relación entre los tweets y noticias que se publican en dos etapas, con el objetivo de separar los efectos de los sentimientos de los inversionistas y el de las noticias. Específicamente, en la primera etapa se construyó una probabilidad de que un tweet sea positivo y negativo, y en la segunda se buscó la relación entre éstas probabilidades y la dirección de los retornos del S&P 500. Se optó por una metodología

4. DATOS

en dos etapas debido a que en la primera etapa se incorporó la noción de usuarios homogéneos y heterogéneos a la hora de agregar el sentimiento en las 13 ventanas media hora.

En este sentido, si bien es plausible realizar un modelo de clasificación en una sola etapa en la que utilice como variable dependiente la dirección de los retornos y variables independientes la matriz TF-IDF de cada una de las noticias y los tweets publicados. Esta no se implementó debido a que no existe una medida clara de agrupación por ventanas, diferenciación de los efectos, incorporación de rezagos, ni el uso de usuarios heterogéneos.

Finalmente se debe resaltar que esta metodología es aplicable para cualquier tipo de acción o índice que disponga de la suficiente información para lograr entrenar los modelos de análisis de sentimiento y análisis de noticias. Este trabajo de grado evaluó esta metodología en el S&P 500, y la corroboró en la acción de Apple. En la sección de resultados y de análisis complementarios se presentan los principales hallazgos.

4 Datos

En primer lugar, para construir los retornos del S&P 500 (SPX) se tomó la serie minuto a minuto desde 01/08/2010 hasta el 01/08/2018 y se tuvieron en cuenta únicamente los días hábiles del mercado. Adicionalmente, la zona horaria de la serie original fue modificada, pasando de GMT-6 a GMT-5 hora de New York. La información del S&P 500 fue adquirida en BackTestmarket, un proveedor de información de mercado para BackTesting.

Posteriormente, con la información del S&P 500 (SPX), se construyeron las 13 ventanas con los retornos del índice. La construcción de cada ventana se realizó a partir del minuto máximo observado en el inicio y el cierre para cada ventana menos 10 minutos. Lo anterior con el objetivo de capturar el minuto más cercano al límite de la ventana, en el caso que no se registraron transacciones en un determinado periodo. Posteriormente, con esta información se procedió a calcular la variable dicótoma que indica la dirección de los retornos del índice.

En segundo lugar, debido a los altos costos computacionales que requiere el almacenamiento y procesamiento de texto, los tweets utilizados en el modelo de sentimiento de los inversionista provienen únicamente del muro del S&P 500 (SPX) y el del ETF (SPY). En total se obtuvieron 3.305.798 tweets entre el 2009-08-17 hasta 2018-08-17, de los cuales el 19,79% estaban clasificados con un sentimiento alcista o bajista. Esta información fue

4. DATOS

recolectada por medio del API de StockTwits²¹. Posterior al almacenamiento, se modificó la zona horaria original de esta información a la zona horaria de los datos del S&P 500 (GTM-5). Finalmente, con esta información se consolidó una base de datos que contiene los siguientes campos para cada mensaje: fecha, usuario, mensaje y sentimiento. En la Tabla 2 se presenta la estructura de la base de datos de los tweets publicados por los usuarios.

Fecha	Usuario	Mensaje	Sentimiento
2018-08-20T03:52:23Z	jimmyroastbeef	\$SPY futures going up bigly boys. ...	Bullish
2018-08-20T03:45	HEROALLMIGHT	\$SPY another week another group of brainless ...	None
2018-08-20T03:44	abacadabra	\$SPY Just let it go right now ...	Bearish
2018-08-20T03:43	Innerbeast	\$SPY So how many accounts did you blow ...	None
2018-08-20T03:38	abacadabra	\$SPY shout out to my fellow peeps with no lives ...	Bearish
2018-08-20T03:38	littlenickysmith	Another big earnings week for retail ...	Bullish

Tabla 2: Estructura de la base de datos inversionistas.

En tercer lugar, para la construcción de la base de entrenamiento del modelo de análisis de sentimiento de noticias se amplió la muestra de tweets extraídos de StockTwits. Esto se realizó debido a que en la muestra de 3.305.798 tweets que hablan acerca del SPX y el SPY solo se lograron recolectar 3.173 noticias clasificadas por los inversionistas. Dado lo anterior, se amplió la muestra de tweets a las primeras 25 acciones con mayor peso en el S&P 500, logrando almacenar un total de 9.182.301 tweets (Ver Anexo A.1). Del total de la muestra se encontró que 1.896.007 de los mensajes hacen referencia a un enlace externo, dentro de los cuales solo se lograron recolectar 6.754 enlaces de noticias clasificadas con un sentimiento alcista o bajista por los usuarios (Ver Tabla 3).

Fuente	Noticias
CNBC	3.149
Bloomberg	384
Business Insider	715
CNN	427
Reuters	523
WSJ	400
New York Times	182
BBC	85
The Street	443
Investopedia	234
The Guardian	59
NBC News	31
Usa Today	80
New York Post	42
Total	6.754

Tabla 3: Noticias de los inversionistas según el portal de noticias.

Luego por medio de webscraping a los enlaces recolectados, se almacenaron las 6.754 noticias, con la fecha de cada noticia²², el título, el contenido, el sentimiento y la fuente. En la Tabla 4 se presenta la estructura de la base de datos con noticias clasificadas por los usuarios.

Por otro lado, para la construcción de la base de datos de noticias sin clasificar se realizó

²¹Para más información consultar Stocktwits for Developers.

²²En este caso, al igual que con los tweets, se modificó la zona horaria original de esta información a la zona horaria de los datos del S&P 500, estandarizando todo en GTM-5.

5. RESULTADOS

Fecha	Título	Noticia	Sentimiento	Fuente
8/23/18 12:59	Trump says the stock market would crash if he were impeached	The stock market has had little reaction so far to Trump's ...	Bearish	CNBC
8/19/18 23:49	Tariffs Will Hurt U.S. More Than Rest of World, Maersk Says	The U.S. economy will be hit many times harder ...	Bullish	Bloomberg
8/19/18 17:07	A Surprising Bulwark for the U.S. Economy: The Personal Saving Rate	On the eve of the last two recessions ...	Bullish	WSJ
8/17/18 18:48	U.S., China Plot Road Map to Resolve Trade Dispute by November	Chinese and U.S. negotiators are mapping ...	Bullish	WSJ
8/16/18 07:51	Emerging market fears hit U.S. stocks and the	Emerging market fears hit U.S. stocks ...	Bearish	Bloomberg

Tabla 4: Estructura de la base de datos noticias.

webscraping a todas las noticias históricas publicadas por boots disponibles en cada uno de los sitios oficiales de los portales seleccionados para construir la base de datos con noticias sin clasificar. En otras palabras, se buscaron los archivos históricos de noticias²³ que envía cada página de noticias a buscadores como Google o Yahoo, y a partir de esta información se construyeron los enlaces de las noticias de cada sitio. Posteriormente, por medio de webscraping se reconstruyó la noticia, el título, la fecha de publicación²⁴. Debido a las restricciones de algunos portales de noticias, no se logró acceder a la información de todos los portales que utilizan los inversionistas de StockTwits.

Fuente	Noticias
Reuters	1.926.119
BBC	717.079
CNBC	535.003
Nypost	416.154
Business insider	218.500
NYT	218.489
Washington	106.439
Investopedia	8.91
Total	4.146.674

Tabla 5: Noticias según el portal de noticias.

Dado lo anterior, se consolidó una base de datos que contiene la fecha de cada noticia²⁵, el título, el contenido, y la fuente de las 4.146.674 noticias que se encuentran sin un sentimiento definido. En la Tabla 5 se presenta a detalle de las noticias que logró almacenar por cada portal.

5 Resultados

En esta sección se presentan los resultados de la metodología descrita. Los modelos que se presentan a continuación fueron calibrados en una submuestra de entrenamiento y se evaluaron en una muestra de evaluación. La muestra de evaluación se utilizó para comparar la categoría asignada por cada uno de los modelos con respecto a la categoría real de los datos. Esta sección se divide en tres partes que abordan de manera independiente los resultados de los modelos propuestos: análisis de sentimiento de los inversionistas, análisis de sentimiento de noticias y pronóstico.

²³Estos archivos se encuentran disponibles en cada página sitio.com/robots.txt, y permiten identificar a los buscadores toda la información histórica disponible del sitio.

²⁴Debido a que la información se extrajo de diferentes portales que utilizan diferentes zonas horarias, se unificaron todas las zonas horarias a GTM-5.

²⁵En este caso, al igual que con los tweets, se modificó la zona horaria original de esta información a la zona horaria de los datos del S&P 500, estandarizando todo en GTM-5.

5 Análisis de sentimiento de los inversionistas

El modelo de análisis de sentimiento de los inversionistas fue calibrado en una base de entrenamiento que correspondió al 95% (621.680 tweets) de la muestra marcada por los usuarios, y evaluado en una base de prueba que correspondió al 5% (32.721 tweets) de la muestra clasificada²⁶. En este sentido, luego de realizar el preprocesamiento a los tweets clasificados y optimizar los parámetros de diferentes modelos, se logró concluir que usando el criterio AUC el mejor modelo es la red CNN. Este resultado también se puede ver en la curva ROC (Ver Figura 2) en la cual la curva de la red CNN se acerca a la curva de 90° que se forma en el punto (0,1) en donde se alcanza la mayor tasa de verdaderos positivos.

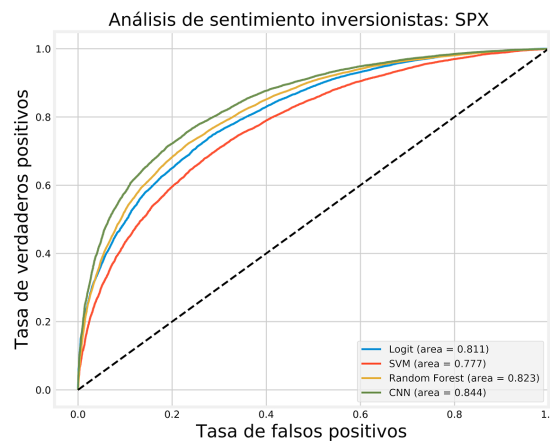


Figura 2: Curva ROC: Fuera de muestra (5%) - Análisis de sentimiento inversionistas.

Siguiendo las métricas de evaluación (Ver Tabla 6) se seleccionó la red CNN, pues superó los demás modelos con un AUC 0,84. En este caso, la brecha entre el AUC de la red CNN con respecto a los otros modelos puede ser explicada por las convoluciones de la red neuronal. Esto se debe a que las convoluciones permiten incorporar diferentes filtros que no solo analizan el sentimiento en bigramas, como los otros modelos, sino que permiten la incorporación del sentimiento en trigramas y cuatrigamas al análisis. En otras palabras, la red CNN agrega más posibilidades semánticas a la hora de analizar un sentimiento.

Modelo	Logit	SVM	Random Forest	CNN
AUC	0,81	0,77	0,82	0,84
Exactitud	73%	70%	74%	76%

Tabla 6: Métricas: Fuera de muestra (5%) - Análisis de sentimiento de los inversionistas.

Por el contrario, los otros modelos por su construcción solo se centran en el sentimiento

²⁶Se evaluó la robustez de los modelos con una muestra de entrenamiento del 70% y una muestra de prueba del 30%. Los resultados no cambiaron significativamente con respecto a una muestra de entrenamiento del 95% y una muestra de prueba del 5%.

5. RESULTADOS

de los inversionistas capturado en bigramas, restringiendo su rendimiento respecto a la CNN. Esta limitación se evidencia en la brecha que tienen los modelos que provienen de aproximaciones teóricas diferentes en comparación de la CNN. Finalmente, el resultado de clasificación es consistente entre todos los modelos y supera un AUC de 0,75, lo que sugiere que esta metodología de clasificación es pertinente para el análisis de sentimiento de los inversionistas.

Posteriormente, utilizando la red CNN, se clasificaron los 2.651.397 tweets no marcados que hablaron sobre el S&P 500 con la probabilidad de ser clasificados como alcista. En el Anexo A.2 se presenta la distribución de la probabilidad de sentimiento de los inversionistas de todos los tweets clasificados. Luego de clasificar todos los tweets, se organizaron en 13 ventanas de tiempo que agregaron el sentimiento de los inversionistas. Esta agregación se realizó desde dos aproximaciones: usuarios homogéneos y usuarios heterogéneos. Los resultados de la agrupación de usuarios heterogéneos y la inclusión de estos en el modelo de pronóstico se presentarán en la sección de análisis complementarios.

5 Análisis de sentimiento noticias

En este caso, al tener 6.754 noticias clasificadas, el modelo de análisis de sentimiento de noticias fue calibrado en una base de entrenamiento que correspondió al 90% (6.054 noticias) de las noticias clasificadas²⁷, y evaluado en una base de prueba que correspondió al 10% (673 noticias) restante. Luego de realizar el procesamiento de las noticias clasificadas y optimizar los parámetros de los modelos, se concluyó que el mejor modelo es la red CNN por el criterio AUC (Ver Figura 3).

²⁷Se evaluó la robustez de los modelos con una muestra de entrenamiento del 70% y una muestra de prueba del 30%. Los resultados no cambiaron significativamente con respecto a una muestra de entrenamiento del 90% y una muestra de prueba del 10%.

5. RESULTADOS

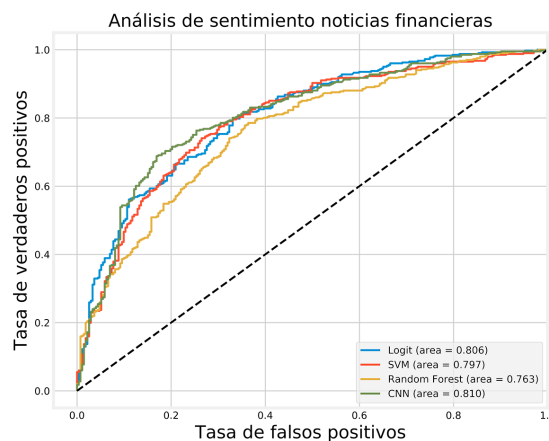


Figura 3: Curva ROC: Fuera de muestra (10%) - Análisis de sentimiento de noticias.

En este caso, si bien la red neuronal alcanza un mayor AUC, la brecha con el Logit y la SVM es menor que en el modelo de análisis de sentimiento de los inversionistas. Lo anterior sugiere que el sentimiento de las noticias se logra capturar directamente en bigramas, a diferencia del sentimiento de los inversionistas. La intuición detrás de este resultado es que al tratarse de noticias redactadas y revisadas por portales de noticias, la relación entre las palabras puede llegar a ser mucho más clara que un tweet escrito por un inversionista. Además, otra de las razones por la cual el desempeño de la red CNN en el corpus de noticias fue inferior al de corpus de tweets es por el tamaño de la muestra. Esto se debe a que el desempeño de una red neuronal está en función del tamaño de la muestra de entrenamiento y el número de veces que se recorre (épocas).

Por otro lado, en la muestra de prueba se evidencia que la mayor exactitud fuera de muestra correspondió a la red CNN con 74%, seguido del Logit y la SVM (Ver Tabla 7). En el análisis de sentimiento de noticias, al igual que el de inversionista, se utilizaron directamente las probabilidades que arrojó el modelo de análisis de sentimiento como insumo para el modelo de pronóstico.

Modelo	Logit	SVM	Random Forest	CNN
AUC	0,80	0,79	0,76	0,81
Exactitud	73%	73%	70%	74%

Tabla 7: Métricas: Fuera de muestra (10%) - Análisis de sentimiento de noticias.

Luego se utilizó la Red CNN para clasificar las 4.146.674 noticias disponibles en los diferentes portales de las agencias informativas seleccionadas con la probabilidad de ser clasificados como alcista. En el Anexo A.3 se presenta la distribución de la probabilidad de

5. RESULTADOS

sentimiento de todas las noticias. Luego de clasificar todas las noticias, se organizaron en 13 ventanas de tiempo que agrupan la probabilidad de sentimiento promedio de todas las noticias durante esa ventana. El resultado de esta etapa se utilizó como variable independiente en el modelo de pronóstico de la dirección de los retornos del S&P 500. En la sección de análisis complementarios se agruparon las noticias por tema y se evaluó el desempeño del modelo de pronóstico controlando por el sentimiento de las noticias de cada uno de los tópicos.

5 Modelo de pronóstico

Con el objetivo de comprobar si el sentimiento de los mensajes de los inversionistas publicados en StockTwits son predictores de los retornos del S&P 500, por medio de un canal constituido a partir de noticias, se corrieron tres modelos de pronóstico. Esta sección está dividida en tres partes, de modo que inicia con la precisión de las métricas que se utilizan para evaluar los modelos pronóstico. Posteriormente, se estima el efecto la inclusión de los modelos de sentimiento los inversionistas y de las noticias de manera independiente. Finalmente, se evalúa de manera conjunta los modelos de sentimiento y se utiliza como control la probabilidad del modelo de análisis de sentimiento y la probabilidad del modelo de noticias.

5.3.1 Métrica de evaluación

Debido a que la variable dependiente de la dirección de los retornos del S&P 500 tiene tres categorías (alcista, neutral y bajista) la métrica de evaluación que se utilizó fue el puntaje F1 ponderado. Puesto que este puntaje se encuentra en función de las proporciones de cada categoría en la muestra de prueba, se simuló 10.000 clasificadores estocásticos que respetan las proporciones históricas de la muestra de entrenamiento. Posteriormente, se calculó el F1-ponderado de los 10.000 clasificadores estocásticos en una muestra de prueba, y se agregó el desempeño de estos clasificadores utilizando el promedio simple.

Es importante resaltar que debido a que la variable dependiente se trata de una serie financiera, el orden de los datos importa. Dado lo anterior, la división de la muestra de entrenamiento y muestra de prueba no se hace de manera aleatoria con validación cruzada, sino que se utilizó una muestra que respetan una cronología temporal. Particularmente, la muestra de entrenamiento que se utilizó para la definición de las métricas de evaluación es el 90% de la serie de la dirección de los retornos del S&P 500, es decir desde el 2011-08-01 hasta el 2017-11-16. Por otra parte, la base de prueba correspondió al 10% restante de la serie de la

5. RESULTADOS

dirección de los retornos del S&P 500, es decir desde el 2017-11-17 hasta el 2018-08-01.

Luego de estimar el F1 ponderado de los 10.000 clasificadores, se encontró un promedio de 0,47 y una exactitud promedio de 0,5. En este sentido, el límite inferior de la métrica del F1 ponderado es de 0,47, resultado equivalente al de un clasificador estocástico. Por otra parte, la exactitud al no utilizar un ponderador de categorías tiene un límite inferior de 0,5. Adicionalmente, se debe resaltar que el puntaje mínimo del F1 ponderado²⁸ aplica siempre y cuando la distribución esté más o menos equilibrada entre las diferentes categorías. En otras palabras, si un modelo estima que todo se concentra en una sola categoría, el puntaje F1 ponderado puede tomar valores menores al de un clasificador estocástico. Particularmente, un clasificador que asigne todas sus observaciones a la categoría bajista obtiene un puntaje F1 ponderado de 0,30 y una exactitud de 0,46. Por otra parte, un modelo que asigne todas sus observaciones a la categoría alcista obtiene un puntaje F1 ponderado de 0,33, y una exactitud de 0,5. Adicionalmente, un modelo que asigne todas sus observaciones a la categoría neutral tiene un F1 ponderado de 0, y una exactitud de 0,03. Finalmente, no es deseable un clasificador que concentre todas sus predicciones en una sola estrategia de inversión. Entre más distribuida sea la predicción, y más se relacione con los datos originales, los criterios F1 ponderado y la exactitud predicha son más altos.

5.3.2 Modelo autorregresivo

Por otra parte, el segundo punto de referencia que se utilizó para evaluar el desempeño de los modelos que incluyen el análisis de sentimiento fue un modelo autorregresivo. Luego de evaluar las diferentes familias de modelos, y los 13 primeros rezagos en las múltiples ventanas se puede concluir que, bajo el puntaje F1 ponderado y la exactitud, el primer rezago aumentó la predictibilidad de la dirección de los retornos del S&P 500 respecto a un modelo de clasificación estocástico. Los otros rezagos no superan al modelo de clasificación estocástico bajo los criterios de exactitud y precisión (Ver Tabla 8). En este caso, al igual que en el anterior, se utilizó una muestra de entrenamiento del 90% desde el 2011-08-01 hasta el 2017-11-16, y una muestra de prueba del 10% comprendida entre el 2017-11-17 hasta el 2018-08-01.

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,45	0,43	0,45
Exactitud	50%	50%	51%

Tabla 8: Métricas: Fuera de muestra (10%) - Modelo autorregresivo orden 1.

²⁸Para más información de esta medida consultar Goutte & Gaussier (2005)

5. RESULTADOS

En este orden de ideas, se puede afirmar que la métrica con la cual se compararon los modelos que incorporan el análisis de sentimiento y el análisis de inversionista fue con un puntaje F1 ponderado promedio de 0,44 y una exactitud promedio de 0,5. Se debe resaltar que, si bien el puntaje F1 ponderado fue mayor que el modelo de clasificación estocástico, la capacidad de predicción del modelo autorregresivo medido por la exactitud fue igual al modelo de clasificación estocástico. Este resultado se ve en la matriz de confusión (Ver Figura 4) en donde la clasificación de los 3 casos estuvo sesgada con una dirección alcista.

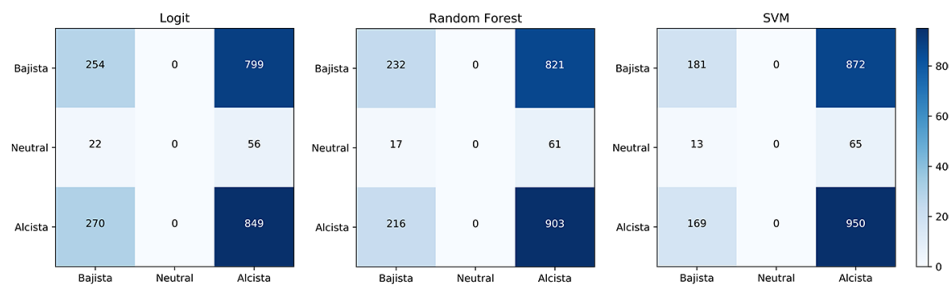


Figura 4: Matriz de confusión: Fuera de muestra (10%) - Modelo autorregresivo de orden 1.

Por otra parte, la diferencia entre el modelo autorregresivo y el modelo estocástico respecto al F1 ponderado se generó porque el modelo autorregresivo al tener el primer rezago predice mejor la proporción de la distribución de las diferentes categorías, pero no la exactitud. La intuición detrás de este resultado es que los modelos estocásticos al no conocer la distribución de los datos originales asignan aleatoriamente una categoría sin respetar proporciones de la muestra original, reduciendo así el puntaje F1 de cada una de las categorías y con esto el F1 ponderado.

Finalmente, es pertinente resaltar que la metodología de Heston et al. (2010) de ventanas de 30 minutos evaluada en modelos discretos aplicados al S&P 500 no aumentó la predictibilidad al incorporar los rezagos de la variable dependiente. Este resultado no es comparable con indicado por Heston et al. (2010), ya que la metodología original toma en cuenta los retornos continuos que permiten capturar patrones de estacionalidad de la serie. En este sentido, siguiendo a Renault (2017) se evaluó si con la inclusión de los sentimientos de los inversionistas agregados en ventanas de 30 minutos aumenta la predictibilidad del S&P 500 aplicado al caso discreto.

5.3.3 Inclusión del sentimiento de los inversionistas

La primera especificación del modelo de pronóstico evaluó la predictibilidad que tiene el análisis de sentimiento de los inversionistas en la dirección de los retornos del S&P 500. En ésta

5. RESULTADOS

se logró determinar que, bajo el criterio F1 ponderado y la exactitud, el modelo Logit alcanzó un mayor desempeño que la SVM y el RF (Ver Tabla 9). La base de entrenamiento correspondió al 90% de la serie de la dirección de los retornos del S&P 500, es decir desde el 2011-08-01 hasta el 2017-11-16. Adicionalmente, la base de prueba correspondió al 10% restante de la serie de la dirección de los retornos del S&P 500, es decir desde el 2017-11-17 hasta el 2018-08-01.

Ventana	Logit	SVM - RBF	Random Forest
F1 ponderado	0,56	0,53	0,49
Exactitud	58%	57%	50%

Tabla 9: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas.

En esta primera especificación se puede afirmar que, bajo las métricas del puntaje F1 ponderado y exactitud, incluir la probabilidad de sentimiento de los mensajes de los inversionistas organizado en ventanas de 30 minutos aumentó la predictibilidad de la dirección de los retornos del S&P 500. Particularmente, el modelo Logit que incorporó el análisis de sentimiento de los inversionistas supera al modelo Logit autorregresivo de orden 1 en un 24,4% bajo el puntaje F1 ponderado y en un 16% bajo la medida de exactitud. Este resultado se puede ver en la diagonal de la matriz de confusión (Ver Figura 5).

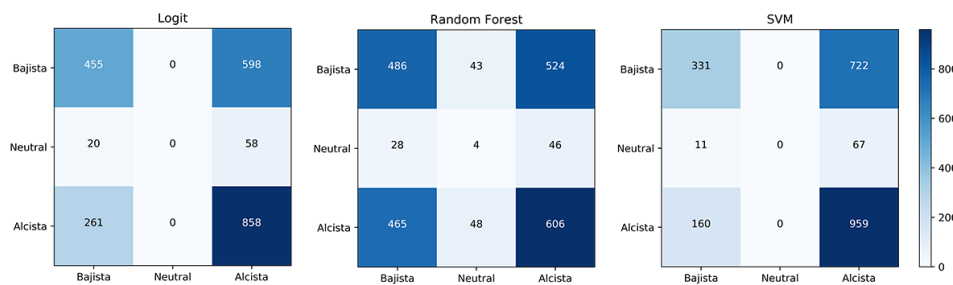


Figura 5: Matriz de confusión: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas.

Respecto a la categoría neutral, si bien son pocas las veces que esto sucede, el modelo no logró predecir ningún verdadero positivo en esta categoría. La explicación al bajo desempeño de Logit en la categoría neutral puede explicarse por la ausencia de una categoría neutral dentro de la clasificación de los sentimientos de los inversionistas. Al no existir una categoría neutral en los mensajes publicados en StockTwits, el modelo de análisis de sentimiento de los inversionistas fue entrenado para asignar una probabilidad alcista o bajista, afectando así la predicción de la dirección neutral en el modelo Logit.

Adicionalmente, se debe resaltar que estos resultados discretos son consistentes con los hallazgos continuos de Renault (2017). Lo anterior sugiere que en el caso discreto la inclusión de sentimiento agregado en ventanas de 30 minutos sí tiene un poder de predicción sobre el desempeño del índice accionario. En este sentido, se puede afirmar que la inclusión de los

5. RESULTADOS

sentimientos de los inversionistas aumentaron la predictibilidad de la dirección de los retornos del S&P 500 en las direcciones alcistas y bajistas. Por otra parte, también se encontró que la mejor forma de modelar la relación no lineal que existe entre la dirección de los retornos y los sentimientos de los inversionistas es por medio de un modelo Logit, teniendo en cuenta la métrica utilizada y los modelos evaluados. Finalmente, en la sección de análisis complementarios se modificó el supuesto de usuarios homogéneos y se probó si el sentimiento de los inversionistas tiene poder de predicción en otras acciones, replicando esta metodología para la acción de Apple.

5.3.4 Inclusión de sentimiento de noticias

La segunda especificación evaluó la predictibilidad que tiene el análisis de noticias financieras en la dirección de los retornos del S&P 500. En esta especificación se determinó que, bajo el criterio del puntaje F1 ponderado y la exactitud, el modelo Random Forest alcanzó un mayor desempeño que la SVM y el Logit (Ver Tabla 10). La base de entrenamiento y de prueba fue dividida en las mismas proporciones que el análisis de sentimiento de los inversionistas.

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,33	0,33	0,49
Exactitud	50%	50%	50%

Tabla 10: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de las noticias.

En esta segunda especificación se encontró que, bajo las métricas de F1 ponderado y exactitud, incluir la probabilidad de sentimiento de las noticias agregadas y organizadas en ventanas de 30 minutos no aumentó la predictibilidad de la dirección de los retornos del S&P 500. Este resultado se puede observar en la diagonal de la matriz de confusión del RF (Ver Figura 6).

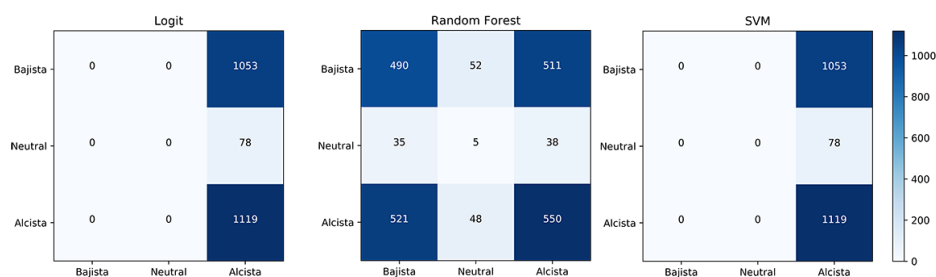


Figura 6: Matriz de confusión: Fuera de muestra (10%) - Inclusión del sentimiento de las noticias.

Respecto al Logit y la SVM, es posible determinar que el puntaje F1 ponderado fue inferior a 0,4, es decir que los dos modelos clasifican todos los movimientos como alcistas. El Logit y

5. RESULTADOS

la SVM se descartaron, puesto que el modelo estocástico y el modelo autoregresvo de orden 1 lograron un mejor desempeño bajo el F1 ponderado.

En este sentido, una de las razones por las cuales la inclusión de las noticias no mejoró la predicción de la dirección de los retornos del S&P 500 se debe a la agrupación de noticias de diversas temáticas. Se esperaba que al utilizar como muestra de entrenamiento noticias citadas por los inversionistas en el muro del S&P 500, las noticias de diferentes temas fueran marcadas como neutrales. Sin embargo, esto no sucedió, ya que existen términos que no son únicos en las noticias financieras que imprimen un sentimiento. Muestra de lo anterior, es la correlación de 3,3% que existe entre la probabilidad de sentimiento de los inversionistas y noticias organizado en ventanas de 30 minutos. En la sección de análisis complementarios se estimó el efecto del sentimiento de las noticias agrupadas por tópicos sobre la dirección de los retornos.

5.3.5 Modelo: Inversionistas & Noticias

La tercera especificación evaluó el efecto de la inclusión conjunta de los sentimientos de los inversionistas y de las noticias sobre la dirección de los retornos del S&P 500. Si bien en la segunda especificación se determinó que el sentimiento de las noticias no aumentó la predicción, en esta aproximación se evaluó si la interacción entre el sentimiento de los inversionistas y de las noticias aumentaba la predicción del modelo.

En esta especificación se logró determinar que, bajo el criterio F1 ponderado y la exactitud, el modelo Logit y la SVM alcanzaron un mayor desempeño que el RF (Ver Tabla 11). La base de entrenamiento y de prueba fue dividida al igual que las especificaciones anteriores. Dado los resultados de la tabla anterior, se puede afirmar que, bajo las métricas de

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,56	0,56	0,52
Exactitud	58%	58%	53%

Tabla 11: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas y noticias. F1 ponderado y exactitud, la inclusión conjunta del modelo de análisis de sentimiento de los inversionistas y noticias iguala a la primera especificación que solo incluye el sentimiento de los inversionistas. En este sentido, al igual que la segunda especificación, se puede determinar que la inclusión del sentimiento de las noticias no aumentó la capacidad de predicción de la dirección de los retornos del S&P 500. Finalmente, este resultado se puede observar al comparar la diagonal de la matriz de confusión de la primera especificación y la tercera especificación (Ver Figura 7).

6. ANÁLISIS COMPLEMENTARIOS

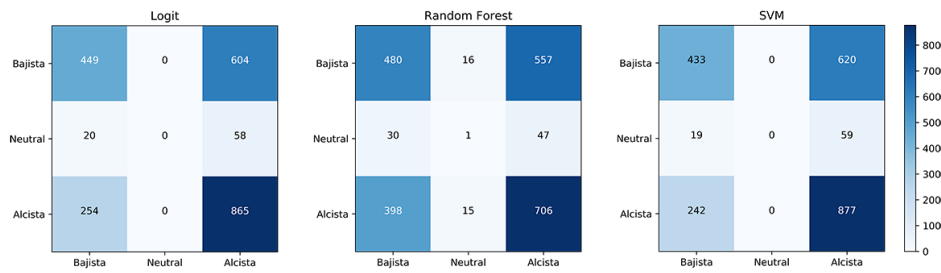


Figura 7: Matriz de confusión: Fuera de muestra (10%) - Inclusión del análisis de sentimiento de los inversionistas y noticias.

Con las tres especificaciones del modelo de pronóstico se puede concluir que, bajo la metodología planteada, la inclusión del análisis de sentimiento de los inversionistas utilizando agentes homogéneos mejoró la predictibilidad de la dirección de los retornos del S&P 500 en 16% bajo medida de exactitud. De esta forma, la predicción de un modelo estocástico con una exactitud del 50% fue superada con una exactitud de 58% al incluir como variable independiente la probabilidad de sentimiento de los inversionistas.

En contraste, se determinó que, bajo el criterio de exactitud, la predicción de la dirección de los retornos no mejoró cuando se incluyó el sentimiento de las noticias agregadas. Este resultado se confirmó con la segunda y tercera especificación, que al estimar de manera individual y conjunta el efecto del análisis de sentimiento de las noticias sobre la dirección de los retornos, dio un resultado similar al modelo autorregresivo de orden 1. En la sección de análisis complementarios se determina si este resultado se debe a la agrupación del sentimiento, probando una agrupación de noticias por temas.

6 Análisis complementarios

En esta sección se profundiza en análisis adicionales que permiten la sofisticación y evaluación de algunos de los supuestos del modelo planteado en la sección de resultados. Esta sección se divide en dos partes: en la primera se ahonda en el modelo de usuarios heterogéneos y la agrupación de noticias. En la segunda parte se evalúa la metodología en la acción de Apple y se modifica las ventanas de tiempo. Cada análisis se realizó de manera independiente, manteniendo los supuestos de la metodología planteada en la sección de resultados a menos que se indique lo contrario.

6 Usuarios heterogéneos

Dado que unos de los objetivos secundarios de esta investigación fue profundizar en el análisis de sentimiento en usuarios heterogéneos, el primer supuesto que se modificó fue el de usuarios homogéneos. De este modo, con el objetivo de identificar los agentes ruidosos y los agentes legítimos, se agruparon los 40.116 usuarios únicos que publicaron en el muro del S&P 500 de StockTwits y se les asignó un grupo según su desempeño histórico al momento de predecir la dirección de los retornos del S&P 500. El resultado de este ejercicio se divide en dos partes: la selección del número de grupos y la agrupación de los usuarios.

La selección del número de grupos se realizó a partir del criterio de la silueta. Al no tener información externa que permitiera validar si la agrupación por usuarios es correcta, se utilizó esta medida que combina el criterio de cohesión²⁹ y el criterio de separación³⁰, ambas medidas emplean la información de cada grupo y evalúan el desempeño respecto a los demás grupos. El criterio de la silueta se define como la diferencia entre la separación y la cohesión sobre el máximo de la cohesión y la separación. La intuición detrás, es que un buen agrupamiento se logra cuando se alcanza la mayor separación entre los grupos respetando la cohesión entre estos. Este criterio toma valores de -1 y 1, -1 significa un mal agrupamiento, 0 indiferente y 1 un buen agrupamiento.

Al utilizar el criterio de la silueta se puede afirmar que el número de grupos que logró la mejor separación y la mayor cohesión entre los grupos fue de 4 con un valor de 0,49. El resultado de esta agrupación y el criterio de la silueta se presentan en la Figura 8. Adicionalmente, debido a que la agrupación se realizó con los usuarios en cuya historia han publicado más de 10 mensajes, existe un quinto grupo que contiene a estos usuarios.

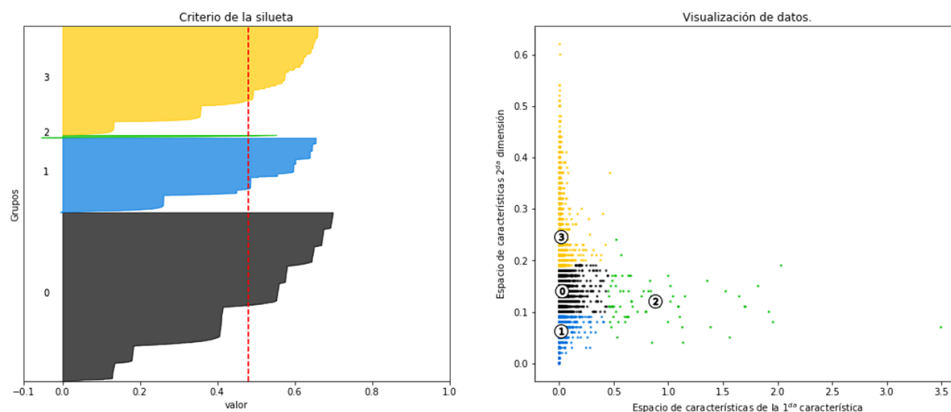


Figura 8: Criterio de la Silueta - K-medias

²⁹Evalúa que el miembro de cada grupo debe ser lo más cercano posible a los otros miembros del grupo.

³⁰Evalúa la separación que existe entre los centroides de cada grupo.

6. ANÁLISIS COMPLEMENTARIOS

Luego de la agrupación de los usuarios, se utilizó una medida de calificación para la jerarquización de los grupos. Esta medida es el número de aciertos por individuo sobre el total de mensajes publicados por éste. En la tabla 12 se presenta a detalle cada uno de los grupos.

Grupo	1	2	3	4	5
Calificación	20%	6,3%	12%	14%	25%
Tweets promedio	4	348	11.382	353	143
Acertos	1	23	1234	49	34
Peso sentimiento	0,5	0,625	0,75	0,875	1

Tabla 12: Estadísticas descriptivas: Grupos usuarios.

De los 5 grupos se debe resaltar que, si bien el grupo 1 tuvo una calificación promedio mayor que el grupo 2 y 3, el número de tweets promedios publicados fue el menor de todos. En este sentido, se puede considerar que el grupo 1 es conformado por usuarios poco frecuentes, cuyo desempeño pronosticando la dirección de los retornos del S&P 500 es bajo.

Luego de identificar los grupos se les asignó un peso al sentimiento según la jerarquización. La intuición detrás de estos pesos es que a los usuarios que publicaron menos de 10 mensajes y no tuvieron un buen desempeño pronosticando el índice, se le castiga el sentimiento de todas sus publicaciones por el peso del sentimiento al grupo que pertenece. A modo de ejemplo, en el grupo 2 se tomó un usuario aleatorio y se identificó por el tipo de mensajes que publica. En la Tabla 13 se presentan 4 mensajes aleatorios de *KolaKong*, si bien estos mensajes parecen escritos por un bot por el tipo de palabras que utiliza, la red CNN asigna una probabilidad cercana a 1. Sin embargo, gracias a la agrupación de usuarios heterogéneos se logra castigar el sentimiento de este tipo de usuarios con un sentimiento ponderado menor al estimado. En este caso en particular el sentimiento de todos los mensajes publicados por este usuario se agregaron asignándole la mitad del peso.

Fecha	Mensaje	Sentimiento	Sentimiento ponderado
2017-08-01T23:12:10Z	SPYSPX AAPLFB NFLXNVDA Here's how he replied ...	0,92	0,57
2017-08-01T23:12:10Z	SPYSPX AAPLFB NFLXNVDA Here's how he replied ...	0,92	0,57
2017-08-01T21:09:37Z	SPYSPX AAPLFB NFLXAMZN \$NVDA ALL newbies be warned about a SHYSTER ...	1.00	0,62
2017-08-01T21:09:37Z	SPYSPX AAPLFB NFLXAMZN \$NVDA ALL newbies be warned about a SHYSTER ...	1.00	0,62

Tabla 13: Mensajes - Usuario aleatorio: *KolaKong*.

Luego de penalizar el sentimiento de los inversionistas por el grupo al que pertenecen, este se utilizó para la construcción de las ventanas de 30 minutos. Con esta aproximación se logró una agregación heterogénea del sentimiento de los inversionistas en la que asigna un mayor peso al sentimiento de los usuarios con un mayor número de aciertos históricos.

Particularmente, el Logit que incorporó el análisis de sentimiento de los inversionistas heterogéneos superó al Logit que utilizó como variable dependiente el análisis de sentimiento de los inversionistas homogéneos en un 3,5%, bajo el puntaje F1 ponderado, y en un 1,72%, bajo

6. ANÁLISIS COMPLEMENTARIOS

la medida de exactitud. (Ver Tabla 14).

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,58	0,56	0,48
Exactitud	59%	58%	50%

Tabla 14: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas heterogéneos.

Adicionalmente, en la matriz de confusión (Ver Figura 9) se observa que con la inclusión del sentimiento heterogéneo cayeron los verdaderos positivos, pero aumentaron los verdaderos negativos con respecto al modelo de sentimiento homogéneo.

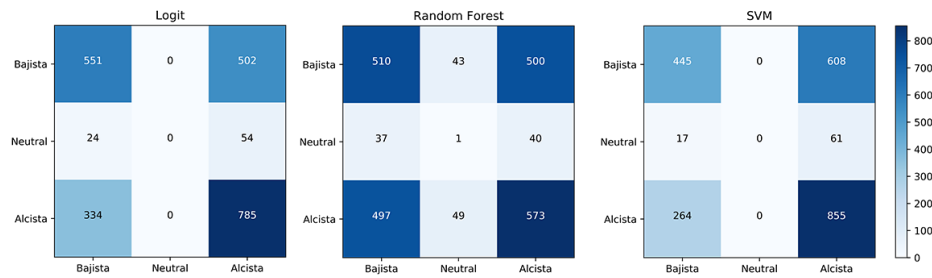


Figura 9: Matriz de confusión: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas heterogéneos.

Finalmente se puede concluir que con la inclusión del análisis de sentimiento de los inversionistas heterogéneos aumentó la capacidad de predicción de la dirección de los retornos respecto al análisis de sentimiento de los inversionistas homogéneos. La intuición detrás de este resultado es que la agrupación de usuarios permite castigar el sentimiento de los usuarios que se pueden considerar como ruidosos o novatos, mejorando la precisión del sentimiento agregado.

6 Agrupación de noticias

Se modificó el supuesto de que todas las noticias, sin importar el tema, afectan a la dirección de los retornos del S&P 500, y se evaluó el efecto del sentimiento de las noticias agrupadas por tópicos. Esto debido a que la hipótesis de la inclusión del canal de noticias no es correcta bajo la metodología que utiliza todas de noticias para la construcción de ventanas de 30 minutos.

Con el objetivo agrupar las noticias lo más parecido a la realidad y que respete la información conocida a la hora de pronosticar, se construyó una muestra de entrenamiento con 3.600.348 noticias entre el 2011-08-01 y el 2017-11-16, y una muestra de prueba con 546.326 noticias entre el 2017-11-16 al 2018-08-01. Con la división de la muestra se respetó la

6. ANÁLISIS COMPLEMENTARIOS

información disponible que se tiene al predecir la dirección de los retornos del S&P 500, siguiendo los mismos lineamientos de la muestra de entrenamiento y de prueba definida anteriormente.

En este sentido, primero se calibró el Modelo LDA que agrupa las noticias de la muestra de entrenamiento en diferentes tópicos, y luego se introdujeron las noticias de la muestra de prueba y se les asignó un tópico que luego fue utilizado como variable independiente en el modelo de pronóstico de la dirección de los retornos. Los resultados de este ejercicio se dividen en tres partes: la selección del número de grupos de temas, la agrupación de los temas y la evaluación de cada tópico como predictor de la dirección de los retornos del S&P 500.

6.2.1 Número de Tópicos

En primer lugar, la selección del número de tópicos se realizó utilizando el criterio de coherencia de validación interna. En este caso, al tratarse de una agrupación de tópicos y al no existir información externa que permitiera validar si los temas escogidos fueron correctos, se utilizó esta medida que toma las N palabras de cada grupo de temas y evalúa la similitud de palabras³¹ por pares de palabras en el tema. En este sentido, una buena agrupación de temas es la que genera una coherencia más alta entre los términos que compone cada uno de los grupos de tópicos. A continuación se presentan los resultados de la coherencia de diferentes grupos de temas creados utilizando la muestra de entrenamiento (Ver Figura 10).

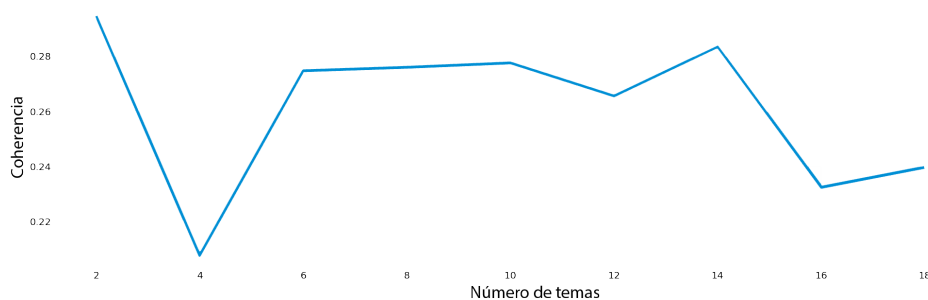


Figura 10: Coherencia - LDA

Utilizando el criterio de la coherencia, se puede afirmar que el número óptimo de tópicos es el que maximiza la coherencia promedio entre las diferentes palabras que compone cada tema. En este caso en particular, la mayor coherencia se logró con 2 tópicos, sin embargo, este número de grupos es reducido para separar una muestra de noticias diversa y amplia. Dado lo anterior, se optó por utilizar 14 temas que corresponden al segundo máximo de la coherencia con un valor

³¹ Esta similitud se evalúa a partir del criterio Pointwise mutual information, para más información de esta medida consultar (Röder, Both & Hinneburg, 2015)

6. ANÁLISIS COMPLEMENTARIOS

de 0,28.

6.2.2 Tópicos

Posterior a la selección del número de temas, se procedió a construir los tópicos utilizando el modelo LDA. En la Tabla 15 se presenta el sentimiento promedio de todas las noticias que componen cada tópico y el número de noticias por tópico. Se puede observar que los temas creados y las palabras que contiene cada uno son consistentes con la realidad. Ejemplo de lo anterior, es el tema 4 que captura las noticias de estados de resultados y anuncios de diferentes compañías que son reportadas en la muestra de entrenamiento.

Tema	Palabras						Sentimiento promedio	Noticias
1	state	court	rule	president	execute	chief	0,612	291.936
2	korea	american	test	target	campaign	senate	0,597	241.987
3	unit	offer	capital	bank	deal	billion	0,711	334.487
4	report	share	profit	financial	earn	loss	0,754	471.136
5	week	risk	partner	health	fiscal	income	0,670	159.598
6	industry	preview	corp	rally	team	giant	0,721	247.098
7	attack	kill	police	shoot	dead	force	0,542	380.185
8	hold	change	euro	time	fitch	increas	0,637	176.401
9	stock	market	bank	share	price	bond	0,655	513.934
10	elect	brexit	government	minister	vote	talk	0,553	391.259
11	trump	russia	win	sharehold	pharmaceutical	mexico	0,583	226.608
12	charge	drug	death	iran	case	annual	0,540	277.803
13	fund	york	city	crash	project	asia	0,608	224.037
14	house	agreement	sign	acquir	start	regulation	0,693	210.204

Tabla 15: Modelo de Tópicos: 6 Palabras más recurrentes en los diferentes documentos por tópico

Por otra parte, se encuentra la medida del sentimiento promedio que se calcula a partir del promedio simple de la probabilidad del sentimiento asignado por la red CNN a las noticias que se encuentra en cada grupo. Esta medida sirve como una aproximación a la percepción de sentimiento que tienen las noticias acerca de un determinado tema. La intuición anterior se captura en el tópico 4 que habla de los beneficios y los estados de resultados de las empresas y tiene un sentimiento promedio de 0,75. Por el contrario, esta medida es menor en el tópico 7 o 12 que hablan sobre las drogas, ataques, disparos, y muerte.

6.2.3 Evaluación de tópicos

Luego establecer los 14 tópicos y clasificar las 4.146.674 noticias, se procedió a agrupar la probabilidad de sentimiento de las noticias de un determinado tema en ventanas de 30 minutos utilizando el promedio simple. Posteriormente, se evaluaron los 14 tópicos por medio de un modelo en el que la variable dependiente capturó la dirección de los retornos del S&P 500 y la variable independiente la probabilidad agregada del sentimiento del tópico j en la ventana $i - 1$ del día t . Utilizando el criterio del puntaje F1 ponderado se determinó que en los 14 tópicos el RF superó al Logit y la SVM. Sin embargo, bajo el criterio de exactitud, el RF no superó al clasificador estocástico, ni al modelo autorregresivo (Ver Tabla 16).

6. ANÁLISIS COMPLEMENTARIOS

	Criterio	Logit	SVM	Random Forest
Tópico 1	F1 ponderado	0,33	0,33	0,47
	Exactitud	50%	50%	50%
Tópico 2	F1 ponderado	0,33	0,33	0,45
	Exactitud	50%	50%	50%
Tópico 3	F1 ponderado	0,33	0,33	0,46
	Exactitud	50%	50%	50%
Tópico 4	F1 ponderado	0,33	0,33	0,46
	Exactitud	50%	50%	50%
Tópico 5	F1 ponderado	0,33	0,33	0,46
	Exactitud	50%	50%	50%
Tópico 6	F1 ponderado	0,33	0,33	0,46
	Exactitud	50%	50%	50%
Tópico 7	F1 ponderado	0,33	0,33	0,48
	Exactitud	50%	50%	50%
Tópico 8	F1 ponderado	0,33	0,33	0,47
	Exactitud	50%	50%	50%
Tópico 9	F1 ponderado	0,33	0,33	0,45
	Exactitud	50%	50%	50%
Tópico 10	F1 ponderado	0,33	0,33	0,47
	Exactitud	50%	50%	50%
Tópico 11	F1 ponderado	0,33	0,33	0,48
	Exactitud	50%	50%	50%
Tópico 12	F1 ponderado	0,33	0,33	0,46
	Exactitud	50%	50%	50%
Tópico 13	F1 ponderado	0,33	0,33	0,48
	Exactitud	50%	50%	50%
Tópico 14	F1 ponderado	0,33	0,33	0,47
	Exactitud	50%	50%	50%

Tabla 16: Evaluación de Tópicos

De esta forma, se puede afirmar que si bien la agrupación de noticias y el sentimiento promedio de cada grupo son consistentes bajo las métricas planteadas, ninguno de estos tópicos aumentaron la predictibilidad de la dirección de los retornos. En este sentido, se puede concluir que bajo esta metodología la inclusión del canal de noticias no mejora la predictibilidad de la dirección de los retornos del S&P 500.

Una conjetura acerca de la irrelevancia de las noticias en este análisis de predictibilidad radica en la generalidad del S&P 500. Al ser un índice que se encuentra en función del desempeño de 500 empresas, puede fluctuar por cientos de razones que distorsionan el canal de noticias. Diferente al análisis de sentimiento de los inversionistas que se encuentra dirigido al S&P 500 por medio de los mensajes de los inversionistas de StockTwits.

Por otra parte, una segunda conjetura por la cual el efecto del análisis de sentimiento de noticias no es significativo es debido a la naturaleza del sentimiento de las noticias. Como se presenta en el Anexo A.3 la distribución de la probabilidad de sentimiento de las noticias se encuentra desbalanceada a favor de una probabilidad cercana a 1, diferente a la distribución de la probabilidad de sentimiento de los sentimientos de los inversionistas. Lo anterior sugiere que aunque el clasificador de sentimiento de noticias tenga un buen desempeño bajo el criterio AUC, este modelo tiende a asignar probabilidades altas de que una noticia sea clasificada como alcista. La intuición detrás de esta conjetura es que las noticias normalmente no dan directamente recomendaciones, ni expresan un sentimiento del mercado que guíen a los inversionistas menos sofisticados a tomar decisiones en un horizonte de corto plazo.

Debido a que las noticias exhiben hechos pasados, estas requieren de un análisis más

6. ANÁLISIS COMPLEMENTARIOS

profundo sobre el efecto que tiene sobre el S&P 500. Este análisis debe ir más allá de un análisis de sentimiento sobre los términos de la noticia y debe incorporar un razonamiento económico que debe incluir una dimensión temporal que permita estimar el efecto real de una noticia sobre el S&P 500. A modo de ejemplo, dentro de la muestra de noticias se encontraron diferentes titulares sobre masacres, asesinatos, y otros atentados contra la población civil de EE.UU que eran clasificadas con una probabilidad de sentimiento negativo. Sin embargo, no es totalmente claro si el efecto de este tipo de noticias es negativo para la dirección del movimiento del S&P 500, o al contrario, este tipo de noticias fortalecen la demanda de compra de armamento, que a su vez afectan otras empresas que fortalecen el S&P 500. Este ejemplo se puede trasladar a cientos de noticias en las cuales no es claro el efecto del acontecimiento sobre la dirección del movimiento del S&P 500.

6 Evaluación de Apple

Este trabajo de grado evaluó su metodología para predecir la dirección de los retornos S&P 500 a partir del efecto de sentimiento de los inversionistas y noticias; sin embargo, esta es aplicable y generalizable a cualquier índice o acción para la cual se tenga la suficiente información. En este sentido, se replicó la metodología con la acción de Apple con el objetivo de evaluar en una última instancia la hipótesis de que existe un canal constituido a partir de noticias que crean opiniones en los inversionistas respecto al desempeño futuro del índice. Para lograr lo anterior se recalibró el modelo de análisis de sentimiento de los inversionistas con los tweets de Apple. Por otra parte, debido a que el modelo de análisis de noticias se entrenó con noticias de 25 acciones, no fue necesario volver a calibrar el modelo, solo se tomaron las noticias que hablaron o hicieron referencia a Apple.

Con la evaluación de la metodología en la acción de Apple se buscó acotar el efecto del canal de noticias y reducir el ruido que existe en un índice tan complejo como el S&P 500. Se seleccionó esta acción, dada la amplitud de la información disponible en StockTwits (2.394.773) y al gran volumen de noticias (398.072) que hablan acerca de Apple. Se debe resaltar que no se tomó un conjunto de acciones más grande debido a los costos computacionales y monetarios que esto exige.

6.3.1 Análisis de sentimiento

El modelo de análisis de sentimiento de los inversionistas fue calibrado con una base de entrenamiento que correspondió al 95% (544.044 tweets) de la muestra marcada por los usuarios, y fue evaluado en una base de prueba que correspondió al 5% (14.317 tweets) restante. Posterior al procesamiento de los tweets de Apple clasificados por los usuarios de StockTwits y de la optimización de los parámetros de diferentes modelos, se logró concluir que el mejor modelo es la red CNN con AUC de 0,825 (Ver Tabla 17).

Modelo	Logit	SVM	Random Forest	CNN
AUC	0,80	0,77	0,77	0,82
Exactitud	76%	70%	74%	77%

Tabla 17: Métricas: Fuera de muestra (10%) - Análisis de sentimiento de los inversionistas (Apple).

Es importante resaltar que luego de evaluar la metodología del análisis de sentimiento de los inversionistas sobre otra muestra de prueba y entrenamiento, se determinó que los resultados de esta aproximación al sentimiento de los inversionistas es robusta bajo el criterio AUC.

Luego de la selección del modelo, se procedió a clasificar los 1.836.412 tweets no marcados que hablan de Apple, para así construir 13 ventanas de tiempo que agregan el sentimiento de los inversionistas. Por otro lado, respecto al sentimiento de las noticias, se tomaron las 398.072 noticias que hablan de Apple y se clasificaron con la red CNN entrenada con las 6.054 noticias en el modelo de análisis de noticias global.

6.3.2 Modelo de pronóstico

Luego de simular 10.000 clasificadores estocásticos para estimar el mínimo puntaje F1 ponderado de la muestra de entrenamiento de Apple, se encontró que debido a la baja proporción de movimientos neutrales de la acción de Apple, el puntaje F1 ponderado mínimo es de 0,47, y la exactitud mínima es 0,5 al igual que el del S&P 500. Adicionalmente, al evaluar el modelo autorregresivo de orden 1, se determinó un puntaje F1 ponderado de 0,50 y una exactitud de 0,51 alcanzado por un RF. En este orden ideas, la métrica con la cual se compararon los modelos que incorporan el análisis de sentimiento y el análisis de inversionista en la acción de Apple fue con un puntaje F1 ponderado promedio de 0,50 y una exactitud promedio de 0,51.

Posteriormente, se evaluó el efecto que tiene el análisis de sentimiento de los inversionistas en la dirección de los retornos de Apple, encontrando que bajo el criterio F1 ponderado y la

6. ANÁLISIS COMPLEMENTARIOS

exactitud, el modelo Logit alcanzó un mayor desempeño que la SVM y el RF (Ver Tabla 18). La base de entrenamiento correspondió al 90%, es decir desde el 2011-08-01 hasta el 2017-11-16. La base de prueba correspondió al 10%, es decir desde el 2017-11-17 hasta el 2018-08-01.

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,57	0,57	0,52
Exactitud	57%	57%	52%

Tabla 18: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento inversionistas (Apple).

En esta primera especificación se puede afirmar que, bajo las métricas de F1 ponderado y exactitud, incluir la probabilidad de sentimiento de los mensajes de los inversionistas organizado en ventanas de 30 minutos aumentó la predictibilidad de la dirección de los retornos de Apple. Particularmente, el modelo Logit que incorporó el análisis de sentimiento de los inversionistas superó al modelo autorregresivo de orden 1 en un 17% bajo el puntaje F1 ponderado y en un 12% bajo la medida de exactitud (Ver Figura 11).

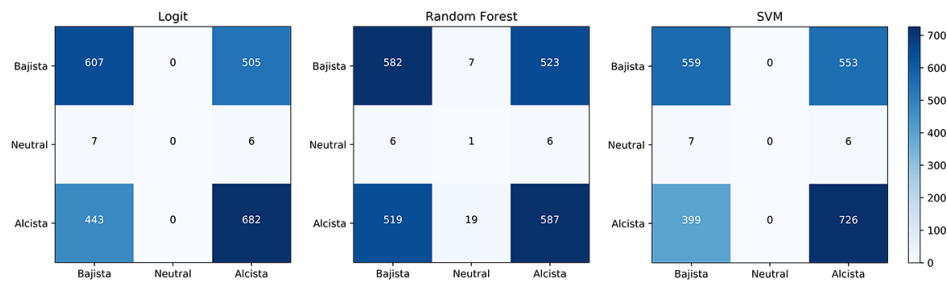


Figura 11: Matriz de confusión: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas (Apple).

En este orden de ideas, respecto a la segunda especificación se puede afirmar que, bajo las métricas de F1 ponderado y exactitud, incluir la probabilidad de sentimiento de las noticias que hablan acerca de Apple organizado en ventanas de 30 minutos aumentó la predictibilidad de la dirección de los retornos de Apple. Particularmente, el modelo RF que incorporó el análisis de sentimiento de las noticias superó al modelo autorregresivo de orden 1 en un 4% bajo el puntaje F1 ponderado y en un 1.96% bajo la medida de exactitud (Ver Tabla 19).

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,33	0,33	0,52
Exactitud	50%	50%	52%

Tabla 19: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de noticias (Apple).

En el caso de Apple, a diferencia de S&P 500, el efecto del canal de noticias aumentó el poder de predicción de la dirección de los retornos. Sin embargo, el efecto de este canal es bajo en comparación del efecto del análisis de sentimiento de los inversionistas (Ver Figura 12).

6. ANÁLISIS COMPLEMENTARIOS

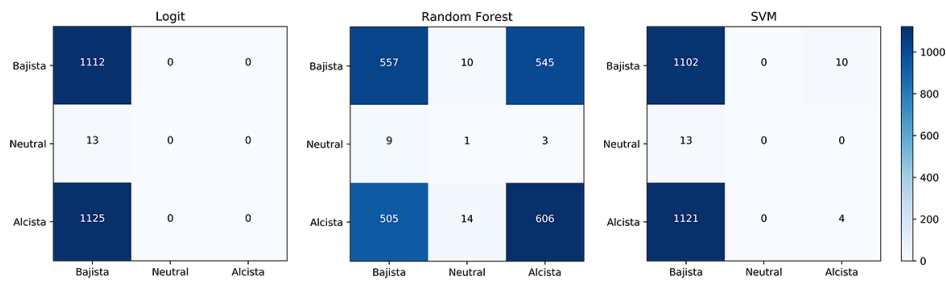


Figura 12: Matriz de confusión: Fuera de muestra (10%) - Inclusión del sentimiento de noticias (Apple).

Finalmente, con la tercera especificación se puede afirmar que, bajo las métricas del puntaje F1 ponderado y exactitud, el efecto de la inclusión conjunta de los sentimientos de los inversionistas y de las noticias disminuyó la predictibilidad de la dirección de los retornos de Apple. Bajo las métricas definidas (Ver Tabla 20) ninguno de los modelos superó al Logit que incorporó el análisis de sentimiento de los inversionistas.

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,50	0,48	0,47
Exactitud	53%	54%	50%

Tabla 20: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas y noticias (Apple).

Si bien se encontró que el análisis de sentimiento de noticias individual aumentó la predictibilidad de la dirección de los retornos de Apple, al incluirlo conjuntamente con el sentimiento de los inversionistas la predictibilidad disminuyó. Lo anterior puede ser explicado por la correlación negativa de -0,4154% que existe entre el análisis de sentimiento de los inversionistas y de noticias. Dado lo anterior, se puede afirmar para el caso particular de Apple que aunque el efecto del análisis de sentimiento de noticias aumentó la predictibilidad de la dirección de los retornos, éste incremento es menor y va en dirección contraria al efecto generado por la inclusión del análisis de sentimiento de los inversionistas.

6 Evaluación de ventanas de tiempo

Esta investigación utilizó en su metodología la aproximación de ventanas de 30 minutos planteada por Heston et al. (2010) para lograr una mayor exactitud a la hora de pronosticar los retornos futuros. Si bien se encontró que con esta metodología la inclusión del análisis de sentimiento mejoró la predictibilidad de la dirección de los retornos, se evaluó el desempeño de este resultado en ventanas de 15 minutos y 1 hora. En este sentido, se mantuvo constante el análisis de sentimiento homogéneo presentado en la sección de resultados y se excluyó el análisis

6. ANÁLISIS COMPLEMENTARIOS

de sentimiento de noticias debido a que no aumentaba la predictibilidad.

6.4.1 Ventana de 1 hora

Para las ventanas de 1 hora se puede afirmar que, bajo las métricas del puntaje F1 ponderado y exactitud, el Logit tiene un desempeño menor que un Logit con ventanas de 30 minutos. Bajo el criterio del puntaje F1 ponderado y el criterio de exactitud, este resultado es menor en 3,5% y % 1,7% respectivamente (Ver Tabla 21). En este caso, al igual que en la metodología original, se mantuvieron las proporciones de la base de entrenamiento y la base de prueba. Al modificar las ventanas de 30 minutos por ventanas de 1 hora disminuyeron las observaciones.

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,54	0,54	0,51
Exactitud	57%	57%	51%

Tabla 21: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas (Ventanas 1 hora).

En este orden de ideas, se puede afirmar que al agrupar el sentimiento en ventanas de 1 hora, se suavizó el efecto del análisis de sentimiento de los inversionistas, disminuyendo la capacidad de predicción del modelo de pronóstico de la dirección de los retornos del S&P 500. La intuición detrás de este resultado es que en una ventana de 1 hora el mercado puede experimentar una gama variada de sentimientos que al agregarlos con un promedio pierden su varianza y se suaviza el efecto. Lo anterior se puede ver en la diagonal de la matriz de confusión (Ver Figura 13).

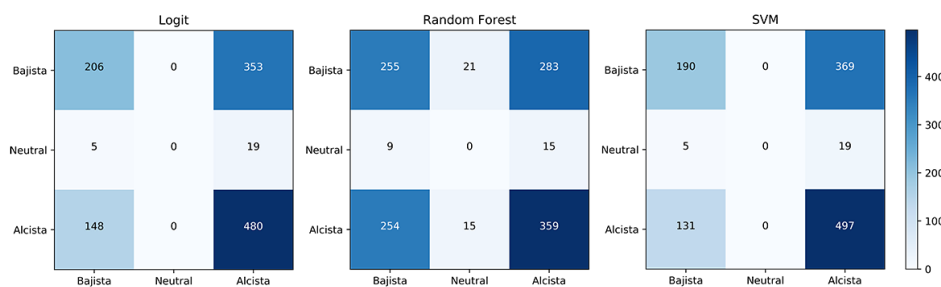


Figura 13: Matriz de confusión: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas (Ventanas 1 hora).

6.4.2 Ventana de 15 minutos

Para las ventanas de 15 minutos se puede afirmar que, bajo las métricas del puntaje F1 ponderado y exactitud, el Logit tiene un mejor desempeño que un Logit con ventanas de 30 minutos. Bajo el criterio del puntaje F1 ponderado y el criterio de exactitud este resultado es

7. CONCLUSIONES

mayor en 1,78% y 3,44% respectivamente (Ver Tabla 22). En este caso, al igual que en la metodología original, se mantuvieron las proporciones de la base de entrenamiento y la base de prueba. Al modificar las ventanas de 30 minutos por ventanas de 15 minutos aumentaron las observaciones.

Ventana	Logit	SVM	Random Forest
F1 ponderado	0,57	0,53	0,48
Exactitud	60%	57%	50%

Tabla 22: Métricas: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas (Ventanas 15 minutos).

En este orden de ideas, se puede afirmar que al modificar las ventanas de 30 minutos por las de 15 minutos el desempeño general de todos los modelos aumentan (Ver Figura 14). La intuición detrás de este resultado es que al agrupar en ventanas de 15 minutos, el sentimiento promedio que se captura en cada ventana es un mejor predictor de la dirección de los retornos del índice.

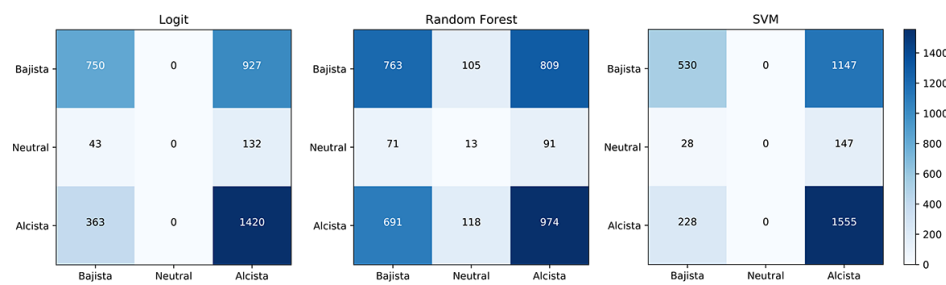


Figura 14: Matriz de confusión: Fuera de muestra (10%) - Inclusión del sentimiento de los inversionistas (Ventanas 15 minutos).

A modo de conclusión, se puede afirmar que al reducir las ventanas de tiempo mejora la precisión del modelo de pronóstico de la dirección de los retornos. Este cambio puede ser explicado por dos efectos, el primero es el cambio generando por un aumento en la correlación de la dirección de los retornos del S&P 500 y la probabilidad de sentimiento de los inversionistas. Particularmente, la correlación disminuye a medida que la ventana se reduce: en ventanas de 1 hora es de 16,9%, en ventanas de 30 minutos 17,5%, y en ventanas de 15 minutos 19,3%. El segundo efecto es explicado por un aumento en el número de datos que ayudan a que los modelos utilizados se ajusten mejor.

7 Conclusiones

Con el objetivo analizar el grado de predictibilidad de la dirección de los movimientos de corto plazo del S&P 500 a partir de los sentimientos de los inversionistas publicados en StockTwits, se construyó una metodología que, a partir de mensajes etiquetados por los

7. CONCLUSIONES

usuarios de StockTwits, logró estimar una medida representativa de sentimiento de los inversionistas y de las noticias. Esta medida luego fue utilizada en un modelo de pronóstico de la dirección de los retornos del S&P 500 y la acción de Apple en ventanas de 30 minutos.

Para definir una medida de sentimiento de los inversionistas y de las noticias se tomó como insumo 621.680 tweets con un sentimiento marcado por los inversionistas para calibrar dos modelos supervisados que se utilizaron para clasificar 2.651.397 de tweets y 4.146.674 noticias con una probabilidad de sentimiento. Posteriormente, se planteó un modelo discreto que evaluó el efecto de la inclusión de los sentimientos de los inversionistas y las noticias sobre la dirección de los retornos del S&P 500 en ventanas de 30 minutos. Adicionalmente, con el objetivo de complementar el análisis, se diseñó una metodología de agentes heterogéneos que permitió castigar el sentimiento de los agentes menos informados, y se exploró una metodología de agrupación por tópicos de noticias para evaluar el efecto del canal de noticias de manera desagregada.

Bajo la metodología planteada se determinó que el modelo que incorporó el análisis de sentimiento de los inversionistas se superó al modelo autorregresivo en un 24,4% bajo el puntaje F1 ponderado y en un 16% bajo la medida de exactitud. Si este modelo se hubiese utilizado desde el 2017-11-17 hasta el 2018-08-01, se habrían predicho correctamente 858 subidas de un total de 1.119, y 455 caídas de un total de 1.053 y 0 movimientos neutrales de 78.

El modelo de pronóstico que incorpora el sentimiento de los inversionistas tiene un mejor desempeño prediciendo movimientos alcistas que bajistas. Este resultado no es desproporcionado, y ha sido abordado por diferentes autores como Stambaugh, Yu & Yuan (2012) que concluyeron que los inversionistas impulsados por el sentimiento participan y comercian más agresivamente en períodos de sentimientos alcistas.

Por otra parte, al aplicar la metodología propuesta en la acción de Apple, se encontró que con la inclusión de los sentimiento de los inversionistas se superó al modelo autorregresivo de orden 1 en un 17% bajo el puntaje F1 ponderado y en un 12% bajo la medida de exactitud. En el caso de Apple, al igual que del S&P 500, se logró un mejor desempeño en el pronóstico de movimientos alcistas que bajistas.

Por otro lado, en la sección de análisis complementario se determinó que la predictibilidad generada por el análisis de sentimiento de los inversionistas con usuarios heterogéneos aumentó respecto al modelo de usuarios homogéneos en un 3,5%, bajo el puntaje F1 ponderado, y en

7. CONCLUSIONES

un 1,72%, bajo la medida de exactitud. Este resultado concuerda con la tesis de Kyle (1985), pues al utilizar la aproximación de usuarios heterogéneos se logra filtrar los agentes ruidosos del mercado y con esto capturar el efecto del sentimiento de los agentes informados.

De otra manera, añadir el análisis de sentimiento de noticias como variable explicativa no mejora la predictibilidad de la dirección de los retornos del S&P 500. La inclusión del análisis de sentimiento de noticias no supera la predicción de un modelo autorregresivo de orden 1 bajo las métricas establecidas. Este resultado se reafirmó en la sección de análisis complementarios, en la cual por medio del modelo de agrupación de noticias se sofisticó la agregación del sentimiento, y se determinó que ninguno de los 14 temas evaluados aumentó la predicción respecto a un modelo autorregresivo.

Una conjetura acerca de la irrelevancia de las noticias en este análisis de predictibilidad radica en la generalidad del S&P 500. Al ser un índice que se encuentra en función del desempeño de 500 empresas, puede fluctuar por cientos de razones que distorsionan el canal de noticias. Este análisis es diferente para el canal de sentimiento, debido a que los inversionistas de StockTwits están opinando directamente sobre la totalidad del índice del S&P 500, distinto a las noticias que presentan diferentes realidades y temas que abordan de manera indirecta a las empresas que componen al S&P 500. Esta conjetura es soportada al analizar individualmente la acción de Apple, en donde el modelo que incorpora únicamente las noticias que hablan acerca de la acción superó en 4% en medida F1 (2% en exactitud) al modelo sin noticias.

Por otra parte, una segunda conjetura por la cual el efecto del análisis de sentimiento de noticias no es significativo es debido a la naturaleza del sentimiento de las noticias. La intuición detrás de esta conjetura es que las noticias normalmente no dan recomendaciones directamente, ni expresan un sentimiento del mercado que guíen a los inversionistas menos sofisticados a tomar decisiones en un horizonte de corto plazo. Debido a lo anterior, el tratamiento de noticias requieren de un análisis más profundo sobre el efecto que tiene sobre el S&P 500. Este análisis debe ir más allá de un análisis de sentimiento sobre los términos de la noticia y debe incorporar un razonamiento económico que debe incluya una dimensión temporal que permita estimar el efecto real de una noticia sobre el S&P 500.

Por otro lado, se encontró que en el caso particular de Apple el efecto de los análisis de sentimiento de los inversionistas y noticias tienen efectos diferentes, pues al agregarlos

7. CONCLUSIONES

disminuyó la predictibilidad de la dirección de los retornos del S&P 500. En este sentido, si bien el S&P 500 y Apple son dos casos particulares de un mercado más amplio, estos dan luces para afirmar que los inversionistas de StockTwits no estructuran sus conjeturas por medio de un canal constituido a partir de noticias que crean opiniones en los inversionistas respecto al desempeño futuro de estos dos índices. Una de las explicaciones de este resultado es que los inversionistas de StockTwits le dan más importancia a otros tipos de contenidos a la hora de estructurar sus opiniones respecto al desempeño futuro de una acción. Este resultado se puede confirmar cuando en una muestra de 9.182.301 tweets solo el 0,073% son enlaces de 6.754 noticias, por debajo de los 22.330 enlaces que hacen referencia a vídeos de Youtube que hablan de análisis técnico, o a los 18.068 enlaces a Chartmill, un sitio de análisis técnico. En este sentido, por el tipo de enlaces a los que hacen referencia los inversionistas en sus mensajes, se puede caracterizar un tercer canal constituido por diferentes tipos de análisis técnico que influye en las opiniones de los inversionistas.

En este orden de ideas, luego de explorar los efectos del análisis de sentimiento de los inversionistas y de las noticias sobre la dirección de los retornos del S&P 500 y Apple, se puede afirmar que los sentimientos de los inversionistas de StockTwits no son determinados por medio de un canal de noticias. Dado lo anterior, se podría afirmar entonces que los sentimientos de los inversionistas son guiados por creencias estocásticas que afectan los precios y la dirección de los retornos como lo plantearon De Long et al. (1990) y Keynes (2014). Sin embargo, este trabajo de grado no exploró todos los canales que podrían afectar el sentimiento de los inversionistas de StockTwits. En este sentido, si bien esta investigación logró encontrar que la inclusión de los sentimientos de los inversionistas mejoran la predictibilidad de la dirección de los retornos, no se logró determinar si los sentimientos de los inversionistas son estocásticos o están guiados por una racionalidad.

Dado lo anterior, se considera que una oportunidad de mejora para trabajos futuros consiste en explorar los 1.896.007 enlaces embebidos en los mensajes para lograr identificar canales alternativos al de noticias. Si bien se considera que es posible etiquetar todos estos enlaces a partir de algoritmos no supervisados, es relevante mencionar que analizar el tipo de contenido y clasificarlo con un sentimiento alcista o bajista es una tarea muy compleja para un algoritmo, debido al tipo y estructura de contenido: videos, podcast, noticias, entre otros formatos. Una segunda oportunidad de mejora de esta investigación radica en explorar otras fuentes de información como Twitter, Facebook, LinkedIn, entre otros portales, que facilitan

7. CONCLUSIONES

la interacción entre diferentes agentes del mercado financiero. La selección de StockTwits se realizó por información etiquetada de los sentimientos de los inversionistas que ofrece la plataforma, y debido a que es ampliamente utilizada en la literatura.

En este sentido, para futuras investigaciones se sugiere explorar nuevos horizontes que permitan generalizar estos resultados y que ayuden a entender los factores que determinan los sentimientos de los inversionistas. A continuación se mencionan otras líneas posibles de trabajo futuro:

- Explorar el canal de análisis fundamental y el efecto que este tiene.
- Crear un índice con el volumen de búsqueda de términos financieros con la información de Google y evaluar el efecto con el análisis de inversionistas.
- Ampliar la medida de sentimiento a plataformas: Twitter, Facebook, LinkedIn, entre otras.
- Evaluar el efecto del sentimiento en el mercado de renta fija y de commodities.

A Anexos

A Análisis de sentimiento: Fuentes Noticias

Acción (Ticker)	Número de Tweets
SPY	2.757.664
APPL	2.394.773
FB	896.993
AMZN	676.833
SPX	548.134
GOOGL	278.978
GOOG	238.362
BAC	217.511
MSFT	183.842
INTC	117.615
JPM	95.184
JNJ	81.926
T	81.391
BA	75.447
MA	75.399
V	65.608
WFC	55.451
CSCO	55.217
XOM	51.575
HD	48.986
VZ	44.759
PFE	40.829
CVX	30.432
MRK	26.039
PG	23.814
UNH	19.539
Total	9.182.301

Tabla 23: Tweets - StockTwits: 2009-2010

A Análisis de sentimiento de los inversionistas

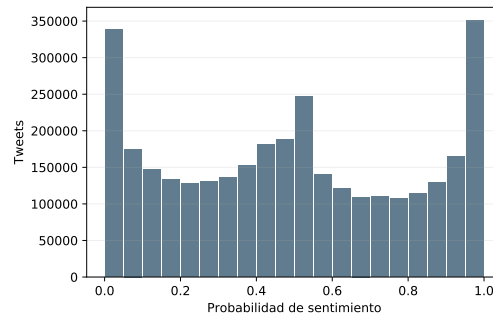


Figura 15: Distribución del sentimiento de inversionistas (Red CNN).

A Análisis de sentimiento de noticias

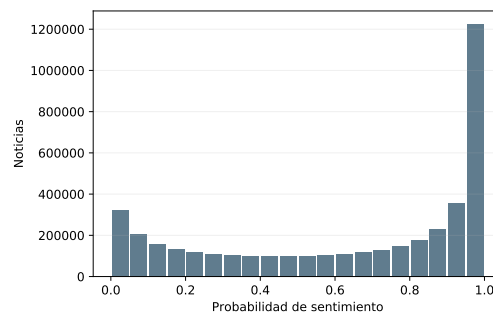


Figura 16: Distribución del sentimiento de noticias (Red CNN).

B Bibliografía

- Akerlof, G. A. & Shiller, R. J. (2010). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton university press.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- Brown, G. W. & Cliff, M. T. (2005). Investor sentiment and asset valuation. *The Journal of Business*, 78(2), 405–440.
- Da, Z., Engelberg, J., & Gao, P. (2014). The sum of all fears investor sentiment and asset prices. *The Review of Financial Studies*, 28(1), 1–32.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4), 703–738.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8), 861–874.
- Gao, L., Han, Y., Li, S. Z., & Zhou, G. (2015). Intraday momentum: The first half-hour return predicts the last half-hour return. *Available at SSRN 2552752*.
- Goutte, C. & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European Conference on Information Retrieval*, (pp. 345–359). Springer.
- Grossman, S. J. (1977). The existence of futures markets, noisy rational expectations and informational externalities. *The Review of Economic Studies*, 44(3), 431–449.
- Grossman, S. J. & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393–408.
- Heston, S. L., Korajczyk, R. A., & Sadka, R. (2010). Intraday patterns in the cross-section of stock returns. *The Journal of Finance*, 65(4), 1369–1407.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Kahneman, D. & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99–127). World Scientific.
- Keynes, J. M. (2014). *Teoría general de la ocupación, el interés y el dinero*. Fondo de cultura

económica.

- Kordonis, J., Symeonidis, S., & Arampatzis, A. (2016). Stock price forecasting via sentiment analysis on twitter. In *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, (pp.36). ACM.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.
- Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69, 14–23.
- Morissette, L. & Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in mathematica. *Tutorials in Quantitative Methods for Psychology*, 9(1), 15–24.
- Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In *Signal Processing, Communication, Power and Embedded System (SCOPES), 2016 International Conference on*, (pp. 1345–1350). IEEE.
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the us stock market. *Journal of Banking & Finance*, 84, 25–40.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, (pp. 399–408). ACM.
- Sabherwal, S., Sarkar, S. K., & Zhang, Y. (2011). Do internet stock message boards influence trading? evidence from heavily discussed stocks with no fundamental news. *Journal of Business Finance & Accounting*, 38(9-10), 1209–1237.
- Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2), 288–302.
- Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5–23.