



**Instituto Católico de Administración y Dirección de Empresas
Facultad de Ciencias Económicas y Empresariales**

Twitter y el Mercado de Valores

**Autor: Alexa Rose Malinowski
Director: José Portela González**

MADRID | junio 2020



ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN	1
1.1 ESTADO DE LA CUESTIÓN	2
1.2 OBJETIVOS	2
1.3 METODOLOGIA	3
1.4 DELIMITACION DEL TEMA	4
2. MACRO TEÓRICO	5
2.1 TWITTER	5
2.1.1 Definición una Red Social	5
2.1.2. Contenido generado por el usuario	5
2.1.3. El nacimiento de Twitter	6
2.1.4. Microblogging	7
2.1.5. Un Canal Nuevo de Información	8
2.1.6. Cómo La Información Se Disemina en Twitter	9
2.1.7. Twitter Financiero	12
2.1.8. Tiempo de reacción del mercado	14
2.2 LA BOLSA	15
2.2.1. Predicción bursátil	15
2.2.2. Finanzas Conductuales	16
2.3 MARCO DE RELACION ENTRE TWITTER Y LA BOLSA	18
2.3.1 Pronóstico relacionado con el volumen de Tweets	18
2.3.2. Indicador de volumen	19
2.3.3. Akamai Technologies	19
2.3.4. Pronosticador Relacionado con Sentimientos	21
2.3.5. Clasificación de Sentimientos	22
2.3.6. Análisis de Sentimientos y Ejemplos del Mercado de Valores .	27
2.4. ESTRATEGIAS DE INVERSION BASADAS EN EL ANÁLISIS DE TWITTER	31
2.4.1. La formulación de las estrategias	31
2.4.2. Algoritmo Predictivo	32
2.4.3. Estrategias Basadas en Algoritmos Predictivos v.s. Índice S&P 500	33
2.4.4. Programación Genética	35

2.4.5. Estrategias Basadas en la Programación Genética de Sentimientos vs. Estrategias Tradicionales	38
3. INVESTIGACIÓN: TWITTER EN LA INDUSTRIA FINANCIERA	40
3.1. PROVEEDORES DE ANÁLISIS DE DATOS DE TWITTER.....	41
3.1.1 Estado actual.....	41
3.1.2. Desafíos para los proveedores de análisis de datos de Twitter	44
3.2. Implementación actual del análisis de Twitter en empresas financieras	46
3.3. LA UTILIZACIÓN DE DATOS DE TWITTER EN LOS <i>HEDGE FUNDS</i>	48
3.3.1. Estado actual.....	48
3.3.2. Desafíos para los <i>hedge funds</i>	50
4. CONCLUSIONES	53
5. BIBLIOGRAFÍA.....	56

Índice de Gráficos

Gráfico 1 Numero de Usuarios Activos de Twitter	7
Gráfico 2 Pequeña Estructura Mundial de Twitter Financiero	13
Gráfico 3 Precio de DJIA.....	14
Gráfico 4 Stock de Akamai Durante la Semana 03-06-2012	20
Gráfico 5 Proceso de Análisis de Sentimientos.....	23
Gráfico 6 Clasificación de los Sentimientos de Tweets	25
Gráfico 7 Rendimiento de Estrategias de Sentimiento y el Índice S&P 500.....	34
Gráfico 8 Mutación de Indicadores Tradicionales y de Twitter.....	36
Gráfico 9 Crossover de Indicadores Tradicionales y de Twitter	37
Gráfico 10 Comparación del Porcentaje de Estrategias Ganadoras	38
Gráfico 11 Comparación de las Diferentes Estrategias de Trading	39
Gráfico 12 Sentimientos y Volumen de Tweets para el S&P 500.....	43
Gráfico 13 La Curva de Difusión de Inversión basada en Twitter en el Sector Financiero	47
Gráfico 14 Retos a los que Se Enfrentan Actualmente los <i>Hedge Funds</i>	50

Índice de Tablas

Tabla 1 Empresas que Realizan Análisis de Datos de Twitter.....	42
Tabla 2 Legislación Vigente para Proteger los Datos de los Consumidores	45

RESUMEN

Desde que se creó la primera bolsa de valores, los inversores han buscado formas de entender sus patrones y predecir sus movimientos. La más mínima ventaja podría multiplicar el valor de los activos de un inversor, mientras que un pequeño error podría resultar en pérdidas devastadoras. En la actualidad, la misma fascinación por el mercado de valores continúa, pero con mayores desafíos. Cada día, miles de millones de activos se negocian en las bolsas de valores de todo el mundo. Una rápida búsqueda en Google sobre la predicción del mercado de valores revela de todo, desde libros, pasando por *webinars*, hasta gurús autoproclamados. Sorprendentemente, el futuro de predecir los movimientos del mercado podría residir en una fuente poco probable: Twitter. Aunque aparentemente no están relacionadas, hay cada vez más pruebas de que estas dos entidades comparten un vínculo predictivo que puede aprovecharse para dar a los inversores una ventaja competitiva. Este trabajo investiga los orígenes de estos vínculos y cómo podría afectar a la industria financiera.

PALABRAS CLAVES: Twitter, el mercado de valores, análisis de sentimientos, fondos de cobertura, canales de información, estrategias de negociación, análisis predictivo, análisis avanzado, programación genética, medios sociales.

ABSTRACT

Since the first stock exchange was established, investors have sought ways to understand its patterns and predict its movements. The slightest advantage could multiply the value of an investor's assets, while a small misstep could result in devastating losses. In the modern day, the same fascination with the stock market continues, but with higher stakes. Every day, billions of assets are being traded on stock exchanges world wide. A quick Google search on stock market prediction reveals everything from books, to webinars, to self-proclaimed gurus. Surprisingly, the future of predicting market movements could reside in an unlikely source: Twitter. While seemingly unrelated, there is mounting evidence that these two entities share a predictive link that can be leveraged to give investors a competitive edge. This work investigates the origins of this links, and how it could impact the financial industry.

KEY WORDS: Twitter, the stock market, sentiment analysis, hedge funds, information channels, trading strategies, predictive analytics, advanced analytics, genetic programming, social media

INTRODUCCIÓN

La empresa Twitter se lanzó en el año 2006, y desde entonces la compañía se ha convertido en una de las redes sociales más conocidas y relevantes en nuestra sociedad. Con un par de teclas, cualquiera puede unirse a Twitter, y ganar acceso a una plataforma que ya tiene más de 326 millones de usuarios (Pew Research Center, 2018). Además, Twitter ha creado un lugar intangible donde todo el mundo desde empresas, consumidores, e incluso presidentes de gobiernos pueden coexistir e interactuar. Cada día, más de 500 millones de Tweets son publicados como media para transmitir ideas, noticias, imágenes, y mucho más (Pew Research Center, 2018). Este flujo de información representa una riqueza de datos, e investigaciones previas han demostrado que las redes sociales están reflejando e influyendo cada vez más en el comportamiento de otros sistemas complejos (Ranco, Aleksovski, Caldarelli, Grčar, & Mozetič, 2015). El vínculo entre Twitter y el mercado de valores en particular ha llamado la atención de investigadores e inversores. La evidencia sugiere que los Tweets publicados en Twitter, pueden directamente influir el movimiento de precios en las bolsas. Un ejemplo reciente que demostró este fenómeno fue cuando el CEO de Tesla, Elon Musk, anunció en Twitter que iba a privatizar Tesla asegurando que la financiación había sido asegurada. Las acciones de Tesla Motors (\$TSLA) terminaron subiendo un 11% (Yahoo Finance, 2018).

Para explicar este tipo de incidentes, muchos equipos de investigadores han intentado a determinar si existe una correlación entre los dos sistemas. Las investigaciones han afrontado la cuestión desde diferentes puntos de vista, y han empleado diferentes métodos para medirlo. Pero casi todos están de acuerdo que hay una dependencia entre el precio de acciones y el sentimiento de Twitter en tweets sobre una empresa (Ranco et al, 2015). A pesar de la exitosa ejecución de estos estudios, muchas preguntas relacionadas con el tema siguen sin respuesta.

Con el presente estudio, el principal objetivo es entender por qué Twitter tiene una relación predictiva con el mercado de valores, y determinar si Twitter puede ser utilizado para ayudar a los inversores a obtener una rentabilidad superior. El trabajo también considerará las implicaciones de la vida real que Twitter podría tener en la industria financiera. Este proyecto se basará en varios estudios y ofrecerá un análisis

exhaustivo de cómo este vínculo se desarrolló, cómo medirlo, y como utilizar la información que nos aporta.

1.1 ESTADO DE LA CUESTIÓN

La explosión de los medios sociales, que incluye el nacimiento y la expansión meteórica de Twitter, ha ocurrido en los últimos 15 años. El descubrimiento del vínculo entre Twitter y el mercado de valores es aún más reciente. Sólo en los últimos años los académicos han comenzado a investigar esta compleja y multifacética relación. La literatura actual sobre este tema ha demostrado con éxito que existe un vínculo predictivo entre Twitter y el mercado de valores. Sin embargo, cada estudio hasta ahora se ha enfocado solamente en un área de nicho de la relación, y no ha proporcionado un análisis exhaustivo. Además, la literatura actual no ha buscado entender por qué existe esta relación y si puede ser aprovechada en el mundo real para ayudar a los inversionistas a obtener un desempeño financiero superior. Este proyecto pretende llenar los vacíos actuales de la literatura utilizando teorías financieras y estudios de caso para entender los orígenes del vínculo entre Twitter y la bolsa. El proyecto también explorará la aplicación actual en la vida real de la inversión en la industria financiera basada en Twitter

1.2 OBJETIVOS

El propósito general de este artículo es explorar la afirmación de que hay un vínculo directo entre la red social Twitter, y la bolsa de valores. Específicamente, esta investigación se enfocará en cómo se formó este vínculo a través del análisis de investigaciones previas, el estudio de casos actuales, y con la ayuda de varias teorías financieras. Además, este trabajo también determinará si el vínculo entre Twitter y el mercado de valores puede ser aprovechado para vencer al mercado

Por lo tanto, los objetivos que este trabajo detalla son los siguientes:

1. Comentar sobre el auge de las redes sociales como Twitter y detallar como facilitaron el intercambio de ideas, opiniones, y noticias a través de contenido generado por el usuario y microblogging
2. Identificar los nuevos canales de información formado por Twitter, y los papeles específicos de cada uno de los actores principales en la diseminación de información en Twitter

3. Entender el papel que cada tipo de usuario juega en la difusión de la información, y cómo el Twitter financiero vincula directamente al Twitter con el mercado de valores
4. Profundizar en ambos tipos de predictores de la bolsa y destacar como se utilizan para extraer información sobre la bolsa
5. Determinar si la información basada en Twitter puede ser utilizada para construir carteras superiores, o estrategias de trading que resulten en mayores ganancias para los inversores
6. Investigar el estado actual de la inversión en la industria financiera basada en Twitter, centrándose específicamente en cómo los *hedge funds* han comenzado a experimentar con los datos de Twitter, lo que les da una ventaja competitiva.

1.3 METODOLOGIA

La metodología de investigación en este proyecto consiste en los siguientes componentes. Se llevará a cabo por separado una extensa investigación sobre las características clave de Twitter y el mercado de valores para asegurar que los lectores entiendan sus respectivas propiedades. Dentro de cada sección se definirán y ampliarán conceptos importantes con la ayuda de la literatura actual sobre cada tema. Estas dos secciones también destacarán por qué estos dos sistemas distintos se han entrelazado cada vez más. La siguiente sección, titulada, Marco de Relación entre Twitter y la Bolsa, se basará en las dos secciones anteriores y analizará la naturaleza interdependiente de la relación a través de los resultados de varios estudios de caso. Esta sección del artículo será de carácter teórico y se centrará en los vínculos entre Twitter y el mercado de valores en un contexto académico. Esta parte del proyecto hará referencia a investigaciones que han sido patrocinadas por universidades líderes y que han sido llevadas a cabo por investigadores de renombre en este campo. Esta sección concluirá con una verificación de si la relación entre Twitter y el mercado puede incorporar un análisis de sentimientos para realizar estrategias de inversión superiores. Se llevará a cabo una comparación entre los resultados de las estrategias basadas en Twitter y las estrategias tradicionales para demostrar la viabilidad o no de estas nuevas estrategias basadas en Twitter y algoritmos de procesamiento del lenguaje.

El último capítulo de este proyecto busca cerrar la brecha entre el mundo académico y el mundo real al determinar si el vínculo entre Twitter y el mercado de valores está siendo utilizado actualmente en el sector financiero. Esta sección consistirá en un análisis macro de dos stakeholders claves en el sector financiero y sus éxitos y desafíos actuales serán evaluados de manera holística. Este proyecto también realizará una reflexión sobre el potencial impacto de la tecnología en el mercado financiero a corto, medio y largo plazo.

La información y los datos en los que se basa esta proyección son los datos recogidos a partir de numerosos estudios que se han llevado a cabo para entender todas las facetas de la relación entre Twitter y el mercado de valores. Estos estudios comenzaron en el año 2011 con los estudios pioneros de autores como Johan Bollen, Huina Mao y Xiaojun Zeng, y se han ido desarrollando hasta la actualidad, donde todavía se sigue investigando la potencial relación entre el análisis de volumen y de sentimiento de Twitter y su potencial predictivo sobre el mercado de valores.

1.4 DELIMITACION DEL TEMA

Aunque este trabajo reconoce la existencia de otros sitios de medios sociales, se centra específicamente en Twitter. No se considera la posibilidad o la existencia de una relación entre el mercado de valores y otros sitios de medios sociales.

2. MACRO TEÓRICO

2.1 TWITTER

2.1.1 Definición una Red Social

Twitter oficialmente se lanzó en marzo de 2006 en San Francisco, EE. UU. (Twitter, 2018). Con su lanzamiento, Twitter se unió a la lista de redes sociales que emergieron en las últimas dos décadas. Antes de profundizar en el nicho que Twitter ocupa en la gama de redes sociales, es importante mencionar brevemente algunas de las características que todas comparten. Una red social es definida como una forma de comunicación donde gente puede crear, compartir, y editar contenido (Kaplan & Haenlein 2010). A diferencia de los años iniciales del internet, el contenido y las aplicaciones ya no son creadas y publicadas por individuos, sino que son modificadas continuamente por todos los usuarios de forma participativa y colaborativa (Kaplan & Haenlein 2010). En los años recientes, las redes sociales se han convertido en una plataforma importante donde consumidores pueden intercambiar opiniones, sentidos, e ideas. El potencial que esta fuente de información tiene para prever tendencias en el mundo real ha sido notado en varias investigaciones. Por ejemplo, un equipo de investigadores analizó el contenido de web blogs para predecir el éxito de películas (Mishne and Glance, 2006); mientras el análisis del datos del buscador podía detectar brotes de gripe (Ginsberg, 2009). En comparación con otros medios de comunicación, una red social es particularmente caracterizada por su capacidad de revelar información útil ya que ésta puede reflejar reacciones directas e inmediatas en el mercado. Además, son capaz de capturar datos simultáneamente sobre una amplia gama de temas (He, Guo, Shen, & Akula, 2016).

2.1.2. Contenido generado por el usuario

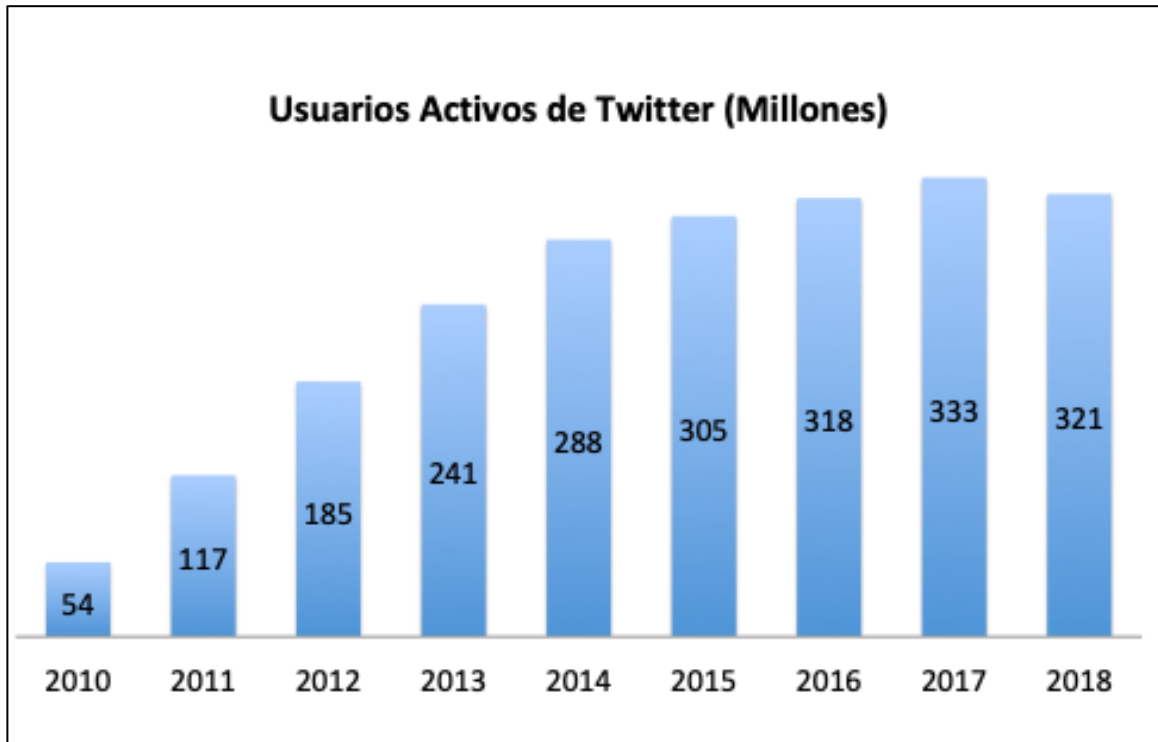
Las redes sociales aportan las plataformas donde usuarios primero crean sus propias cuentas o perfiles, y después ganan la capacidad de generar y publicar su propio contenido. El contenido generado por el usuario (CGU), es una de las principales características que distingue a una red social de otras comunicaciones online, o tipos de media. Según la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2007), hay tres criterios para cumplir para definir CGU. Primero, tiene que

ser publicado en un sitio web de acceso público, o en un sitio de redes sociales accesible para un grupo selecto de personas; segundo, necesita mostrar una cierta cantidad de esfuerzo creativo; y finalmente, debe haber sido creado fuera de rutinas y prácticas profesionales (OCDE, 2007). CGU es un fenómeno que ha crecido rápidamente y en sincronización con el auge de las redes sociales. Debido al hecho de que CGU es generado principalmente por otros usuarios en vez de empresas o anunciantes, consumidores lo consideran más fiable que otras formas de media (He et al., 2016).

2.1.3. El nacimiento de Twitter

Ahora que se ha proporcionado el contexto de cómo se definen las redes sociales en general, este artículo se centrará exclusivamente en la red social Twitter. La misión de Twitter es: dar a todos el poder de crear y compartir ideas e información al instante, sin barreras (Twitter, 2018). Desde la perspectiva de un usuario, es fácil unirse a Twitter. Solo tiene que elegir un usuario y una contraseña. Se puede acceder a la plataforma desde cualquier dispositivo siempre que tenga conexión a internet. Twitter empezó en los EE.UU., pero debido a la naturaleza ubicua de la red, se ha extendido a casi todas las esquinas del mundo y tiene una presencia global. Actualmente, Twitter tiene en promedio 321 millones de usuarios activos, y casi 46% de los usuarios acceden a la plataforma diariamente (Pew Research Center, 2018). Como se puede ver en el gráfico 1 que se encuentra en la siguiente página, desde el lanzamiento de la plataforma, el número de usuarios activos ha crecido significativamente.

Gráfico 1 Numero de Usuarios Activos de Twitter



Fuente: adaptado de (Statista, 2018)

Cuentas de Twitter pueden ser controladas por individuos, o por entidades como empresas y gobiernos. Entonces, la plataforma de Twitter es un ecosistema virtual donde comunidades que anteriormente han estado separadas por el tiempo y espacio se pueden comunicar e interactuar en tiempo real.

También es importante notar como la plataforma de Twitter se organiza. Primero, cada individuo puede seguir y ser seguido por otros usuarios. Esta función crea la red de Twitter, donde los usuario están conectados a través de una relación explícita entre ellos (Zhang, Han, Yang, & Zhang, 2017). Además, cada usuario tiene su propia página donde los Tweets se publican. En cuanto un Tweet es publicado, aparece en el feed, (un flujo continuo de mensajes en Twitter), de Twitter de cada un de sus seguidores. Esta combinación de características hace Twitter especialmente útil para la transmisión de información en tiempo real a una audiencia diversa.

2.1.4. Microblogging

Entre las otras redes sociales más populares, Twitter se destaca como una plataforma de microblogging. Microblogging es un tipo de CGU nuevo que se define como: una

forma de blogging que le permite escribir breves actualizaciones de texto sobre su vida en movimiento y enviarlas a amigos y observadores interesados a través de mensajes de texto, mensajes instantáneos (IM), correo electrónico o Internet. (Java, Song, Finin, & Tseng, 2007). También, el microblogging se caracteriza por su ciclo de vida corto y su velocidad de producción rápida y repetitiva (Gussenhoven, 2014). Twitter encaja muy bien con estas características. El límite de caracteres en Twitter es solo 320 por mensaje. Además, alrededor de 500 millones de Tweets son publicados cada día (Pew Research Center, 2018). Este hecho indica que los usuarios de Twitter acceden la plataforma diariamente y están publicando Tweets con frecuencia. Los motivos más populares por las que publicar un Tweet son: conversaciones, compartir información, y informar sobre las noticias (Java et al., 2007).

2.1.5. Un Canal Nuevo de Información

Estudios anteriores han concluido que existe una relación predictiva entre la difusión de nueva información sobre una empresa en Twitter y la reacción de los mercados financieros. Antes de profundizar en este vínculo directo entre Twitter y la bolsa, es importante comprender como Twitter ha cambiado como la información se disemina entre usuarios. Ya se ha indicado que cuentas de Twitter no son limitadas a personas individuales. Empresas, gobiernos, y otras entidades también pueden abrir and manejar su propia cuenta de Twitter. Entonces, Twitter aporta una plataforma única donde varios stakeholders se intercambian información que ha llevado a la creación de nuevos canales de información. Por ejemplo, en el pasado, cuando una empresa quería emitir un comunicado de prensa, escribía el anuncio internamente y luego dejaba que la información fuera difundida por un tercero. Ahora, las empresas utilizan su cuenta de Twitter para publicar información directamente en el mercado sin necesidad de intermediarios. A su vez, los usuarios de Twitter pueden reaccionar y comentar sobre dicho anuncio en una plataforma compartida (Blankespoor, Miller, & White, 2014). Debido a la naturaleza interconectada y global de Twitter, investigaciones han demostrado que información e influencia se propagan con más fuerza a través de Internet que a través de los canales tradicionales (Li et al., 2013). Esto significa que información se puede diseminar a través de una audiencia más amplia y exigir más influencia. Por lo tanto, Twitter no es una red social típica; sus características topológicas lo hacen más parecido a un medio de transmisión (Cha, Benevenuto, Haddadi, & Gummadi, 2012).

Además, hay dos características únicas que se utilizan con frecuencia en Twitter: el hashtag y el retweet. Un hashtag se crea cuando alguien pone una palabra o frase clave detrás del símbolo #. Esto asocia una "etiqueta" rastreable con el tweet. Si alguien quiere encontrar rápidamente información relacionada con un tema específico en Twitter, puede poner, por ejemplo, #Google en la barra de búsqueda. Cada tweet que contenga este hashtag específico aparecerá en los resultados de la búsqueda. Además, los hashtags que se utilizan con más frecuencia se llaman "trending". Estos populares hashtags se muestran en la página principal de cada usuario de Twitter. Por lo tanto, en el momento en que alguien se conecta a su cuenta, está al tanto de cualquier tema que sea popular en Twitter. Esto permite a los individuos que no tienen una relación explícita en Twitter intercambiar información. También significa que la cantidad de acceso a la información que un usuario de Twitter tiene no está restringida a los individuos que están siguiendo formalmente (Zhang et al., 2017).

Un retweet es cuando alguien publica en su propia cuenta un tweet que tiene un autor original diferente. Un retweet mantiene el formato y la información original, pero le da a la persona que realiza el retweet la oportunidad de añadir sus propios pensamientos u opiniones. Un retweet es frecuentemente interpretado como una aprobación de la información. Por lo tanto, el número de veces que un tweet ha sido "retweeted", es un indicación del nivel de influencia que el autor original ejerce en la plataforma (Sprenger, Tumasjan, Sandner, & Welpe, 2014). También asegura que la información en el tweet original sea ahora visible para su propia red de seguidores y difunde la información.

2.1.6. Cómo La Información Se Disemina en Twitter

La propagación de información en Twitter se ha distinguido por la rapidez y la difusión global de la información. Esto es debido al hecho de que la transmisión de información a través del boca a boca, y desde medios de comunicación tradicionales coexisten en Twitter. Esta sección va a mostrar la importancia de cada una, y explicar su papel en la diseminación de información. Para empezar, se distingue entre tres tipos de difusores de información.

Medios de comunicación:

El primer difusor de información son los medios de comunicación que consisten de fuentes de información tradicionales como la BBC, o CNN. Diferente a los usuarios individuos, los medios tradicionales no utilizan Twitter por motivos sociales, o para compartir opciones e intercambiar ideas. En su lugar, los medios tradicionales aprovechan el ciclo de vida corto y rápido para tener el mayor impacto con noticias de última hora a todos sus seguidores. Por ejemplo, la cuenta de Twitter de “CNN Breaking News”, solo se centra a publicar noticias de gran importancia y generalmente publica un tweet cada dos horas. Este papel es destacado por cómo las cuentas de los medios tradicionales se relacionan con otros usuarios. Cuando una cuenta de una red social gana un seguidor, sólo ocasionalmente ese seguidor sigue a su vez a la otra cuenta de forma recíproca. De hecho, una fuente tradicional de medios de comunicación sólo responde a un seguidor el 1,84% del tiempo (Cha, Benevenuto, Haddadi y Gummadi, 2012). El patrón de publicación de los medios de comunicación tradicionales también proporciona información clave sobre su papel como propagador de información. La cuenta media de un medio de comunicación publica 78,4 tweets a la semana, que es la más alta de cualquier grupo en Twitter. (Cha, Benevenuto, Haddadi y Gummadi, 2012). Sin embargo, debido a la cantidad limitada de las cuentas de los medios de comunicación de masas en comparación con el uso de los tipos de usuario, los tweets de las cuentas de los medios de comunicación de masas sólo representan el 1% de todos los tweets enviados durante una semana. Sin embargo, el alto grado de conectividad de las cuentas de los medios de comunicación las hace vitales en la cadena de difusión de noticias. De hecho, un sistema que fue diseñado para recrear la difusión de información en Twitter estimó que un Tweet publicado por una cuenta de medios tradicionales sería capaz de llegar al menos al 70% de los usuarios de Twitter (Cha et al., 2012).

Figuras públicas:

El segundo difusor principal de información en Twitter son las figuras públicas. Las figuras públicas pueden definirse como usuarios de Twitter que son bien conocidos en la esfera pública. Por ejemplo, las celebridades, los políticos y los individuos que son influyentes dentro de su campo dado caerían en esta categoría. Para decirlo numéricamente, una figura pública tiene al menos 100.000 seguidores en Twitter. A

diferencia de las cuentas de los medios de comunicación, las figuras públicas toman un papel más activo en la construcción de su red social y recíprocamente le siguen el 46,4% del tiempo (Cha et al., 2012). El papel de esta categoría es la difusión de información en Twitter se basa en su motivación para tener una presencia en Twitter en primer lugar, utilizan la plataforma para expresar sus opiniones, influir en los demás, y atraer a más seguidores. Mientras que los medios de comunicación realizan un trabajo exhaustivo de difusión de información relacionada con titulares internacionales y noticias de última hora, las figuras públicas pueden utilizar su influencia para llevar temas especializados a la primera línea de Twitter e influir en la opinión pública. Del mismo modo, pueden ampliar el alcance de los medios de comunicación tradicionales en un 25% twitteando sobre su opinión o sesgo a favor de un acontecimiento digno de una noticia en particular (Cha et al., 2012). Las figuras públicas son también prolíficas generadoras de contenido original. Dado que la mayoría de los usuarios de Twitter consumen en lugar de generar contenido, la influencia de la información es mayor en el contenido original (Gussenhoven, 2014).

Usuarios ordinarios:

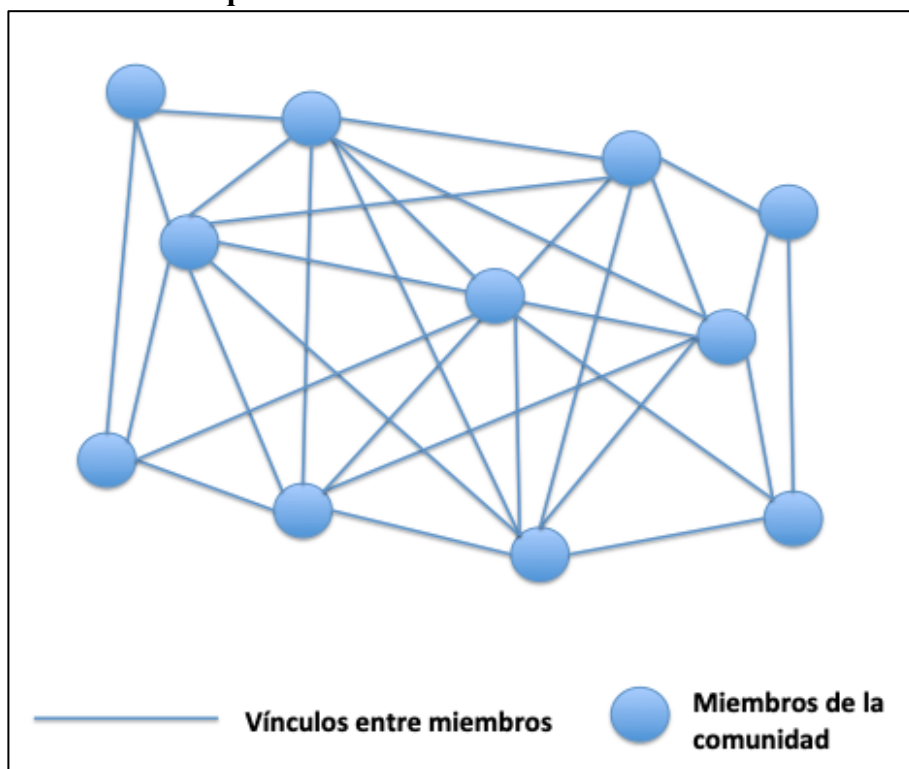
El último grupo principal de difusores de información en Twitter son los usuarios comunes. Un usuario ordinario tiene menos de 200 seguidores, y en promedio corresponderá a un seguimiento el 40% del tiempo (Cha et al., 2012). Individualmente, un usuario ordinario no está bien conectado, sin embargo, la tendencia de los usuarios a seguir a las mismas figuras públicas o cuentas de medios tradicionales los mantiene bien informados. Aunque la influencia relativa de cada usuario individual ordinario es baja, constituyen el 98,6% de todas las cuentas de Twitter, y por lo tanto tienen la mayor presencia en Twitter. El papel único de los usuarios ordinarios no puede ser subestimado. Debido a su gran número, los usuarios ordinarios son capaces de crear un efecto de "incendio forestal" cuando un interés colectivo en un tema específico puede causar la rápida y amplia difusión de la información. Esto es especialmente cierto en lo que respecta a los temas que los medios tradicionales no cubren. De hecho, algunos expertos equiparan la difusión de información entre los usuarios ordinarios de Twitter a la difusión de rumores, o boca a boca. Es posible que cada usuario sólo difunda información dentro de una red pequeña o limitada, pero el gran volumen de usuarios ordinarios garantiza una amplia difusión del tema.

2.1.7. Twitter Financiero

La sección previa se trata de cómo la información se disemina en Twitter, pero no se enfoca en las dinámicas de las comunidades específicas. Una característica importante de Twitter es que los usuarios pueden formar parte simultáneamente de numerosas comunidades de Twitter. En la sección anterior se presentaba la estructura general de la difusión de la información popular a través de los canales de información más utilizados. La información popular puede definirse como temas que suscitan un interés generalizado y no se consideran temas de nicho que requieran conocimientos específicos para poder comprender o expresar una opinión. En contraste, esta sección hablará sobre la dinámica de una sola comunidad en particular: Twitter financiero. Cualquier comunidad en Twitter se define como un conjunto de nodos que interactúan entre sí con mayor frecuencia que con los nodos fuera del grupo (Yang et al., 2015). Además de esta definición, Twitter financiero es un grupo de usuarios relevantes de Twitter cuyos intereses personales están alineados con los mercados financieros y están incorporando la información que encuentran en la plataforma en las decisiones financieras (Yang et al., 2015). La relación de Twitter con la comunidad financiera está bien documentada. Una encuesta realizada por Thomas Reuters reveló que la mayoría de los profesionales de la industria financiera, así como los clientes de la industria, utilizan las redes sociales como Twitter por razones profesionales. Desde la perspectiva de la firma, un estudio revela que el 77% de las empresas de Fortune 500 tienen una cuenta de Twitter corporativa frecuentemente activa, y que el 94% de las empresas de la industria de los Bancos Comerciales tienen una presencia corporativa en Twitter (Barnes y Wright 2013).

Es importante mencionar varias de las características que definen a esta comunidad específica de Twitter. De los numerosos patrones de redes sociales que pueden vincular a los usuarios de una comunidad específica, el Twitter financiero más se asemeja a la estructura de la comunidad conocida como el fenómeno del *small-world*. El modelo de *small-world* se caracteriza por muchos vínculos entre conocidos que hacen que la comunidad esté estrechamente entrelazada y conecta a individuos que de otro modo estarían muy alejados unos de otros (Yang et al., 2015). En el gráfico 2 en la página siguiente se puede ver una réplica a pequeña escala de este modelo.

Graficó 2 Pequeña Estructura Mundial de Twitter Financiero



Fuente: adaptado de (Yang et al., 2015)

Así como los participantes de Twitter pueden ser clasificados en 3 grupos de usuarios distintos, el Twitter financiero tiene un desglose de usuarios similar que categoriza a los usuarios en base a sus patrones de reciprocidad y nivel general de influencia en la red. Para empezar, también hay una notable presencia de fuentes de medios tradicionales en la comunidad financiera de Twitter. Sin embargo, todos ellos están orientados a las finanzas e incluyen proveedores de noticias bien conocidos como: Bloomberg, Forbes y CNNMoney. Twitter financiero también tiene algunos participantes que se definirían como figuras públicas. Por ejemplo, la figura pública más seguida en la comunidad es actualmente Josh Brown, que es el director ejecutivo de Ritholtz Wealth Management. Actualmente tiene más de un millón de seguidores, y es conocido por compartir sus opiniones sobre la política y la economía (MarketWatch, 2018). Por último, la mayor parte del Twitter financiero consiste en usuarios ordinarios que no ejercen mucho poder en la red, pero que son importantes para la cohesión general de la comunidad. Además de usar el hashtag, esta comunidad también utiliza con frecuencia una variante de un hashtag conocido como el cashtag. Escribir el ticker de una acción o índice detrás de su respectiva moneda se crea un cashtag. Por ejemplo, si se escribe \$TSLA en el motor de búsqueda de Twitter, aparecerán todos los Tweets relacionados con la empresa Tesla.

2.1.8. Tiempo de Reacción del Mercado

Para destacar la velocidad y el alcance que puede alcanzar la información difundida en Twitter, este artículo demuestra los efectos en tiempo real que un Tweet falso tiene en el mercado. La cuenta de Twitter de Associated Press (AP) sería clasificada como una cuenta tradicional de medios de comunicación porque utiliza su plataforma para informar a sus seguidores sobre noticias de última hora. Además, la cuenta de AP tiene 106 millones de seguidores, pero sólo está siguiendo 38 cuentas. El 23 de abril de 2013 a la 1:07pm EST, los hackers sirios obtuvieron acceso a la cuenta de Twitter de AP y publicaron el siguiente mensaje falso:

"Noticia: Dos explosiones en la Casa Blanca y Barack Obama herido"

Gráfico 3 Precio de DJIA



Fuente: (Google Finanzas, 2013)

Un minuto después, a la 1:08pm EST el Dow Jones Industrial Average (DJIA) comenzó a caer en picado como se puede ver en el gráfico 3. Bajó 150 puntos, de 14697.15 a 14548.58, y sólo dejó de caer a la 1:10pm EST, cuando empezaron a circular noticias de que el Tweet era falso. En tres minutos, el DJIA volvió a 14690. Sin embargo, durante la caída temporal, se eliminaron 136 mil millones de dólares en valor de mercado (The New York Times, 2013). La transferencia de información entre Twitter y el mercado de valores se realizó en cuestión de minutos. Amplificando

los efectos fue la reacción agresiva de los sistemas de comercio de alta frecuencia que están programados para operar basados en palabras clave casi instantáneamente. Si bien este ejemplo es extremo, demuestra el flujo directo de información de Twitter a los mercados financieros.

2.2 LA BOLSA

2.2.1. Predicción bursátil

En 1970, Eugene Fama publicó su influyente trabajo sobre la eficiencia del mercado, y la hipótesis de mercado eficiente (HME) se convirtió en la teoría dominante del mercado durante varias décadas. En un mercado eficiente, las cotizaciones bursátiles deben incorporar instantáneamente toda la información relevante tan pronto como estén disponibles (Fama, 1970). Como resultado, las acciones nunca se infravaloran o sobrevaloran, y los inversores no pueden "vencer al mercado" realizando arbitrajes, o utilizando técnicas de *market timing*. Estrechamente relacionado con la HME es la hipótesis del *Random Walk Hypothesis* (RWH) que se basa en los inquilinos fundamentales del HME, y establece que las cotizaciones bursátiles son impulsadas por nueva información, como las noticias. Dado que las noticias son impredecibles, los precios de las acciones subirán y bajarán en un patrón de *Random Walk* impredecible y no pueden predecirse con una precisión superior al 50% (Bollen, Mao y Zeng, 2011). A pesar de tener una influencia duradera en el campo de la economía financiera, la HME ha sido criticada tanto empírica como teóricamente. Para que un mercado funcione eficientemente, la HME supone: (1) la mayoría de los inversores son racionales (inversores racionales) y son capaces de valorar las acciones por su valor fundamental, (2) cuando algunos inversores irracionales *noise traders* realizan operaciones irracionales, sus decisiones son aleatorias y no están correlacionadas y se anulan mutuamente (3) las acciones de los *noise traders* serán contrarrestadas por inversores racionales hasta que se alcance el equilibrio del mercado (Fama, 1970). Los críticos se han apresurado a destacar que varios de los supuestos clave de la HME no se mantienen en el mercado real y tampoco explican anomalías bien documentadas en los movimientos de las cotizaciones bursátiles (Nofer & Hinz, 2015). Por ejemplo, los inversores racionales no pueden impedir que los *noise traders* puedan apartar los precios de su valor fundamental cuando los *noise traders* de grandes grupos siguen a otros *noise traders* y causan caídas o subidas predecibles en los precios que crean

oportunidades para el arbitraje (Nofer & Hinz, 2015). Además, si los principios de la HME se aplicaran perfectamente en el mercado de valores, Twitter no sería capaz de proporcionar una visión predictiva del comportamiento de los mercados de valores de forma coherente y precisa. De hecho, los sitios de medios sociales como Twitter deberían teóricamente tener un impacto positivo en la eficiencia del mercado porque difunden rápidamente la información, lo que debería significar que la nueva información se refleja plenamente en los precios y que no hay oportunidad de arbitraje (Ali, 2018). La siguiente sección explicará esta paradoja proporcionando una teoría financiera alternativa a la HME, y utilizándola para explicar por qué Twitter es una herramienta ideal para la predicción del mercado de valores.

2.2.2. Finanzas Conductuales

La Teoría de Finanzas de Comportamiento (BFT) surgió en la década de 1990, y sostiene que el mercado de valores está impulsado principalmente por la psicología de los inversores. En particular, los investigadores de Finanzas Conductuales intentan explicar la relación entre las emociones de los inversores y su comportamiento comercial (Nofer & Hinz, 2015). En términos sencillos, la idea subyacente del BFT es que la mayoría de los inversores no son racionales. En lugar de basar la decisión financiera en un razonamiento puramente fáctico o fundamental, los inversores toman decisiones basadas en lo que piensan, y lo que pensamos se deriva de cómo sentimos. Estos sentimientos, que constituyen la base de nuestras acciones, están fuertemente influenciados por las interacciones con los demás (Nofsinger & Nofsinger, 2010). En los casos en que las emociones o los estados de ánimo están correlacionados entre grandes grupos de inversores, el estado de ánimo colectivo afectará a las decisiones financieras y puede dar lugar a fenómenos que afectan a todo el mercado (Nofsinger & Nofsinger, 2010). La correlación entre el estado de ánimo social, también conocido como sentimiento, y la actividad del mercado se ha observado en numerosas anomalías relacionadas con el estado de ánimo que contradicen las premisas de la HME, pero que apoyan las ideas centrales de la BFT. Por ejemplo, los eventos deportivos influyen en el estado de ánimo de la gente, y un estudio realizado por Edmans et al. (2007) reveló que los mercados bursátiles nacionales reaccionan negativamente a las pérdidas de los equipos nacionales de fútbol en competiciones internacionales como la Copa Mundial. Del mismo modo, existe la bien documentada "anomalía de ahorro de luz diurna". Se ha notado consistentemente que el lunes

después del horario de verano tiene menos devoluciones de acciones que los lunes regulares durante todo el año. Se cree que esto se debe a que los inversores se encuentran en un estado de ánimo menos optimista y adverso al riesgo debido a la falta de sueño (Nofer & Hinz, 2015). Ambas anomalías están en contradicción directa con la HME porque implican que el comportamiento del mercado puede predecirse basándose en el comportamiento irracional o influenciado por el estado de ánimo de los inversores.

En el contexto de Twitter, la predicción bursátil podría ser factible porque los medios de comunicación social desempeñan un papel importante a la hora de influir y, a veces, liderar el estado de ánimo y la respuesta emocional del público a cualquier nueva información o noticia que afecte al mercado (Ali, 2018). Este artículo ya ha discutido la eficacia de Twitter en la agregación y difusión de información entre diversos grupos de usuarios en tiempo real. Además, este trabajo ha puesto de relieve que existe una cantidad sustancial de usuarios de comunidades, como Twitter financiero, que se inclinan a incorporar la información que encuentran en sus acciones financieras. El BFT enfatiza que además del intercambio de información, también hay un importante intercambio de emociones. Como se ha mencionado anteriormente, Twitter cuenta con canales de difusión de información eficientes y de gran alcance que comparten opiniones en tiempo real. Por lo tanto, a medida que se difunde el ambiente social generalizado a través de Twitter, una gran cantidad de *noise traders* podrían reaccionar simultáneamente, lo que provocaría cambios en los precios de mercado. Las implicaciones de las Finanzas Conductuales cuando se aplica a Twitter son las siguientes: (1) El humor social determina los tipos de decisiones que toman los consumidores (2) Debido a la naturaleza eficiente y sentimental de las transacciones bursátiles, el propio mercado bursátil puede utilizarse para medir el humor social (3) Las tendencias de Twitter pueden ayudar a pronosticar la actividad futura del mercado bursátil (Nofsinger & Nofsinger, 2010).

2.3 MARCO DE RELACION ENTRE TWITTER Y LA BOLSA

Hasta ahora, este artículo ha cubierto algunas de las características únicas e importantes de Twitter. En particular, que se trata de un sitio de microblogging muy popular que permite a los usuarios proporcionar actualizaciones sobre sus pensamientos y sentimientos, compartir noticias e intercambiar ideas. En la sección anterior se demostró el eficiente y amplio flujo de información a través de Twitter, y se explicó el importante papel de cada tipo de usuario en la propagación de la información. También mostró la rapidez con la que la información podía ser incorporada al mercado de valores. Combinados, estos dos factores hacen de Twitter una fuente ideal para obtener información sobre el sentimiento popular y de los consumidores hacia las empresas, los productos y las tendencias (Tafti, Zotti y Jank, 2016). El valor de esta información con respecto a los pronósticos del mercado de valores se ha enfocado de dos maneras diferentes: indicadores relacionados con el volumen de tweets e indicadores relacionados con el sentimiento de tweets.

2.3.1 Pronóstico Relacionado con el Volumen de Tweets

De los dos predictores de mercado, un predictor basado en el volumen requiere métodos de recopilación de datos menos intensivos que los indicadores basados en el sentimiento porque es agnóstico al sentimiento expresado en el tweet. En cambio, se basa en la lógica de que un aumento en la charla en Twitter sobre cualquier tema, podría ser una indicación de un mayor interés público en ese tema. En el contexto del mercado de valores, un indicador basado en el volumen busca picos inusuales en Twitter para tratar de predecir la actividad de los mercados financieros en tiempo real. La teoría es plausible por varias razones. En primer lugar, el aumento de la información accesible en línea en sitios como Twitter tiene un efecto sobre el comercio, porque las noticias afectan a los mercados financieros (Gussenhoven, 2014). Adicionalmente, la capacidad del mercado de valores para incorporar rápidamente la información publicada en Twitter ha sido notada en incidentes tales como los efectos que el falso tweet detallado anteriormente tuvo en el DJIA.

2.3.2. Indicador de Volumen

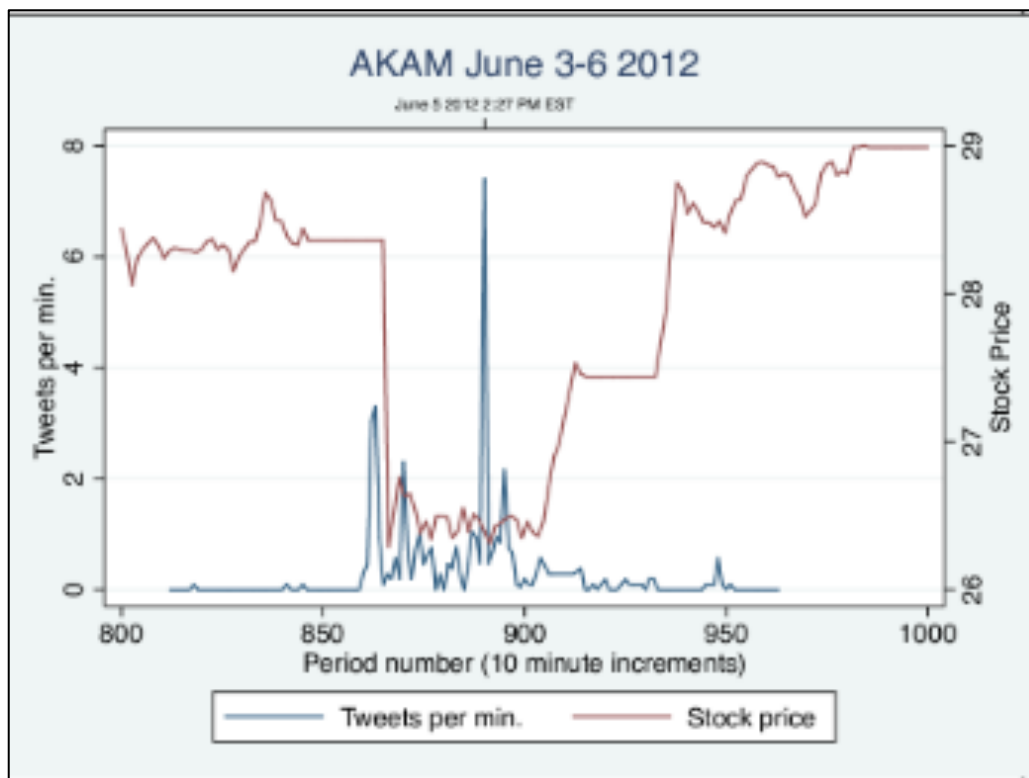
Con el fin de determinar si los indicadores basados en el volumen en Twitter podrían proporcionar información fiable sobre los movimientos del mercado, un equipo de investigadores analizó la relación entre las conversaciones en Twitter y el volumen de operaciones de NASDAQ 100 empresas durante 193 días. Contaron el número de Tweets que mencionaban a la empresa en intervalos de 10 minutos, y recogieron el precio y el volumen de operaciones de cada empresa en las ventanas correspondientes de 10 minutos para comprender la respuesta intradiaria del mercado en relación con las noticias difundidas en Twitter (Tafti et al., 2016). El análisis en profundidad de los tweets en tiempo real fue factible debido al hecho de que Twitter ha puesto a disposición del público una corriente de datos gratuitos conocida como "manguera de jardín" para su análisis (Tafti, Zotti, & Jank, 2016). Para cada empresa, se estableció un número de referencia de tweets sobre la empresa por segundo, de modo que se pudiera detectar un aumento de tweets en un segundo después de su aparición. Para ilustrar cuán efectivo fue este análisis en mostrar una correlación directa entre las conversaciones de Twitter y los movimientos del mercado, los resultados de un incidente específico que ocurrió durante el experimento realizado por Tafti et al. 2016 serán resumidos a continuación.

2.3.3. Akamai Technologies

Akamai Technologies (AKAM): El 4 de junio de 2012, la popular compañía de streaming Netflix anunció que desarrollaría su propia red de entrega de contenidos. Hasta ese momento, Netflix había dependido de la empresa Akamai para sus servicios de entrega de contenido. La noticia tuvo un impacto repentino en el precio y el volumen de comercio de las acciones de Akamai y el precio de las acciones cayó más de un 10%, antes de recuperarse por completo en varias horas. Curiosamente, un pico en la actividad de Twitter haciendo referencia a Akamai en Tweets ocurrió aproximadamente 10 minutos antes de que se produjera cualquier cambio en el volumen de operaciones o en el precio de las acciones de Akamai a medida que la información se difundía a través de varios canales de Twitter. En menos de una hora, hay otro pico en Twitter que menciona a Akamai, y el mercado comienza lentamente a recuperar los precios de la empresa y los volúmenes de operaciones hasta los niveles de preanuncio.

El Gráfico 4 muestra el precio de las acciones de Akamai junto con los tweets por minuto sobre la empresa. En el gráfico se puede ver el fuerte aumento del volumen de los tweets, como lo indica la línea azul, junto con la caída del precio de las acciones, como lo indica la línea roja.

Gráfico 4 Acciones de Akamai Durante la Semana 03-06-2012



Fuente: (Tafti et al., 2016)

Este estudio de caso revela varias facetas sobre cómo los indicadores relacionados con el volumen podrían ayudar a anticipar los movimientos del mercado. Para empezar, sugiere que el aumento de la actividad de Twitter es muy pronunciado en un corto período de tiempo, y ocurre rápidamente después de que se anuncia nueva información o nueva información relacionada con una empresa. Mientras tanto, el mercado requiere más tiempo para reaccionar a las noticias e incorporar la información antes de la autocorrección. Esto sugeriría que el monitoreo de Twitter para detectar picos en las conversaciones sobre empresas específicas podría proporcionar a los individuos información ventajosa sobre la actividad futura del mercado (Tafti et al., 2016).

2.3.4. Pronosticador Relacionado con Sentimientos

El método de análisis de información que se utiliza con mayor frecuencia para encontrar predictores de mercado dentro del flujo constante de información en Twitter es el análisis de sentimientos. La investigación en el campo de la economía del comportamiento ha demostrado que las emociones pueden afectar profundamente el comportamiento individual y la toma de decisiones (Bollen, Mao y Zeng, 2011). Los estudios en finanzas conductuales han demostrado que las decisiones financieras están influenciadas significativamente por la emoción y el estado de ánimo (Nofsinger & Nofsinger, 2010). Por lo tanto, hay razones para creer que el sentimiento público podría ser una fuerza motriz de los movimientos en el mercado de valores. El sentimiento en sí mismo se define como "una opinión que se mantiene o se expresa" (Oxford Dictionary 2019). El análisis de sentimientos ofrece a los investigadores la capacidad de procesar un conjunto de resultados de búsqueda para un elemento determinado, generando una lista de atributos del producto y agregando opiniones sobre cada uno de ellos (Pang y Lee, 2008). En otras palabras, el análisis de sentimientos es capaz de procesar grandes cantidades de datos y clasificar los sentimientos predominantes expresados en los datos hacia un tema específico. Se ha utilizado cada vez más para extraer indicadores que predicen cambios en la actividad comercial y económica porque la investigación psicológica muestra que la información y las emociones juegan un papel importante en el proceso de toma de decisiones (Bollen et al., 2011). La naturaleza colaborativa y la estructura de la comunidad de Twitter se han detallado en secciones anteriores. Se ha observado que estos atributos hacen de la plataforma un lugar ideal para que los usuarios compartan sus pensamientos (sentimientos) sobre cualquier tema de una manera ubicua y de gran alcance. Este intercambio de información se conoce oficialmente como "boca a boca" electrónico (eWOM), y es esencial para la formación y el seguimiento de los sentimientos de los consumidores (Bollen et al., 2011). El boca a boca tradicional (WOM) es anterior a Internet y consistía en que los consumidores compartían opiniones, actitudes y reacciones hacia las empresas o productos con otras personas. Con el auge de los sitios de medios sociales como Twitter, eWOM se ha extendido aún más. Las investigaciones han demostrado que es más probable que los consumidores confíen en las percepciones y opiniones de las personas que están dentro de su red social para basar sus decisiones de compra, más que en las aportaciones de las fuentes de marketing tradicionales (Jansen, Zhang, Sobel y

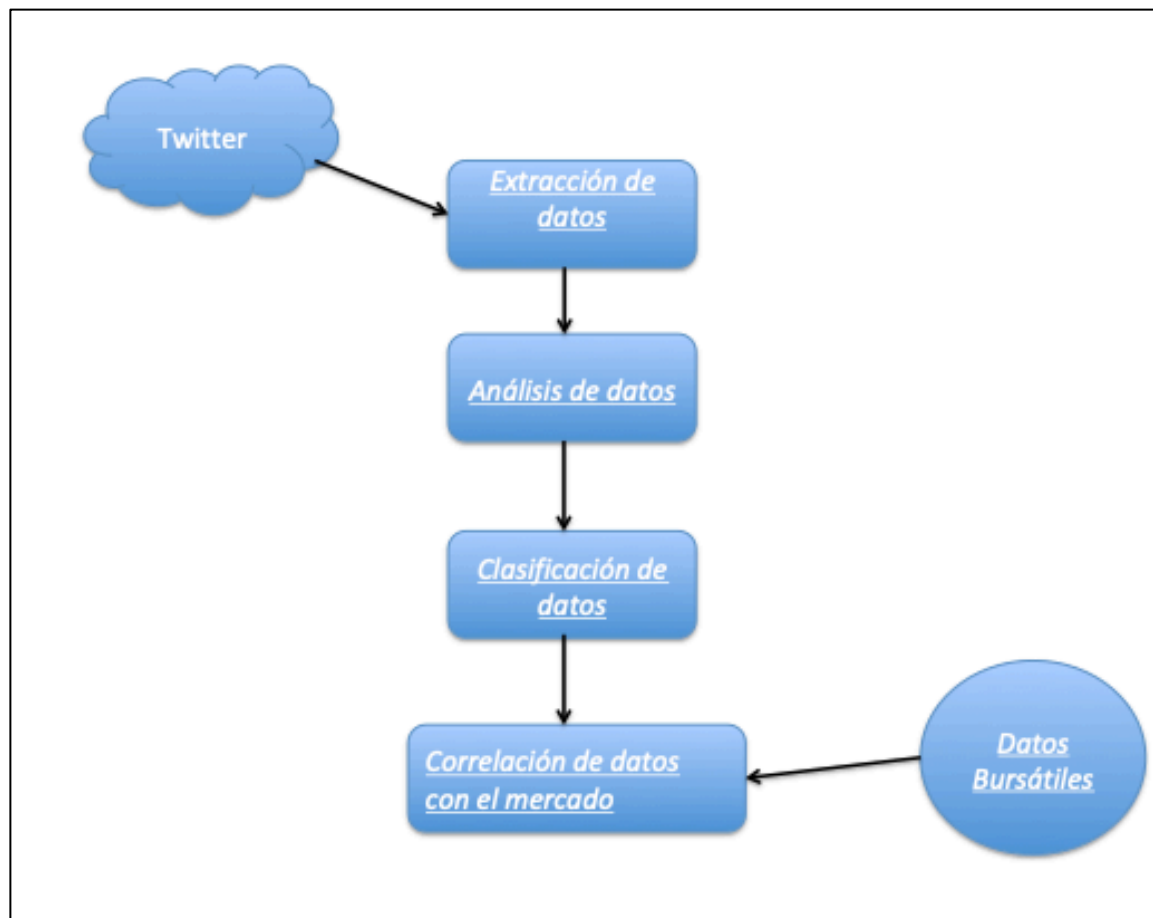
Chowdury, 2017). Los microblogs como Twitter son un flujo de datos ideal para el análisis de sentimientos porque proporcionan evaluaciones tempranas instantáneas, confiables y escalables del estado de ánimo del público.

2.3.5. Clasificación de Sentimientos

Numerosos equipos de investigación han tratado de demostrar que el análisis de sentimientos de Twitter puede utilizarse para predecir los movimientos del mercado de valores. Esta es una pregunta amplia que puede ser abordada desde muchas perspectivas diferentes. Mientras que algunos equipos se han centrado en la correlación entre el sentimiento de Twitter y una empresa específica, otros se han centrado en cómo el sentimiento del público puede afectar al mercado en su conjunto. Sin embargo, cada equipo de investigación debe deducir primero cómo extraer, analizar y categorizar el sentimiento de Twitter. Esta sección proporcionará una visión general de las diversas técnicas y tecnologías utilizadas para realizar el análisis de sentimientos.

El Gráfico 5 muestra una visión general paso a paso de las diversas fases del análisis de sentimientos que también se tratarán en esta sección.

Gráfico 5 Proceso de Análisis de Sentimientos



Fuente: adaptado de (Bollen et al., 2011)

Extracción de datos:

El primer paso de los datos de sentimientos es extraer los datos relevantes para ser analizados. En el caso de Twitter, se puede acceder a la información a través de la API de Twitter. API son las siglas de *Application Programming Interface* (API). Es una herramienta que facilita la interacción entre los programas informáticos y los servicios web. Por ejemplo, los investigadores pueden utilizar una aplicación que solicita datos específicos de Twitter. La API devolverá los datos solicitados en base a solicitudes específicas en el formato solicitado por la aplicación (Li, Zhou y Liu, 2016). La API de *streaming* de Twitter se utiliza específicamente para proporcionar datos en tiempo real desde Twitter. Las funciones de flujo público y de flujo del sitio son particularmente útiles para el análisis de sentimientos porque permiten que la

aplicación monitoree los datos públicos y las fuentes de Twitter de los usuarios simultáneamente (Li et al., 2016).

Análisis de datos:

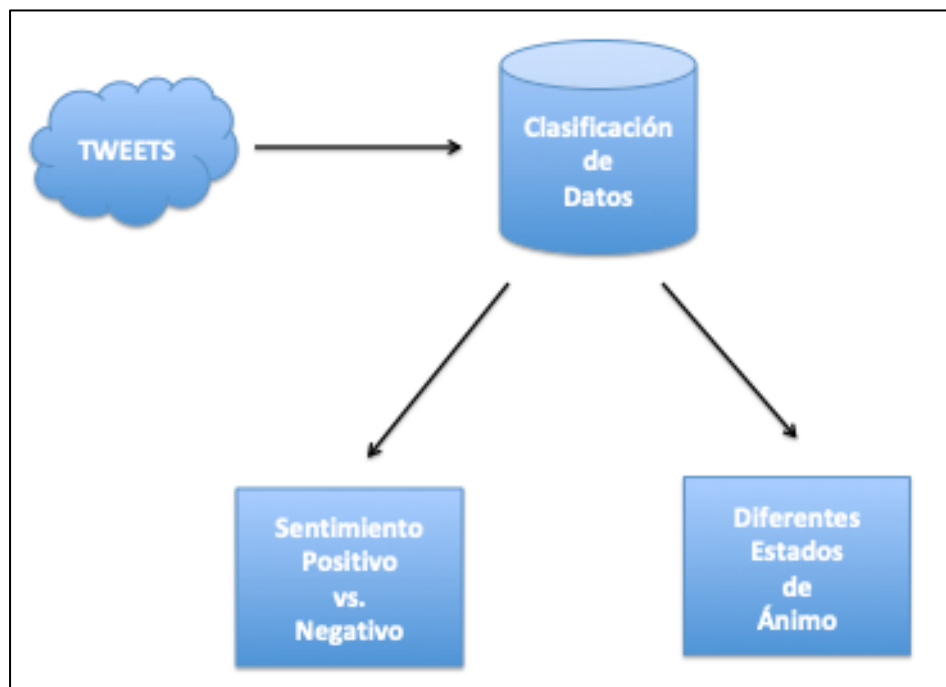
Los objetivos del análisis de sentimientos son extraer las señales emocionales expresadas en Tweets y cuantificarlas. Hacerlo manualmente requeriría mucha mano de obra, consumiría mucho tiempo y estaría sujeto a la parcialidad humana. Afortunadamente, el desarrollo de las técnicas de minería de textos y la lingüística computacional han hecho posible el análisis automatizado de sentimientos (He, Guo, Shen y Akula, 2016). El primer paso en el análisis de los datos es la clasificación de los textos mediante la extracción de indicadores de sentimientos. Consiste en una clasificación de textos supervisada y no supervisada. Algunos estudios utilizaron aprendizaje automático supervisado, como el algoritmo de clasificación de texto de Naïve Bayes (Sprenger, Tumasjan, Sandner y Welp, 2014). Las técnicas supervisadas requieren que los investigadores primero clasifiquen manualmente una muestra de tweets, y luego "entrenen" el algoritmo para clasificarlos de forma independiente. Otros investigadores han utilizado un enfoque no supervisado que se basa en algoritmos que identifican palabras y frases clave de un léxico predeterminado para clasificar los sentimientos (Oliveira, Cortez y Areal, 2017).

Clasificación de datos:

Después de analizar el contenido de los tweets, el siguiente paso es clasificar el sentimiento. Basándose en la técnica de análisis de sentimientos que se emplea y en el objetivo de la investigación, los investigadores pueden clasificar los datos según varios parámetros diferentes. Por ejemplo, el clasificador de sentimientos "Opinion Finder" (OF), disponible públicamente y sin supervisión, es capaz de determinar la subjetividad de los tweets. Esencialmente, esto significa que es capaz de identificar la polaridad emocional de los tweets y clasificarlos como positivos o negativos (Bollen et al., 2011). Aunque este método es útil para determinar el sentimiento general en Twitter, se limita a una clasificación bidimensional y binaria. El sentimiento humano actual es mucho más diverso y multidimensional. Por ejemplo, otro algoritmo no supervisado llamado Google Profile of Mood States (GPOMS) utiliza un léxico específico para clasificar el estado de ánimo en seis dimensiones: Calma, Alerta,

Seguro, Vital, Amable, Feliz (Bollen et al., 2011). El gráfico 6 muestra las dos formas diferentes en que se clasifica el sentimiento.

Gráfico 6 Clasificación de los Sentimientos de Tweets



Fuente adaptado de: (Bollen et al., 2011)

Las dimensiones emocionales específicas dependen de la discreción de cada equipo de investigación. Otro equipo de investigación utilizó un proceso similar para clasificar los sentimientos como: Miedo, Ira, Sorpresa, Felicidad, Tristeza, Disgusto (Li et al., 2016). La clasificación de los sentimientos como estados de ánimo específicos abre la posibilidad de un análisis más complejo y profundo entre los sentimientos y los movimientos del mercado.

Correlación de datos con el mercado:

Capturar y clasificar el sentimiento de Twitter es sólo la primera mitad del proceso de análisis de sentimientos. Para relacionar el sentimiento con el mercado y establecer el grado en que uno puede influir en el otro, también es necesario extraer los datos de mercado correspondientes. Afortunadamente, muchos sitios web acreditables como Yahoo Finance, Morning Star y MarketWatch proporcionan acceso a una amplia gama de información sobre el rendimiento pasado de las empresas y los índices. Esto no sólo incluye los rendimientos históricos de las empresas e índices, sino también

otras métricas importantes como el volumen de comercio y la volatilidad. Con todos estos datos, los investigadores pueden rastrear en tiempo real el sentimiento de Twitter hacia un índice o empresa específica, y simultáneamente monitorear cualquier cambio perceptible en el mercado. Para cuantificar el grado de correlación entre el sentimiento de Twitter y los movimientos del mercado de valores, se han adoptado diferentes enfoques para analizar la relación entre los dos conjuntos de datos. Uno de los enfoques más básicos es calcular las correlaciones de Pearson entre el sentimiento de Twitter y los datos de mercado. En primer lugar, es necesario seleccionar una variable de cada conjunto de datos. La variable independiente sería el sentimiento de Twitter, y la variable dependiente sería la métrica del mercado de valores. La correlación de Pearson establecerá si hay un nivel de asociación entre las dos variables en un día dado. Esta correlación es denotada por r , y significa que cuando hay un cambio en la variable independiente, hay un cambio correspondiente en la variable dependiente (Nisar & Yeung, 2018). Los valores de r pueden estar entre $(-1,1)$, y una puntuación de 1 indica la correlación positiva más fuerte posible, mientras que una puntuación de -1 indica la correlación negativa más fuerte posible (Nisar & Yeung, 2018). Si no existe una relación lineal entre las dos variables, no se puede utilizar la correlación de Pearson para probar la relación entre los dos conjuntos de datos. Otra herramienta estadística que se basa en la regresión lineal es el análisis de regresiones múltiples. Este método explica la relación que tiene cada variable independiente individual, sobre la variable dependiente, cuando también se toman en consideración las otras variables independientes (Nisar & Yeung, 2018). Un escenario posible en el que se utilizaría esta prueba es cuando múltiples variables independientes, tales como diferentes estados de ánimo, se están probando simultáneamente para determinar su efecto holístico sobre una variable de mercado. Los resultados de esta herramienta incluyen el valor R^2 , que indica la proporción de varianza en la variable dependiente que puede ser explicada por las variables independientes (Investopedia, 2018). Otro análisis lineal es la prueba de Causalidad de Granger. La prueba de Causalidad de Granger difiere ligeramente de las formas de análisis estadístico descritas anteriormente porque busca específicamente determinar la correlación entre las series temporales. El análisis de causalidad de Granger se basa en el supuesto de que si la variable independiente afecta a la variable dependiente, entonces los cambios en la variable independiente ocurrirán sistemáticamente antes que los cambios en la variable dependiente (Bollen et al., 2011). Es importante

recordar que mientras que el análisis estadístico puede demostrar altos niveles de correlación entre variables, la correlación en sí misma no prueba la causalidad.

La correlación de Pearson, el análisis de regresión múltiple y la causalidad de Granger son sólo formas de análisis aplicables cuando se prueban variables que tienen una relación lineal. Una relación no lineal significa que un cambio en la variable independiente no provoca un cambio correspondiente en la variable dependiente (Investopedia, 2018). Una relación no lineal podría implicar que una relación entre las dos variables es inexistente o impredecible. Sin embargo, existen métodos de análisis que pueden superar este reto, y que son capaces de tener en cuenta factores más complejos a la hora de buscar una relación predictiva entre dos variables. Esto significa que podrían descubrir una relación que antes no se hubiera detectado. Uno de estos métodos es utilizar algoritmos de aprendizaje predictivo y ver si su capacidad para predecir el rendimiento del mercado de valores mejora cuando se les da acceso a los resultados del análisis de sentimientos.

2.3.6. Análisis de Sentimientos y Ejemplos del Mercado de Valores

Numerosos equipos de investigación han utilizado el análisis de sentimientos para investigar un posible vínculo entre Twitter y el mercado de valores. El interés generalizado por este tema ha llevado a una plétora de variaciones en la forma en que los académicos han abordado y llevado a cabo sus investigaciones. Debido a la aplicación flexible del análisis de sentimientos, los parámetros emocionales específicos en los que los equipos se han centrado han sido tan básicos como los sentimientos positivos frente a los negativos. Mientras tanto, otros investigadores han buscado capturar una amplia gama de emociones, y han rastreado varios estados emocionales en Twitter a través del análisis de sentimientos. Del mismo modo, el "vínculo" con los mercados de valores se ha definido de numerosas maneras. Diversos estudios han seguido los efectos que tiene el sentimiento en todo un índice, mientras que otros estudios se han centrado en la correlación entre el sentimiento y una empresa específica. De manera similar, algunos equipos de investigación rastrearon los efectos de los sentimientos sobre los rendimientos y las variaciones en el precio de las acciones, mientras que otros buscaron determinar la influencia de los sentimientos sobre variables tales como el volumen de operaciones. Curiosamente, diferentes

estudios también han definido el tamaño de su muestra basándose en los tweets de los usuarios de Twitter en comunidades específicas. Por ejemplo, algunos estudios siguen el sentimiento de Twitter en su conjunto, mientras que otros tratan de limitar el alcance de su análisis a Twitter financiero solamente. Sin embargo, cada estudio es único y aporta nuevos conocimientos. En la sección siguiente se ofrece una visión general de la variedad de formas en que se ha investigado la correlación entre el análisis de sentimientos y los mercados bursátiles.

El Sentimiento de Twitter y el Dow Jones Industrial Average

Bollen et al (2011), realizaron el estudio posterior que se centró específicamente en la capacidad de predicción del sentimiento de Twitter. Se buscó determinar si las mediciones de los estados de ánimo colectivos derivados de los feeds de Twitter están correlacionados con el valor de cierre del Dow Jones Industrial Average (DJIA). Las herramientas de análisis de sentimientos que utilizaron fueron el *opinión finder* (OF) mencionado anteriormente y el perfil de Google de los estados de ánimo. Del 28 de febrero al 3 de noviembre de 2008, el equipo recogió el valor de cierre diario de la DJIA según lo informado por Yahoo! Finance. Además, llevaron a cabo un análisis de series cronológicas sobre el sentimiento de Twitter. No distinguían entre los tipos de usuarios de Twitter, y recogieron Tweets de todas las comunidades de Twitter para realizar el análisis de sentimientos. Utilizando ambos conjuntos de datos, realizaron una prueba de causalidad de Granger y, basándose en los resultados, pudieron rechazar la hipótesis nula de que las series temporales de sentimientos no predicen los valores del DJIA. Sin embargo, esto sólo era cierto para un estado de ánimo: la calma. La relación predictiva fue más fuerte con un desfase de tres días, con cambios de tres días en el estado de calma en Twitter prediciendo los aumentos y disminuciones de las corrientes en el DJIA. Sin embargo, los otros estados de ánimo no tuvieron una relación casual significativa con los cambios en la DJIA. Tampoco lo hizo la serie de tiempos OF que seguía el sentimiento positivo vs. negativo.

El equipo observó que, aunque la prueba de causalidad de Granger sugiere una relación predictiva entre un estado de ánimo "tranquilo" en Twitter y el DJIA, se basa en una regresión lineal. La relación real entre el estado de ánimo del público y los valores del mercado de valores no es lineal. Para tener en cuenta los aspectos no lineales de la relación y evaluar el valor de utilizar la evaluación del estado de ánimo

público en la construcción de modelos predictivos de los valores del DJIA, el equipo realizó un experimento básico utilizando una *self-organized fuzzy neural network* (SOFNN). La SOFNN es un algoritmo de aprendizaje que fue desarrollado específicamente para problemas de análisis de series de tiempo. El experimento consistió en comparar la capacidad de la SOFNN para predecir con precisión los valores del DJIA basándose en dos conjuntos de entradas. El primer conjunto sólo consistió en los últimos 3 días de valores DJIA, mientras que el segundo conjunto combinó los valores de los últimos 3 días con las entradas de la serie de sentimientos de Twitter anteriormente descrita. Los resultados indicaron que la SOFNN tenía la mayor precisión de predicción del 87,6% cuando los valores pasados se combinaban con las entradas del análisis de sentimientos "tranquilo".

La principal contribución de esta investigación al estudio de Twitter y el mercado de valores es que proporciona evidencia concreta de que el estado de ánimo del público puede ser rastreado con precisión a partir del contenido de los feeds de Twitter a gran escala, y que los cambios en dicho estado de ánimo corresponden a cambios en los valores del DJIA. Por su propia admisión, esta investigación no proporciona información sobre el mecanismo causal que explicaría la conexión entre el estado de ánimo del público y los valores del DJIA.

El Sentimiento de Twitter y Precios a Nivel de Firma

La investigación subsiguiente que se discutirá en esta sección fue realizada por (He et al., 2016). Este equipo de investigación también utilizó el análisis de sentimientos para probar el valor predictivo de Twitter en los mercados de valores. Sin embargo, en lugar de centrarse en un índice completo como el DJIA, este equipo redujo su alcance a las cotizaciones bursátiles a nivel de cada empresa. Específicamente, este estudio optó por centrarse en siete empresas de servicios financieros bien conocidas: American Express, Capital One, Citibank, HSBC, JPMorgan, MasterCard y Wells Fargo. Estas compañías fueron seleccionadas debido a su relativa popularidad en Twitter de ser mencionadas frecuentemente en los tweets. Para el componente de análisis de sentimientos de este estudio, el equipo empleó la ayuda del algoritmo de análisis de sentimientos disponible comercialmente, Lexalytic, para clasificar las frases emotivas en tweet en una escala de (-1 a +1). Después, la puntuación de cada frase se combina para discernir el sentimiento general positivo o negativo en cada

tweet. A continuación, el sentimiento promedio positivo y el promedio negativo se calculó sumando el valor absoluto de cada categoría en un intervalo diario. Los datos financieros correspondientes a la cotización de las acciones de cada empresa se recogieron de la base de datos Bloomberg. A continuación, se realizó una regresión lineal para ver si el sentimiento positivo o el sentimiento negativo mostraban un efecto estadísticamente significativo sobre las cotizaciones bursátiles. Se analizaron tanto los sentimientos contemporáneos como los rezagados hacia las siete compañías en cuanto a su capacidad para predecir los cambios en los precios de las acciones de la compañía. En última instancia, se descubrió que sólo el sentimiento negativo de un día de retraso de Twitter tuvo un efecto estadísticamente y económicamente significativo sobre el precio de las acciones. De hecho, el equipo pudo determinar que si el valor absoluto del sentimiento negativo aumenta en un 1%, al día siguiente el precio de la acción disminuirá en un 3,72%. Ni el sentimiento positivo contemporáneo ni el rezagado de Twitter tuvo un efecto positivo significativo desde el punto de vista estadístico o económico en la cotización de las acciones.

Los resultados de este estudio proporcionan más pruebas de que el sentimiento de Twitter está significativamente relacionado con los precios futuros de las acciones de una empresa. Este equipo contribuyó al área de estudio al ser el primer grupo en analizar específicamente el poder predictivo de Twitter a nivel de cada empresa. Específicamente, fueron capaces de proporcionar pruebas concluyentes de que el sentimiento negativo de Twitter predice significativamente el precio futuro de las acciones de las empresas. Aunque este resultado contradice el hallazgo de (Bollen et al., 2011), es importante señalar que ambos estudios se realizaron en diferentes niveles de análisis (nivel de empresa vs. nivel de índice).

Twitter alcista y precios de las acciones

El último estudio que se discutirá en este artículo fue realizado por (Sprenger et al., 2014). Este estudio en particular exploró un nicho interesante dentro del campo más amplio del análisis de sentimientos y Twitter. A diferencia de estudios anteriores, este equipo de investigadores fue más selectivo en la forma en que eligieron qué tweets utilizar en su análisis de sentimientos. En lugar de incluir todos los tweets que mencionaban una empresa o un índice en particular, sólo tenían en cuenta los tweets que también incluían temas bursátiles. Es decir, ignoraron los tweets de la población

general de Twitter, y trataron de identificar el sentimiento que se produce en las comunidades de Twitter que se dedican a temas financieros. Por lo tanto, este estudio está libre de cualquier ruido que no esté relacionado con el mercado de valores. La hipótesis que este equipo de investigadores trató de explorar con su análisis de sentimientos fue si el aumento de la tendencia alcista de los micro blogs de valores se asocia con mayores rendimientos. Para recoger el tamaño de la muestra de los tweets relevantes, el equipo sólo recogió los tweets que habían sido etiquetados con el símbolo de una empresa en el S&P 100. Esto equivale a un tamaño de muestra de 249.533 tweets en un período de seis meses. Con el fin de determinar el valor al alza de cada tweet individual, se entrenó un algoritmo de clasificación de texto para distinguir entre señales de compra, retención, o venta. La relación entre las señales de compra y venta se utilizó para determinar la tendencia alcista de cada una de las empresas. Los datos financieros correspondientes se descargaron a intervalos diarios desde el DataStream de Thomas Reuters. Se realizó una correlación por pares entre los datos de stock y tweet, y los resultados de la regresión apoyaron la hipótesis de que existe una relación estadísticamente significativa entre la tendencia al alza y la subida de las cotizaciones bursátiles.

Este estudio proporcionó información interesante sobre la forma en que el sentimiento encontrado en las comunidades financieras en Twitter podría correlacionarse con las cotizaciones bursátiles. Además de detectar una relación entre el sentimiento al alza en Twitter y los precios de las acciones, este estudio destaca la influencia que la comunidad financiera en Twitter tiene en la toma de decisiones de inversión y, en consecuencia, en los movimientos del mercado de valores.

2.4. ESTRATEGIAS DE INVERSION BASADAS EN EL ANÁLISIS DE TWITTER

2.4.1. La formulación de las estrategias

La sección anterior de este trabajo demostró claramente el poder predictivo que tiene Twitter con respecto al mercado de valores. Aunque la manera en que se llevó a cabo cada estudio fue diferente, llegaron a conclusiones similares: El sentimiento de Twitter puede predecir de forma precisa y consistente los cambios en el mercado de valores. La modularidad de los estudios y sus alentadores resultados no han pasado

desapercibidos. De hecho, han fomentado un nuevo conjunto de preguntas y nuevas vías de investigación. En particular, los investigadores han tratado de determinar si pueden basarse en las ideas obtenidas del análisis de sentimientos de Twitter para formular estrategias de trading que tengan mejores resultados que las estrategias de trading tradicionales. La lógica detrás de la formulación de la estrategia es simple: si el sentimiento de Twitter está mostrando un fuerte impulso en una dirección específica, es probable que el mercado lo siga. Por lo tanto, un inversor puede establecer una posición larga cuando el indicador de sentimiento genera dicha señal y salir cuando aparece la inversión (Beigi, Hu, Maciejewski y Liu, 2016). Ya se han llevado a cabo varias investigaciones que han puesto de manifiesto el potencial de utilizar los sentimientos para desarrollar e implementar estrategias de trading con la ayuda de métodos estadísticos y algoritmos avanzados. Concluyeron que la interacción entre las redes sociales y el sentimiento de las noticias, y los movimientos del mercado financiero podrían aprovecharse con técnicas de inteligencia computacional para desarrollar estrategias de inversión avanzadas (Beigi et al., 2016). En la siguiente sección se comentará brevemente la metodología utilizada en estos estudios y se destacarán las principales conclusiones del estudio.

2.4.2. Algoritmo Predictivo

En la sección anterior, se detallaron varios algoritmos y sus respectivas funciones con respecto al análisis de sentimientos. El estudio realizado por (Bollen et al., 2011) concluyó que un algoritmo de aprendizaje podría predecir los movimientos subsiguientes en el mercado de valores con una precisión del 87,6% cuando se le dio acceso a los datos de sentimiento de Twitter. Un estudio posterior de (Makrehchi, Shah, & Liao, 2013) se basa en los hallazgos de estudios anteriores formulando estrategias comerciales basadas en los resultados que reciben de su algoritmo predictivo. En un intento de reforzar el rendimiento de sus estrategias basadas en Twitter, el equipo utilizó un novedoso método basado en eventos para predecir el sentimiento de los futuros tweets a través de un algoritmo predictivo basado en eventos. Antes de detallar los resultados de las estrategias de negociación, es importante tener un sólido entendimiento de cómo el equipo utilizó los eventos y sentimientos de Twitter para desarrollar sus estrategias.

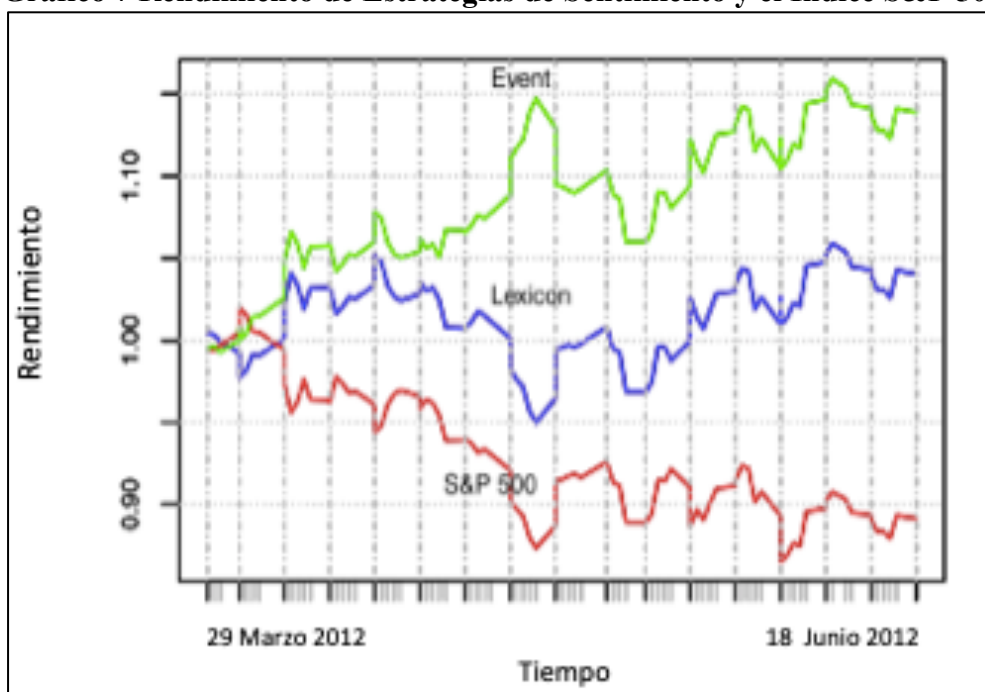
Con el fin de extraer el sentimiento de los Tweets, la mayoría de los investigadores han utilizado algoritmos de sentimiento que pueden derivar el sentimiento de los tweets mediante el análisis del lenguaje y la comparación con un léxico proporcionado (Li et al., 2016). Basado en ciertas palabras clave, los estados de ánimo específicos o la polaridad de los tweets pueden ser identificados. Además de realizar esta extracción de sentimientos estandarizada, (Makrehchi et al., 2013) también realizó una clasificación de sentimientos basada en eventos. La clasificación por sentimiento basada en eventos consistió en analizar el rendimiento relativo diario de los valores individuales que se encuentran en el S&P 500. La suposición clave que el equipo hizo fue que si el precio de una acción sube/disminuye significativamente en relación con el mercado de valores en general, entonces la mayoría de los tweets sobre dicha compañía deben ser positivos/negativos y debe estar ocurriendo un evento significativo. Por lo tanto, en los días en los que hay un evento positivo significativo (el precio de una acción está subiendo en relación con el mercado), todos los tweets se clasifican como con sentimientos positivos, y viceversa para los eventos negativos. Esta información se introdujo en un algoritmo de aprendizaje de sentimientos supervisado basado en eventos. Los investigadores creían que el sistema sería capaz de predecir el sentimiento de los futuros tweets, y luego agregar el sentimiento para hacer predicciones del mercado de valores. La idea detrás de las capacidades predictivas de los algoritmos es que sería capaz de reconocer patrones en los tweets que son indicativos de futuros rendimientos del mercado de valores (Makrehchi et al., 2013). Además, el equipo también entrenó otro algoritmo predictivo que fue alimentar el sentimiento de Twitter que se derivó del enfoque del léxico estándar.

2.4.3. Estrategias Basadas en Algoritmos Predictivos v.s. Índice S&P 500

En cuanto a la comparación de estrategias de trading, el objetivo del equipo de investigación era doble. En primer lugar, querían determinar si las estrategias de trading basadas en el sentimiento podían superar al índice S&P 500. En segundo lugar, querían determinar si los datos de Twitter derivados del enfoque de sentimiento de léxico, o los datos de Twitter derivados del enfoque de sentimiento de evento resultarían en una estrategia de trading superior. La selección del índice S&P 500 fue significativa por varias razones. Como todos los fondos indexados, se gestiona de forma pasiva y busca replicar el rendimiento de un índice de referencia. El S&P 500

consiste en las 500 acciones más grandes que cotizan en la Bolsa de Valores de Nueva York. Una estrategia de inversión pasiva común es simplemente comprar y mantener el índice, en lugar de negociar activamente las acciones individuales que lo componen. El equipo de investigación planteó la hipótesis de que podrían superar una estrategia de inversión pasiva ejecutando operaciones de las acciones individuales basadas en el sentimiento de Twitter. Para probar esta teoría, implementaron una simple estrategia de trading direccional que tomó una posición larga si el sentimiento de Twitter del día anterior era positivo neto, y una posición corta del sentimiento neto de Twitter del día anterior era negativo (Makrehchi et al., 2013).

Gráfico 7 Rendimiento de Estrategias de Sentimiento y el Índice S&P 500



Fuente: (Makrehchi et al., 2013)

Repitieron el experimento diariamente durante un poco más de tres meses. Como se observa en el gráfico 7, los resultados mostraron que ambas estrategias basadas en el sentimiento fueron capaces de predecir y capitalizar sistemáticamente los movimientos futuros de las acciones que forman parte del índice S&P 500, lo que dio lugar a rendimientos significativamente más altos que el simple hecho de mantener el índice.

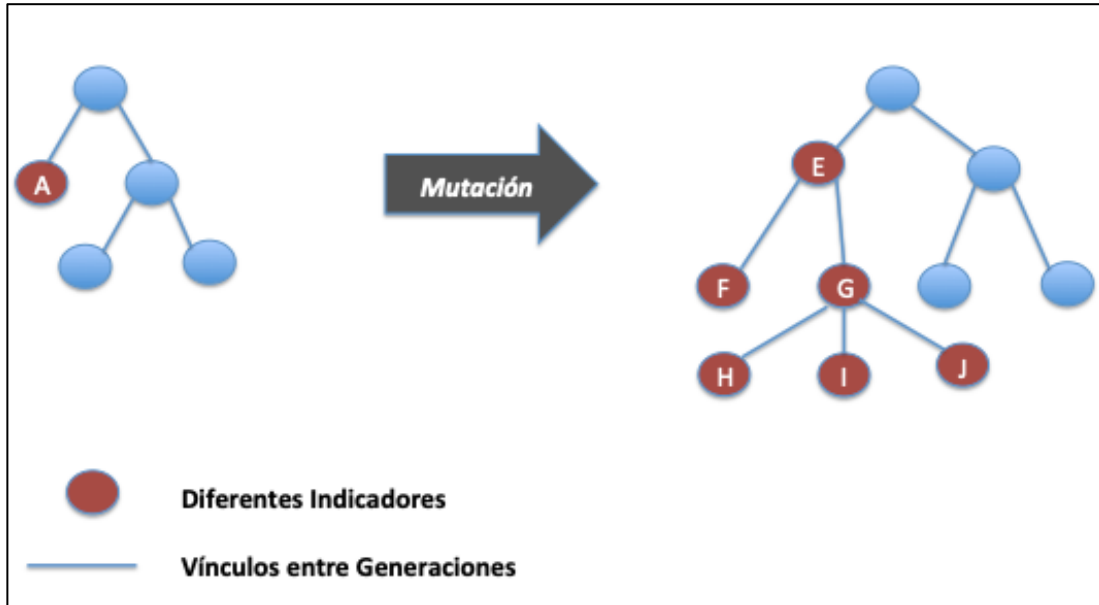
Acumulativamente, las estrategias basadas en el sentimiento superaron al Índice S&P 500 en un 20% en el período de tres meses (Makrehchi et al., 2013). Los retornos en este experimento fueron medidos por las ganancias y pérdidas teóricas que serían

generadas por las operaciones ejecutadas en cada estrategia. Curiosamente, el experimento también mostró que el enfoque basado en el sentimiento de eventos también superó sistemáticamente al enfoque del sentimiento del léxico.

2.4.4. Programación Genética

Con el fin de determinar si una técnica de trading basada en Twitter y el sentimiento de las noticias podría superar a una basada en técnicas tradicionales, también se utilizan técnicas de programación genética. La programación genética es una clase especial de algoritmos genéticos que se basan en los mismos principios que la selección natural. En la naturaleza, los miembros más aptos de una población sobreviven y transmiten su información genética a la siguiente generación. El objetivo de la programación genética en este caso es determinar la estrategia de trading óptima basada en los indicadores basados en el sentimiento proporcionado y en los indicadores basados en la técnica. Los indicadores son sinónimos de genes y se almacenan como datos que el algoritmo puede procesar. Los indicadores basados en los sentimientos se derivaron de Twitter, así como de artículos de noticias a través del análisis de sentimientos. Mientras tanto, los indicadores de base técnica se derivan de cálculos a los que se hace referencia con frecuencia en la literatura técnica tradicional. Para determinar la idoneidad relativa de cada combinación, se calculará el *Sterling ratio* para cada combinación hipotética. El ratio del *Sterling ratio* representa el total de los rendimientos acumulados a lo largo de la estrategia de trading, que son a menudo los criterios más referenciados para medir el éxito financiero. (Yang et al., 2017). Después de comparar la relación de el *Sterling ratio* para cada combinación de indicadores, sólo las combinaciones más adecuadas de indicadores pueden pasar a la siguiente generación de pruebas. A lo largo de numerosas generaciones, la combinación de indicadores evolucionará con el objetivo de optimizar el *Sterling Ratio* y encontrar una estrategia de trading ideal.

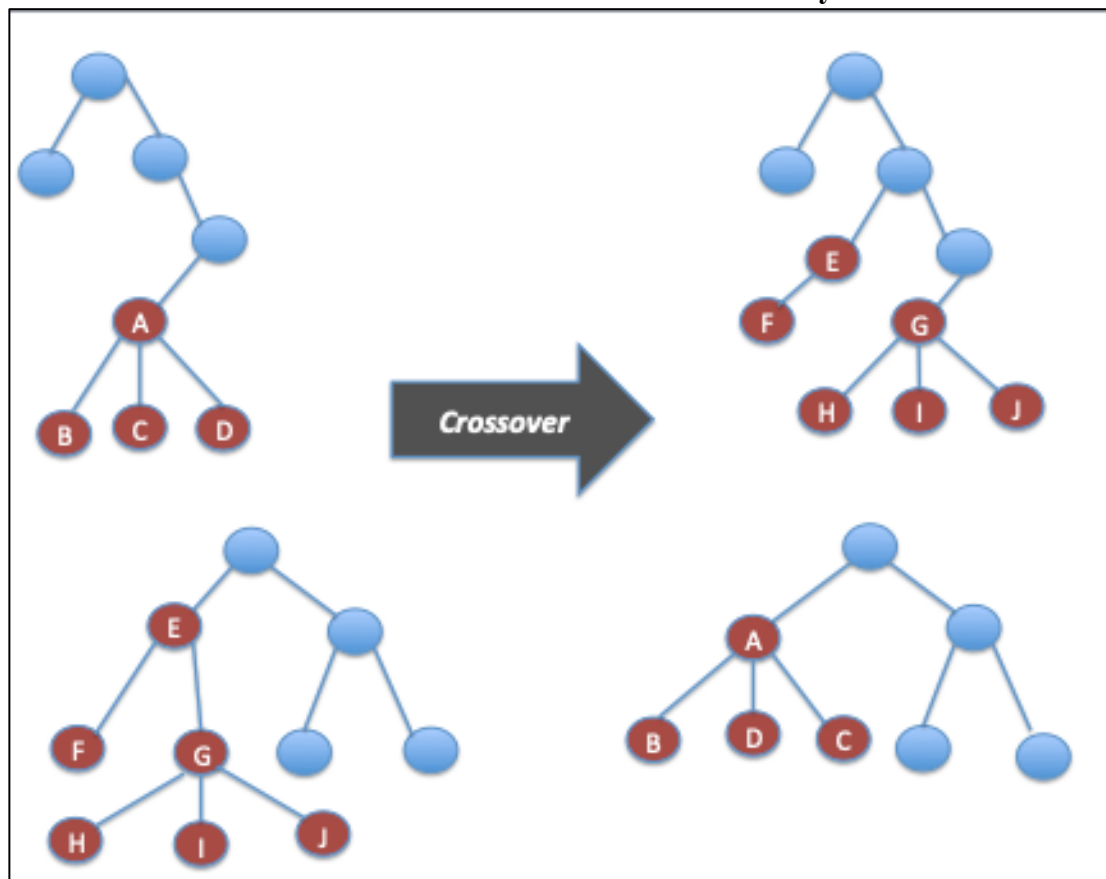
Gráfico 8 Mutación de Indicadores Tradicionales y de Twitter



Fuente: adaptado de (Yang, Yin, Mo, Liu, & Kirilenko, 2017)

En línea con el componente genético de este algoritmo, las combinaciones de sentimientos e indicadores técnicos son capaces de mutar y crossover para compartir material genético y explorar nuevas combinaciones de datos (Beigi et al., 2016). Un ejemplo de una mutación puede verse en el gráfico 8. El gráfico representa la mutación durante la evolución generacional de las combinaciones de indicadores. Cada círculo es una combinación única, y sólo el círculo que produce el mayor *Sterling Ratio* es capaz de tener descendencia y contribuir con sus datos a la siguiente generación (Beigi et al., 2016). Los círculos rojos son los descendientes de la combinación más exitosa. El objetivo de la mutación es aumentar la diversidad de la población mediante la introducción de un cambio aleatorio en los genes.

Graficó 9 Crossover de Indicadores Tradicionales y de Twitter



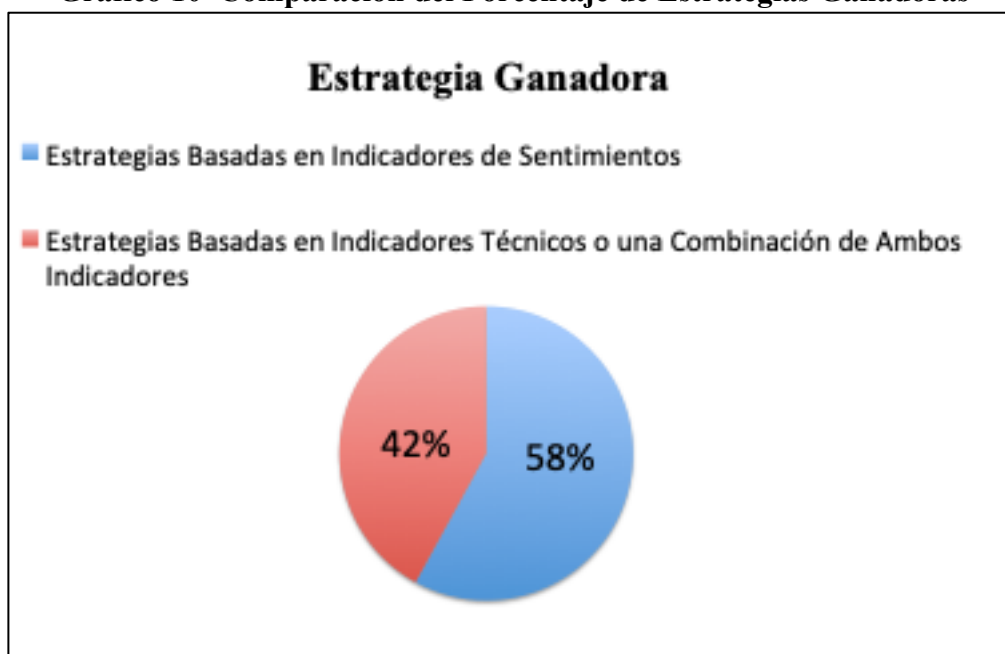
Fuente: adaptado de (Yang et al., 2017)

En el graficó 9, se está produciendo un crossover. El objetivo de un crossover es también aumentar la diversidad mediante el intercambio de material genético entre los individuos de una población. En este escenario, el material genético de los círculos rojos se intercambia para crear y probar nuevas combinaciones. El algoritmo continuó creando nuevas generaciones tanto de Twitter como de estrategias de trading tradicionales hasta que alcanzó lo que se consideraba la combinación ideal de los indicadores. Las diversas estrategias se probaron en un sistema de comercio controlado para determinar qué estrategias tenían los mejores resultados. En el estudio realizado por (Yang et al., 2017), su simulador de operaciones se basó en información financiera real del S&P 500 ETF.

2.4.5. Estrategias Basadas en la Programación Genética de Sentimientos vs. Estrategias Tradicionales

Los resultados de los estudios revelaron que las estrategias de trading que incorporan indicadores de sentimiento derivados de Twitter superan a las estrategias de trading tradicionales. El algoritmo de programación genética encontró que cuando se comparan estrategias de trading cuyas entradas provenían de indicadores basados en sentimientos, indicadores basados en técnicas o una combinación de ambos indicadores, los indicadores basados en sentimientos conducían a la estrategia ganadora el 58% del tiempo como se puede ver en el gráfico 10 (Yang et al., 2017).

Gráfico 10 Comparación del Porcentaje de Estrategias Ganadoras

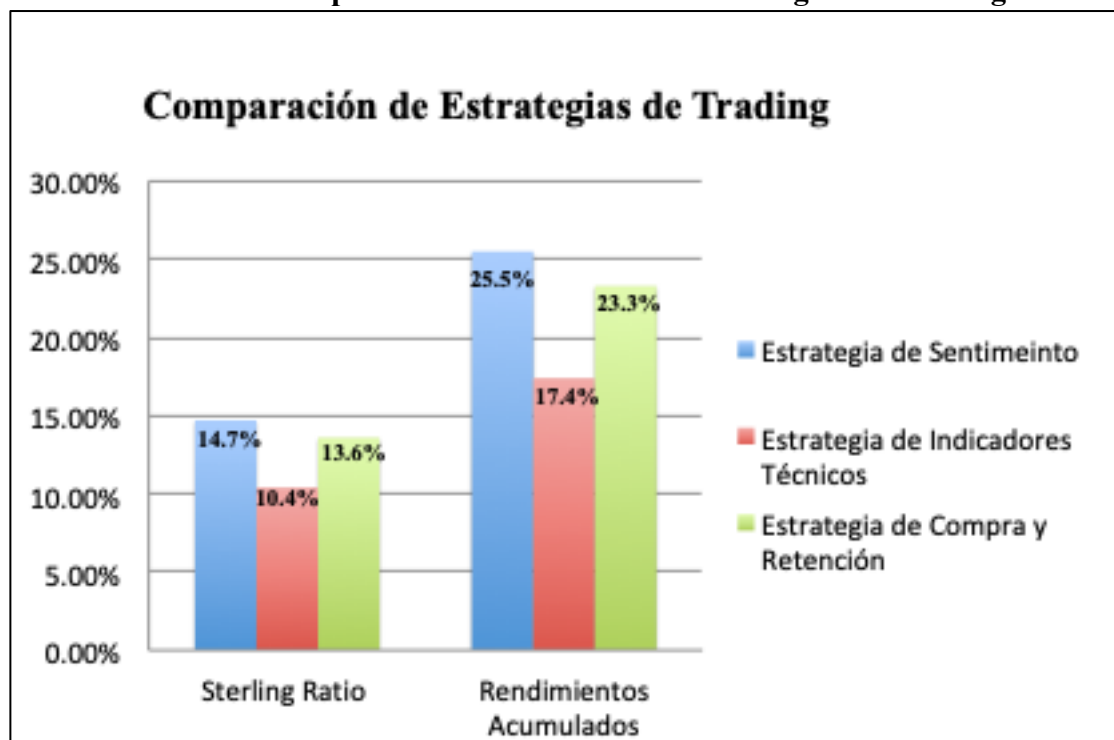


Fuente: adaptado de (Yang et al., 2017)

Con el fin de probar los resultados de la estrategia basada en el sentimiento, se probó contra dos estrategias de referencia. El primer punto de referencia era una estrategia típica basada en indicadores técnicos, y el segundo era una estrategia de compra y retención que es frecuentemente utilizada por pequeños inversores y fondos de inversión. La estrategia de sólo sentimiento óptimo generó un *Sterling Ratio* del 14,7%, en comparación con el 10,4% y el 13,6% de la estrategia de indicadores técnicos y la estrategia de compra y retención, respectivamente (Yang et al., 2017). Para una mayor comparación de los resultados de cada estrategia durante el período de evaluación, las estrategias basadas en la retroalimentación de sentimientos arrojan

los mejores resultados con rendimientos acumulados superiores al 25,5%, en comparación con el 17,4% y el 23,3% de las dos estrategias de referencia respectivamente (Yang et al., 2017). El gráfico 11 resume estos resultados.

Gráfico 11 Comparación de las Diferentes Estrategias de Trading



Fuente: adaptado de (Yang et al., 2017)

Para hacer los resultados más realistas, también se consideró el impacto de los costos de trading en la rentabilidad respectiva de dos de las estrategias de trading más rentables (basadas en el sentimiento frente a la combinación). Sin considerar los costes de trading, la rentabilidad relativa de cada estrategia, medida por el *Sterling Ratio*, es similar. Si se incluyen los costos de trading, los costos de punto de equilibrio para la estrategia basada en sentimientos y la estrategia de combinación, se encuentran en 92 bps y 10 bps, respectivamente (Yang et al., 2017). El mayor costo del punto de equilibrio para la estrategia basada en el sentimiento indica que la estrategia puede soportar fluctuaciones de mercado más grandes y aún así ser más rentable que la estrategia de combinación en transacciones reales (Yang et al., 2017).

3. INVESTIGACIÓN: TWITTER EN LA INDUSTRIA FINANCIERA

Como se ha dicho al principio de este trabajo, el objetivo de este proyecto es determinar si el poder predictivo de Twitter puede ser aprovechado para ayudar a los inversores a vencer al mercado de valores. Hasta ahora, este documento se ha dividido en varias secciones. La primera sección explora cómo Twitter ha llevado a una explosión de datos en tiempo real que pueden diseminarse a través de canales de comunicación globales. La segunda sección utilizó varias teorías financieras para proporcionar la base de un vínculo predictivo entre el sentimiento de Twitter y los movimientos del mercado de valores. La tercera sección de este trabajo revisó los resultados de tres casos académicos que utilizaron análisis de sentimientos y análisis avanzados para demostrar un vínculo predictivo y concreto entre Twitter y el mercado de valores. Además, el trabajo también investigó si las estrategias de trading basadas en el sentimiento de Twitter podrían dar lugar a mayores rendimientos que las estrategias de trading tradicionales. Hasta ahora, los estudios que han explorado el vínculo entre Twitter y el mercado de valores se han realizado en un entorno totalmente controlado. Numerosos investigadores han podido reproducir los resultados de los demás y llegar a conclusiones similares. En otras palabras, basado en el consenso de la comunidad académica, la incorporación de los conocimientos de Twitter en las estrategias de inversión y trading debería teóricamente conducir a mayores rendimientos y mejores resultados. Sin embargo, actualmente existe una gran brecha entre los beneficios teóricos que Twitter puede aportar al sector financiero y la implementación en la vida real. Es importante recordar que Twitter no se lanzó hasta 2006. El documento, que muchos consideran el primero en proponer y demostrar un vínculo predictivo entre Twitter y el mercado de valores, fue escrito en 2011. A su vez, la investigación académica que determinó que Twitter podría conducir a estrategias comerciales superiores fue publicada en 2016. Todo esto busca enfatizar que si bien la incorporación de datos de Twitter en las estrategias de inversión es vanguardista y prometedora, también se encuentra en su infancia.

Hasta ahora, la literatura sobre Twitter y el mercado de valores ha pasado por alto en gran medida el análisis de la viabilidad y los impactos potenciales que la implementación real del análisis de Twitter podría tener en el sector financiero. El ámbito académico ha proporcionado información importante al demostrar que existe

una relación predictiva y que puede ser aprovechada para obtener beneficios económicos. Sin embargo, ahora es el momento de determinar el impacto que la inversión basada en Twitter podría tener en caso de ser ampliamente adoptada por la industria financiera. El objetivo principal de este documento es determinar si Twitter puede utilizarse para ayudar a los inversores a predecir los movimientos de los mercados bursátiles, a través de la implementación de dicha información en estrategias superiores de inversión y trading. Para que este escenario tenga lugar, es necesario que se apliquen en la vida real los hallazgos teóricos que este proyecto ha explicado en detalle. Por lo tanto, este artículo proporcionará un análisis detallado del estado actual del análisis de Twitter en la comunidad financiera. Este análisis incorporará dos actores importantes en el mercado: las empresas que están analizando los datos de Twitter, y las instituciones que están implementando los datos en sus decisiones de inversión. Este análisis también incluirá los principales retos y limitaciones a los que se enfrenta cada parte interesada, y concluirá con una mirada al futuro sobre cómo el sector financiero podría verse afectado en los próximos años por la implementación generalizada de estrategias de trading basadas en Twitter.

3.1. PROVEEDORES DE ANÁLISIS DE DATOS DE TWITTER

3.1.1 Estado actual

Para salvar la brecha entre el mundo académico y la industria financiera, es necesario que exista un mecanismo mediante el cual los inversores puedan acceder a los indicadores basados en Twitter. Una sección de este documento detalla el complejo proceso por el cual los datos de Twitter pueden ser analizados con el fin de extraer indicadores basados tanto en el volumen como en el sentimiento. Los métodos incluyen el uso de aprendizaje automático y algoritmos, combinados con análisis estadísticos y analíticos avanzados. La ejecución exitosa de este proceso requiere una formación técnica específica que la mayoría de la gente no posee. Además, las investigaciones son lentas y costosas. Esto plantea la siguiente pregunta: ¿cómo podría un inversor que quisiera incorporar el análisis de Twitter en sus estrategias comerciales tener acceso a información precisa y fiable? Seguramente, es irrazonable y casi imposible esperar que los propios inversores puedan llevar a cabo sus respectivos análisis de Twitter. Esto requeriría experiencia y recursos que no son accesibles a la gran mayoría de los inversores. Sin embargo, el mercado responde a

las necesidades y en la última década ha habido una explosión de nuevas empresas de tecnología financiera que utilizan las mismas técnicas de análisis de datos que se describieron en este proyecto para obtener información financiera de fuentes de datos no tradicionales. La Tabla 1 proporciona una breve visión general de las empresas más exitosas en esta nueva industria. Estas nuevas fuentes de datos se han denominado acertadamente "datos alternativos". Debido a las características únicas que se mencionan en la primera sección de este documento, Twitter es especialmente valorado por su estrecha vinculación con el mercado de valores y la rápida difusión de la información. Por lo tanto, una parte importante del negocio de estas empresas se deriva de la extracción de información de Twitter en tiempo real.

Tabla 1 Empresas que Realizan Análisis de Datos de Twitter

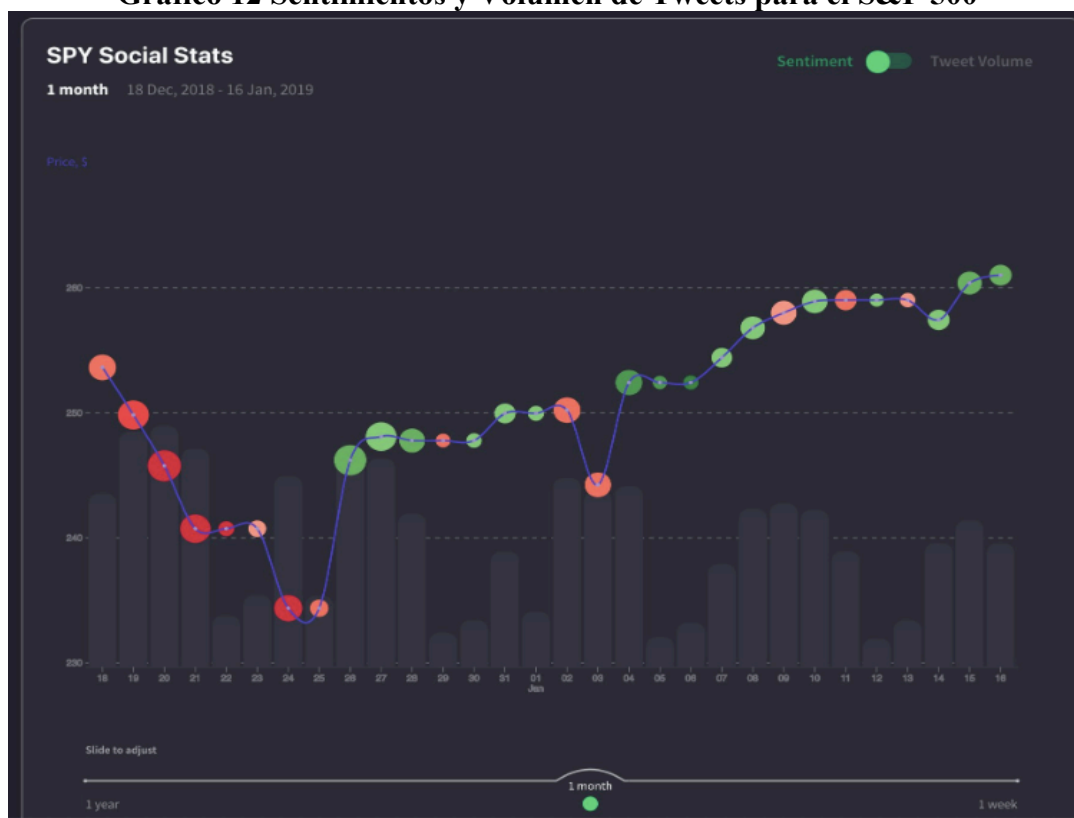
Empresa	Fecha de Fundación	Misión	Servicios Ofrecidos	Clientes	Tecnología Empleada
PsychSignal	2011	Es una API y plataforma líder en análisis de datos sociales y sentimientos que permite a los inversores tomar decisiones más informadas sobre la inversión y la negociación en los mercados financieros	Además de los indicadores en tiempo real, la plataforma PsychSignal proporciona herramientas de gráficos analíticos, índices de estado de ánimo del mercado, chat privado, noticias de tendencias, alertas y una gran cantidad de otras funciones	Inversores institucionales	Algoritmos de procesamiento de lenguaje natural e índices de sentimiento social centrados en el mercado financiero
iSentium	2008	iSentium le ofrece hoy las noticias de mañana. El flujo de datos único e inmediato ofrece una ventaja directa a las instituciones financieras que han luchado por encontrarla en los últimos años	Indicadores Direccionales Diarios (IDDs): indicador de sentimiento diario para pronosticar el movimiento del mercado entre la apertura y el cierre; Sentimiento derivado en tiempo real del acceso completo a los feeds de Twitter + otras plataformas de medios sociales	Inversores institucionales	Algoritmos de procesamiento de lenguaje natural, analítica avanzada, inteligencia artificial
MarketPsych	2012	Líder en investigación y consultoría de finanzas del comportamiento, integrando las ventajas competitivas derivadas de la psicología moderna en la industria financiera global	La creación y el aumento de estrategias de negociación, la previsión de la volatilidad, la gestión de riesgos. El producto final se distribuye a través de una asociación con Thomson Reuters, que también realiza extensos controles de calidad de los datos.	Bancos de inversión, hedge funds y traders	Procesamiento del lenguaje específico del dominio, análisis multidimensional de los sentimientos

Fuente: adaptado de (PsychSigna, iSentium, MarketPsych, 2019)

El concepto de terceras empresas que sintetizan información bursátil compleja en nombre de los inversores no es un concepto nuevo o novedoso. El análisis técnico, que se basa en el precio pasado y las tendencias comerciales, ha sido tradicionalmente un método popular que los inversores utilizan para basar sus predicciones de mercado.

Al igual que el análisis de sentimientos, la extracción de indicadores técnicos requiere el análisis de datos. Aunque el análisis técnico es significativamente menos complejo que el análisis de sentimientos, sigue siendo lento e intensivo. Los sitios de información financiera comunes, como Yahoo Finance, incluyen indicadores de análisis técnico como Bollinger Bands, RSI y MACD para cada empresa que cotiza en bolsa. Para que las inversiones basadas en análisis de Twitter tengan el mismo uso generalizado, es necesario que exista una fuente de información equivalente y de fácil acceso. Afortunadamente, la gran cantidad de nuevas empresas cuyo modelo de negocio consiste en el análisis avanzado de datos con el fin de proporcionar información financiera basada en Twitter sugiere que la investigación académica a la que se hace referencia en este proyecto se ha filtrado con éxito en el mundo de los negocios. Además, las empresas han destilado con éxito datos complicados y voluminosos en herramientas de análisis financiero fáciles de interpretar. El gráfico 12 muestra una herramienta incluida en los servicios que PsychSignal ofrece a los inversores. El gráfico es el rendimiento diario del S&P 500, y los círculos son el sentimiento agregado diario de Twitter. Los círculos rojos indican un sentimiento negativo, mientras que los círculos verdes indican un sentimiento positivo.

Gráfico 12 Sentimientos y Volumen de Tweets para el S&P 500



Fuente: (PsychSignal, 2018).

Los datos se obtienen directamente de la API de Twitter y permiten una visualización en tiempo real de cómo interactúan el precio de las acciones, el sentimiento de Twitter y el volumen de Twitter. Las herramientas que estas empresas ofrecen a sus clientes facilitan la incorporación de Twitter basado en el análisis de las decisiones financieras, haciendo que la información sea de fácil acceso y de fácil interpretación.

3.1.2. Desafíos para los proveedores de análisis de datos de Twitter

A pesar del exitoso lanzamiento de numerosas empresas que se dedican a la obtención de datos alternativos, el sector no está exento de desafíos. El análisis tradicional de las inversiones sigue siendo dominante, y el interés en implementar los indicadores de Twitter en las decisiones financieras ocupa un nicho específico dentro de la comunidad de inversiones. Sin embargo, un desafío aún más apremiante que podría deshacer los años de progreso que estas empresas han hecho es el componente legal/cumplimiento. El Tabla 2 que se encuentra en la pagina siguiente ofrece un resumen de la legislación vigente en materia de protección de datos. La tabla incluye una visión general de dónde se aplica la legislación geográficamente, una descripción de la ley, la fecha en que fue promulgada y las sanciones en caso de violación de la ley.

Tabla 2 Legislación Vigente para Proteger los Datos de los Consumidores

País	Ley	Fecha	Multas
Unión Europea	El Reglamento General de Protección de Datos: ha sido creada con la meta de reglamentar el modo en el que los datos personales de los ciudadanos europeos son protegidos por las empresas, y con la de reemplazar la Directiva 95/46/EC sobre la Protección de Dato	2018	Nivel 1: 2% del volumen de ingreso anual de la empresa, o 10 millones de euros, el que sea mayor Nivel 2: 4% del volumen de ingreso anual de la empresa, o 20 millones de euros, el que sea mayor
Reino Unido	<u>Data Protection Act:</u> Una ley nacional que complementa el Reglamento general de protección de datos de la Unión Europea	2018	Nivel 1: 2% del volumen de ingreso anual de la empresa, o 10 millones de euros, el que sea mayor Nivel 2: 4% del volumen de ingreso anual de la empresa, o 20 millones de euros, el que sea mayor
Estados Unidos	En la actualidad, los Estados Unidos no cuentan con una regulación cohesiva a nivel federal que regule la forma en que las empresas utilizan los datos de sus clientes o que proteja la privacidad de los consumidores		

Fuente: adaptado de (GDPR.ORG, 2018)

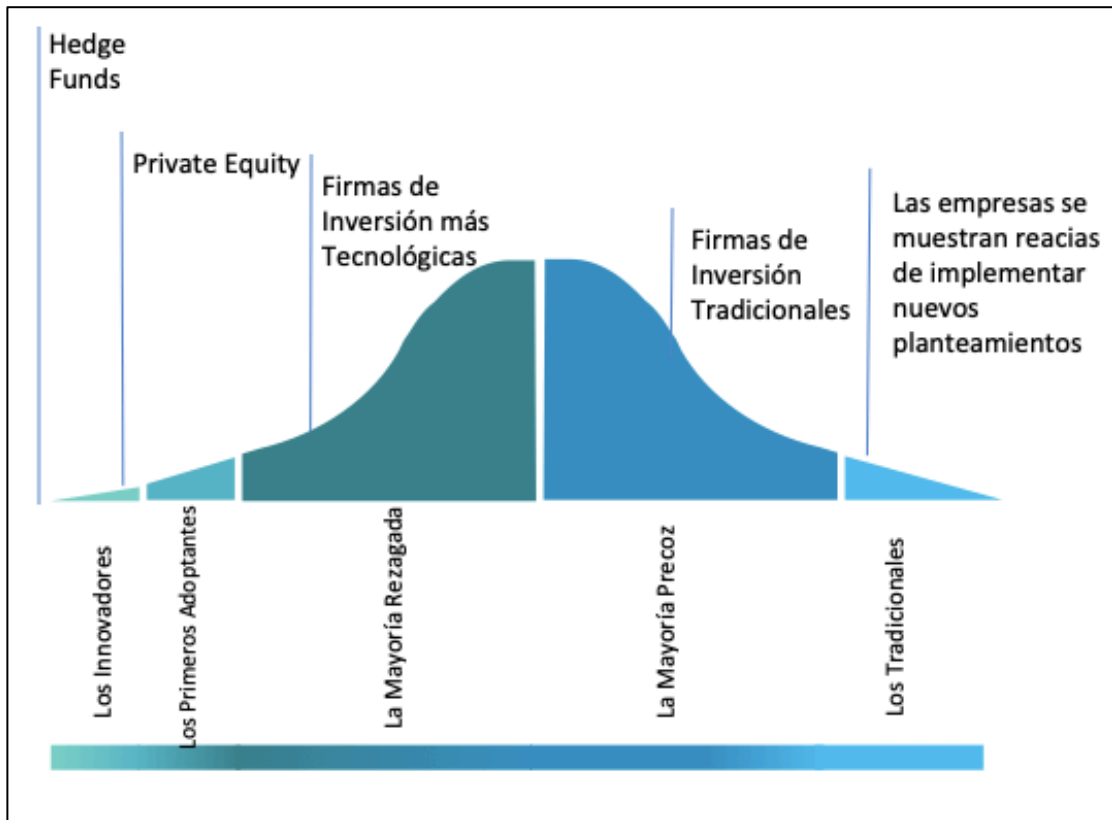
En estos tiempos, el uso y mal uso de los datos está siempre presente. Cuando una persona publica un tweet en Twitter, es cierto que está poniendo información en el dominio público. Sin embargo, existe una ambigüedad significativa en cuanto a la forma en que los terceros pueden utilizar esos datos. En el caso de las empresas alternativas de tecnología de búsqueda de datos, están accediendo a millones de tweets al día sin el consentimiento explícito de los editores de tweets originales. Esto plantea importantes cuestiones legales, como la violación de los derechos de autor, y el uso intencionado o no de información personal identificable. Para complicar aún más las cosas, Twitter es una plataforma ubicua que trasciende las fronteras internacionales. Mientras tanto, diferentes países tienen leyes y regulaciones específicas sobre el uso de datos que difieren significativamente. Los efectos de esta zona gris legal son negativos por varias razones. Para empezar, cuestiona la viabilidad a largo plazo de las empresas. La legislación sobre protección de datos cambia constantemente, y las tendencias recientes, como lo demuestra el *General Data*

Proteccion Regulation (GDPR) 2018, sugieren que la legislación para proteger los datos de los consumidores se está endureciendo en lugar de aflojarse. Además, los clientes que compran datos de empresas de datos alternativos también se exponen a riesgos legales. Esto podría amortiguar aún más la incorporación de los datos de Twitter en las prácticas de inversión más comunes.

3.2. Implementación actual del análisis de Twitter en empresas financieras

Aunque las empresas están procesando y analizando datos de Twitter, esto no significa necesariamente que los inversores estén implementando la información en sus decisiones financieras. Como se ha señalado en el resumen de las actuales empresas de datos alternativos, su mercado objetivo es el de los inversores institucionales. Esto significa que en lugar de intentar involucrar a todo el mercado, estas empresas están centrando sus esfuerzos en un segmento específico de inversores. Los inversores institucionales son conocidos como los gigantes del mundo de las inversiones. En comparación con los inversores habituales que venden/compran activos en su propio nombre, los inversores institucionales son organizaciones que agrupan fondos e invierten en nombre de otros. Ejemplos comunes de un inversionista institucional serían los fondos mutuos, los fondos de dotación, los fondos de pensiones y los *hedge funds*. Los *hedge funds* el mayor tipo de inversor institucional, y actualmente tienen 2,9 billones de dólares en activos bajo gestión (Statista, 2018). Su enorme tamaño y riqueza, combinados con intensas presiones competitivas, hacen que los *hedge funds* sean los más propensos a buscar y ser receptivos a fuentes de datos alternativas como Twitter. Sin embargo, esto no significa que otros tipos de inversores nunca utilizarán fuentes de datos alternativas como Twitter. En cambio, implica que la posición que los *hedge funds* mantienen en los mercados los impulsa a ser más innovadores, y también les proporciona los recursos necesarios para ocupar ese papel. En el futuro, es probable que más tipos de inversores empiecen lentamente a adoptar e implementar los datos de Twitter en sus decisiones de inversión. La Curva de Difusión de Innovaciones de Everett M. Rogers ilustra además cómo la implementación de fuentes de datos alternativas como Twitter se extenderá y difundirá por todo el sector financiero.

Graficó 13 La Curva de Difusión de Inversión basada en Twitter en el Sector Financiero



Fuente: adaptado de (Dannemiller and Kataria, 2017)

El gráfico 13 muestra la curva de difusión de la innovación de los datos de Twitter en el sector financiero para cada tipo de inversor. Según Rogers, los innovadores son los participantes en un sector que está más dispuesto a asumir riesgos. Ocupan una posición de prestigio dentro de la industria respectiva, y tienen fuentes dentro de la comunidad científica. Su tolerancia al riesgo y sus tendencias competitivas les permiten adoptar nuevas tecnologías que no han sido probadas y que, en última instancia, podrían fracasar. Sus recursos financieros les permiten cancelar cualquier empresa fracasada. Esta descripción encapsula perfectamente los *hedge funds* y ayuda a destacar por qué las empresas de datos alternativos los eligieron como su mercado objetivo. Después de los *hedge funds*, el siguiente grupo dentro de la industria financiera que tiene más probabilidades de incorporar datos de Twitter en sus estrategias de inversión son las empresas de private equity. En la curva de difusión, se denominan los primeros adoptantes, y se definen por su liquidez financiera y su educación avanzada. Sin embargo, a diferencia de los innovadores, son más discretos en sus opciones de adopción y necesitan más pruebas antes de estar dispuestos a

adoptar nuevas ideas. El siguiente grupo en el modelo de difusión se conoce como la mayoría temprana. Constituyen una gran parte del sector financiero, y si este grupo decidiera implementar el pronóstico basado en Twitter, significaría que la técnica está empezando a convertirse en algo habitual en la industria. La primera mayoría en la industria financiera serían probablemente empresas de gestión de inversiones con conocimientos técnicos que son más receptivas a las últimas tecnologías. En este punto, la inversión basada en Twitter será común en el mercado y sólo será cuestión de tiempo antes de que los escépticos participantes del mercado, conocidos como mayoría tardía y rezagados, adopten finalmente las nuevas técnicas de datos alternativos.

3.3. LA UTILIZACIÓN DE DATOS DE TWITTER EN LOS *HEDGE FUNDS*

3.3.1. Estado actual

Actualmente, el mercado se encuentra en las primeras etapas de adopción y sólo los innovadores han comenzado a implementar la información de Twitter en sus estrategias de inversión. En esta sección se analizará el nivel de implementación y los beneficios actuales que los datos alternativos están proporcionando a estas empresas. Para un *hedge funds*, el uso de tácticas para obtener una ventaja informativa sobre la competencia no es un fenómeno nuevo. Antes de la era de Internet, se rumoreaba que los fondos de cobertura enviarían becarios para contar el número de coches en los concesionarios de automóviles, o el número de clientes que entraban o salían de las grandes tiendas minoristas con el fin de obtener información sobre el mercado. Por lo tanto, no es sorprendente que el 70% de los inversores creen que los datos en tiempo real les dan una ventaja (Greenwich Associates, 2017). Desafortunadamente, la naturaleza competitiva de los *hedge funds* también los hace infamemente reservados. Sus estrategias, tecnologías y actuaciones reales están estrechamente vigiladas y no están a disposición del público. Sin embargo, todavía es posible determinar cómo las fuentes de datos alternativos como Twitter están afectando a los fondos de cobertura a nivel macro. En 2017, EY entrevistó a los principales gestores de *hedge funds* y descubrió que el 60% de ellos tienen sus oficinas centrales experimentando con fuentes de datos alternativos. En el momento de la encuesta, el 27% de los gestores de *hedge funds* estaban utilizando datos alternativos en sus herramientas de trading y los

medios sociales eran la fuente de datos alternativos más popular (Dannemiller and Kataria, 2017). Una anécdota de 2013 también sugiere que los *hedge funds* están analizando agresivamente los medios sociales para encontrar oportunidades de vencer al mercado. En este caso específico, un hombre creó una cuenta falsa en Twitter para hacerse pasar por firmas financieras de renombre, y publicó tweets que afirmaban que la compañía Audience estaba siendo investigada por fraude, y que la FDA ha incautado los registros de ensayos clínicos de Sarepta Therapeutics. Las acciones cayeron un 28% y un 16% respectivamente (Wieczner, 2015). A pesar de haber engañado con éxito al mercado, el autor sólo pudo sacar \$97 USD de su plan cuando hizo sus transacciones 10 minutos después de publicar los tweets. Esto implica que los inversores que comercian en grandes volúmenes, como un *hedge fund* fueron capaces de detectar la estafa, y mover los precios de mercado de vuelta a sus valores fundamentales en menos de 10 minutos.

3.3.2. Desafíos para los *Hedge Funds*

Graficó 14 Retos a los que Se Enfrentan Actualmente los *Hedge Funds*



Elaboración Propia

Aunque la evidencia muestra que los *hedge funds* están implementando fuentes de datos alternativas en sus estrategias de inversión, aún deben superarse importantes desafíos. A nivel macro, los desafíos pueden dividirse en cuatro categorías: regulación, talento, costos y escepticismo como se puede ver en el graficó 14.

Reglamento:

En la sección anterior se enumeraron los problemas reglamentarios a los que se enfrentan los proveedores de datos alternativos. La misma incertidumbre de los cambios en la regulación y en las leyes de noticias también afecta a las empresas que utilizan fuentes de datos alternativas. Tanto el vendedor como el comprador se enfrentan al mismo riesgo de procedencia de los datos. El riesgo de procedencia de los datos se refiere al origen de los datos y a la forma en que fueron recopilados. Si los datos no se obtienen o utilizan de acuerdo con los términos de la fuente original, podrían emprenderse acciones legales. Las empresas de gestión de inversiones

también están obligadas a no utilizar datos que contengan información de identificación personal. Un riesgo regulador adicional que los *hedge funds* deben tener en cuenta es el importante riesgo de la información no pública. Esta información podría tener un impacto en el precio de las acciones de una empresa y en las decisiones de inversión. La utilización de esta información antes de que se haga pública podría considerarse como uso de información privilegiada, lo que conlleva severas penalizaciones en la mayoría de los países. Las fuentes de datos alternativas como Twitter tienen capacidades de predicción que pueden poner en duda cuándo se hace pública exactamente la información.

Talento:

La naturaleza compleja e intrincada de los datos alternativos requerirá que los *hedge funds* contraten personal con antecedentes especializados. Para las empresas de gestión de inversiones, los beneficios de los datos alternativos dependerán de lo bien que puedan analizar, interpretar e incorporar los datos en las estrategias de trading o en la construcción de la cartera. Esto significa que podría producirse una competencia entre los *hedge funds* para contratar a los mejores talentos. Algunas evidencias ya indican que los *hedge funds* están reclutando agresivamente a científicos de datos y seduciéndolos con altos salarios (Hope & Zuckerman, 2017). La rotación de talento entre empresas también supone un riesgo para los *hedge funds*, ya que podría dar lugar a la pérdida de información y conocimientos técnicos propios.

Costes:

Los enormes costos que implica la implementación de datos alternativos también plantean un desafío para los *hedge funds*. Como se menciona en la Curva de Difusión de Innovaciones, los *hedge funds* son considerados innovaciones debido a sus vastos recursos económicos. Sin embargo, el costo asociado con los datos alternativos podría considerarse excesivo, especialmente si la gerencia superior no está convencida de sus beneficios. Para empezar, el acceso al flujo completo de feeds de Twitter cuesta \$30,000 USD al mes, o alrededor de \$1.5 millones USD al año (Wieczner, 2015). Además, habría que gastar fondos en todo el software para integrar los datos de Twitter con los modelos financieros y los esquemas de comercio existentes. Las

empresas también tendrían que invertir en la contratación y formación de nuevo personal, lo que también implica importantes costes de material y tiempo.

Escepticismo:

Mientras que el mundo académico está de acuerdo en que existe un vínculo predictivo entre Twitter y el mercado de valores, y que esta información puede ser utilizada para crear estrategias comerciales superiores, algunas partes del mundo de los *hedge funds* siguen siendo escépticas. Shanta Punchteler es el co-director de inversiones de Man Numeric, uno de los mayores *hedge funds* del mundo con 77.000 millones de dólares bajo gestión. En una entrevista declaró que su equipo de inversión no ha podido obtener beneficios fiables de las inversiones basadas en Twitter. Su opinión sobre los datos alternativos implica que el bombo es mayor que los beneficios reales: "Existe la romántica idea de que los tweets de Twitter son invendibles y que se puede ganar mucho dinero si se salta sobre ellos. Tienes que hacerte la pregunta, ¿Dónde está el valor?" (Wieczner, 2015). El escepticismo de los gestores de *hedge funds* no está exento de méritos. En 2011 se abrió el *hedge funds* británico Derwant Capital. Contaba con una estrategia de inversión completamente derivada de Twitter, que se cerró un mes más tarde. En 2013 se realizó un segundo intento, con resultados similares.

Dada la naturaleza segura y protegida de los *hedge funds*, es casi imposible evaluar si la incorporación de los datos de Twitter en las estrategias de inversión o en la creación de carteras ha dado lugar a rendimientos definitivamente más altos. Sin embargo, dados los retos actuales de la industria y el fracaso de los anteriores *hedge funds* basados en Twitter, es probable que el sector aún se esté adaptando al uso de las fuentes de datos de Twitter. En otras palabras, las técnicas de inversión tradicionales siguen siendo dominantes, más confiables y mejor implementadas. Como se ha señalado en la sección anterior, es necesario realizar inversiones significativas en términos de infraestructura y conocimientos técnicos. Una diferencia clave entre los estudios a los que se hace referencia en este proyecto y la implementación en la vida real de los datos de Twitter es la escala y la complejidad. Los investigadores académicos concentraron sus esfuerzos en probar/desprobar una hipótesis específica. Si bien esto sentó las bases teóricas críticas, el éxito de la aplicación en la vida real de estos estudios requerirá un período similar de prueba y error.

4. CONCLUSIONES

Hace diez años, la afirmación de que existe un vínculo entre Twitter y el mercado de valores fue recibida con escepticismo. En la década siguiente, la popularidad de Twitter entre los usuarios se ha multiplicado, y numerosos estudios han reafirmado que este vínculo existe de forma ubicua. Sin embargo, el origen de la relación, o cuáles podrían ser sus ramificaciones para la industria financiera, ha permanecido en gran medida sin investigar. Este proyecto se ha sumado a la literatura abordando estos dos temas con el fin de profundizar nuestra comprensión de cómo interactúan estos dos sistemas y cómo el sector financiero podría verse afectado por estos nuevos descubrimientos.

Este trabajo determinó que las características únicas que Twitter y el mercado de valores poseen por separado son, en última instancia, lo que los une. Como plataforma virtual, Twitter es capaz de trascender fronteras y conectar una red global de usuarios. Los propios tweets se producen a un ritmo constante, lo que demuestra un pulso en tiempo real en los pensamientos y sentimientos de la gente. En muchos sentidos, Twitter se comporta como un noticiero porque difunde información de forma rápida y eficaz. La psicología moderna sostiene que los estados de ánimo a menudo se difunden entre los usuarios junto con la información.

El mercado de valores, por otro lado, ha existido mucho más tiempo que Twitter. Sin embargo, en muchos sentidos se entiende aún menos. En este trabajo se presentaron dos teorías financieras en competencia: la Hipótesis de Mercado Eficiente y la de Finanzas de Conductuales. El HME afirma que es imposible superar el mercado de valores porque las acciones siempre cotizan a su valor justo. Mientras tanto, Las Finanzas Conductuales sostiene que el mercado es inherentemente ineficiente debido a la irracionalidad del comportamiento humano. Como resultado, si un estado de ánimo se comparte entre una gran cantidad de inversores, sus acciones pueden mover colectivamente los precios de las acciones. Por lo tanto, las emociones de los inversores proporcionan una visión predictiva de la actividad futura del mercado.

Con estas características clave en mente, el vínculo entre Twitter y el mercado de valores se hace más evidente. En términos concisos, el estado de ánimo social de los inversores puede extraerse de Twitter y predecir los movimientos del mercado. La

siguiente pregunta que se planteó en este proyecto fue si estos conocimientos podrían utilizarse para crear estrategias de trading superiores. Un análisis entre una estrategia comercial tradicional y una basada en Twitter determinó que la incorporación de la información de Twitter en las decisiones financieras resulta en carteras con mayores ganancias, menos riesgo y mayor resistencia a la volatilidad del mercado.

Otro enfoque importante de este proyecto fue entender cómo la inversión basada en Twitter está impactando la industria financiera. La investigación descubrió que ha habido una explosión de startups de tecnología financiera que proporcionan servicios ad hoc de análisis de Twitter. Esta tendencia sugiere que las herramientas de análisis avanzado que se analizaron en este documento se están comercializando actualmente. También indica que la relación predictiva entre Twitter puede ser replicada con éxito en un entorno real.

Desde la perspectiva del inversor, este trabajo determinó que los *hedge funds* son actualmente el único tipo de inversor que tiene los recursos e incentivos adecuados para empezar a incorporar datos basados en Twitter en sus protocolos de inversión. En concreto, el carácter competitivo de los *hedge funds*, combinado con sus recursos económicos, les permite ser más innovadores que el resto del sector. Actualmente, la comunidad de *hedge funds* parece ser receptiva a las fuentes de datos de Twitter. Numerosos indicadores revelan que la mayoría de los *hedge funds* están al menos experimentando con la implementación de datos en Twitter. Además, los *hedge funds* han invertido cantidades significativas de tiempo, dinero y esfuerzo en el desarrollo de la infraestructura necesaria para el análisis a gran escala de Twitter. En general, los inversores han tomado nota de la investigación académica líder, y están ansiosos por ver si se puede aprovechar en el mundo real.

Sin embargo, el futuro a largo plazo de la inversión basada en Twitter está lejos de ser seguro. La tendencia actual hacia una mayor protección de datos pone en tela de juicio la legalidad de las fuentes de datos alternativas y disuade tanto a los proveedores como a los usuarios de datos de Twitter de invertir en su desarrollo. Sin embargo, el mayor desafío al que deben enfrentarse las estrategias de inversión basadas en Twitter es el continuo escepticismo del sector. Los detractores no creen que Twitter pueda proporcionar valor de una manera real y consistente. Sus preocupaciones no carecen de mérito. Todos los *hedge funds* basados en Twitter han

cerrado debido a los malos resultados. Dado que los *hedge funds* no publican sus retornos o estrategias públicamente, es difícil decir si las estrategias de Twitter han sido refinadas y han dado lugar a mejores resultados. Se requerirá un análisis continuo para determinar si los datos de Twitter son sólo una moda temporal en el sector financiero, o si se convertirán en la norma.

5. BIBLIOGRAFÍA

- Admin. (n.d.). Win With The Wisdom Of The Crowd. Retrieved from <http://isentium.com/protected-win-with-the-wisdom-of-the-crowd/>
- Alternative data: After the gold rush, what are the risks for investors? (n.d.). Retrieved from <https://www.locktoninternational.com/gb/articles/alternative-data-after-gold-rush-what-are-risks-investors>
- Assets managed by hedge funds globally 1997-2018 | Statistic. (n.d.). Retrieved from <https://www.statista.com/statistics/271771/assets-of-the-hedge-funds-worldwide/>
- Barnes, N.G. and Wright, S (2013), Fortune 500 are bullish on social media: Big companies get excited about Google+, Instagram, Foursquare and Pinterest. *Working Paper*, University of Massachusetts Dartmouth,
- Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). Sentiment Analysis and Ontology Engineering (Vol. 639). <https://doi.org/10.1007/978-3-319-30319-2>
- Blankespoor, E., Miller, G. S., & White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of TwitterTM. *Accounting Review*, 89(1), 79–112. <https://doi.org/10.2308/accr-50576>
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Cha, M., Benevenuto, F., Haddadi, H., & Gummadi, K. (2012). The world of connections and information flow in twitter. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 42(4), 991–998. <https://doi.org/10.1109/TSMCA.2012.2183359>
- Company. (n.d.). Retrieved from https://about.twitter.com/en_us/company.html
- Company History. (2019). Retrieved March, 2019, from <https://old.marketpsych.com/about/>
- Dannemiller, D., & Kataria, R. (2017). *Alternative data for investment decisions: Today's innovation could be tomorrow's requirement*(Rep.). Retrieved March 23, 2019, from Deloitte website: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/financial-services/us-fsi-dcfs-alternative-data-for-investment-decisions.pdf>

- Edmans, A., Garcia, D., & Norli, O. (2006). Sports Sentiment and Stock Returns. *SSRN Electronic Journal*. doi:10.2139/ssrn.677103
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383. doi:10.2307/2325486
- GDPR FAQs. (n.d.). Retrieved from <https://eugdpr.org/the-regulation/gdpr-faqs/>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014. doi:10.1038/nature07634
- Gussenhoven, C. J. (2014). Twitter 's effect on share price movements of the Johannesburg Stock Exchange ., (January).
- Hayes, A. (2019, March 12). R-Squared. Retrieved from <https://www.investopedia.com/terms/r/r-squared.asp>
- He, W., Guo, L., Shen, J., & Akula, V. (2016). Social Media-Based Forecasting. *Journal of Organizational and End User Computing*, 28(2), 74–91. <https://doi.org/10.4018/JOEUC.2016040105>
- Hope G., Zuckerman, B., (2017) “The Quants Run Wall Street Now,” *Wall Street Journal*
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis - WebKDD/SNA-KDD '07*, 56–65. <https://doi.org/10.1145/1348549.1348556>
- Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2169-2188. doi:10.1002/asi.21149
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59-68. doi:10.1016/j.bushor.2009.09.003
- Kollmeyer, B., & Shaw, J. M. (2018, December 20). Finance Twitter: The 50 most important people for investors to follow. Retrieved from <https://www.marketwatch.com/story/finance-twitter-the-50-most-important-people-for-investors-to-follow-2018-12-13>
- Lee, M. (2017, November 15). How will you use innovation to illuminate competitive advantages? Retrieved from https://www.ey.com/en_gl/wealth-asset-

management/how-will-you-use-innovation-to-illuminate-competitive-advantages

- Li, C. Te, Kuo, T. T., Ho, C. T., Hong, S. C., Lin, W. S., & Lin, S. De. (2013). Modeling and evaluating information propagation in a microblogging social network. *Social Network Analysis and Mining*, 3(3), 341–357.
<https://doi.org/10.1007/s13278-012-0082-8>
- Makrehchi, M., Shah, S., & Liao, W. (2013). Stock Prediction Using Event-Based Sentiment Analysis. 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT).
doi:10.1109/wi-iat.2013.48
- MARKETPSYCH OUR MISSION. (2019). Retrieved from
<https://www.marketpsych.com/>
- Mishne, G., & Glance, N. (2006). Predicting Movie Sales from Blogger Sentiment. AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, 155-158.
- McPartland, K. (2017). *Alternative Data for Alpha*(Rep.). Retrieved March 23, 2019, from Greenwich Associates website:
<https://www.dnb.com/content/dam/english/dnb-solutions/alternative-data-for-alpha-final.pdf>
- Nisar, T. M., & Yeung, M. (2018). ScienceDirect Twitter as a tool for forecasting stock market movements : A short-window event study. *The Journal of Finance and Data Science*, 4(2), 101–119. <https://doi.org/10.1016/j.jfds.2017.11.002>
- Nofer, M. (2015). Using Twitter to Predict the Stock Market: Where is the Mood Effect? *The Value of Social Media for Predicting Stock Returns*, 63-88.
doi:10.1007/978-3-658-09508-6_4
- Nofsinger, J. R. (2010). Social Mood and Financial Economics Social Mood and Financial Economics. *Journal of Behavioral Finance*, 7560(2005).
doi:10.1207/s15427579jpfm0603
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis, 2(1).
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PLoS ONE*, 10(9), 1–21.
<https://doi.org/10.1371/journal.pone.0138441>

- Rogers, E. M. (2003). *Diffusion of Innovation*. New York: The Free Press.
- Sentiment | Definition of sentiment in English by Oxford Dictionaries. (n.d.). Retrieved from <https://en.oxforddictionaries.com/definition/sentiment>
- Smith, A., Anderson, M., Smith, A., & Anderson, M. (2018, September 19). *Social Media Use 2018: Demographics and Statistics*. Retrieved from <https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and Trades : the Information Content of Stock Microblogs, *20(5)*, 926–957. <https://doi.org/10.1111/j.1468-036X.2013.12007.x>
- Tesla, Inc. (TSLA) Stock Price, Quote, History & News. (2019, April 04). Retrieved from <https://finance.yahoo.com/quote/TSLA/>
- Twitter: Number of active users 2010-2018. (n.d.). Retrieved from <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>
- Tafti, A., Zotti, R., & Jank, W. (2016). Real-time diffusion of information on twitter and the financial markets. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0159226>
- WIECZNER, J. (2015, December 7). How Investors Are Using Social Media to Make Money. *Fortune*.
- Wunsch-Vincent, S., & Vickery, G. (2007). DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY COMMITTEE FOR INFORMATION, COMPUTER AND COMMUNICATIONS POLICY. Retrieved from <https://www.oecd.org/sti/38393115.pdf>
- Yang, S. Y., Yin, S., Mo, K., Liu, A., & Kirilenko, A. A. (2017). Neurocomputing Genetic programming optimization for a sentiment feedback strength based trading strategy. *Neurocomputing*, *264*, 29–41. <https://doi.org/10.1016/j.neucom.2016.10.103>
- Yang, S. Y., Yin, S., Mo, K., Liu, A., Yang, S. Y., Yin, S., ... Liu, A. (2015). Twitter financial community sentiment and its predictive relationship to stock market movement Twitter financial community sentiment and its predictive relationship to stock market movement. *Quantitative Finance*, *7688*, 1–17. <https://doi.org/10.1080/14697688.2015.1071078>

Zhang, X., Han, D. D., Yang, R., & Zhang, Z. (2017). Users' participation and social influence during information spreading on Twitter. *PLoS ONE*, *12*(9), 1–17.
<https://doi.org/10.1371/journal.pone.0183290>