

FSAL: Lexicón financiero de sentimiento en español rioplatense diseñado para “Bolsas y Mercados Argentinos” (BYMA)

» Juan Pablo Braña, Alejandra M. J. Litterio, Alejandro Fernández
Centro de Altos Estudios en Tecnología Informática (CAETI) /
juan.brana@uai.edu.ar, alejandra.litterio@uai.edu.ar, alejandro.fernandezc@uai.edu.ar

Resumen

En la última década, se ha estudiado cómo el Análisis de Sentimiento basado en lexicones en combinación con técnicas de *Machine Learning* puede ser utilizado para optimizar estrategias de Trading Algorítmico. El presente trabajo tiene como objetivo mostrar que un lexicón de dominio específico en finanzas (FSAL) diseñado para Bolsas y Mercados Argentinos obtiene mejores resultados que un lexicón de propósitos generales (SDAL). Primero, proponemos un lexicón a medida en finanzas. Segundo, mostramos que nuestro lexicón supera los resultados obtenidos en comparación a los resultados de un lexicón de propósitos generales aplicado sobre un corpus compuesto por tweets de cuentas de comunidades de confianza de los mercados argentinos, previamente clasificado de manera colaborativa por expertos en finanzas. Luego, realizamos un estudio comparado de los lexicones aplicando diferentes técnicas de *Machine Learning*. Finalmente, presentamos algunos resultados preliminares y conclusiones.

PALABRAS CLAVE: ANÁLISIS DE SENTIMIENTO, LEXICÓN FINANCIERO, TRADING ALGORÍTMICO, MACHINE LEARNING

FSAL: A Tailor-made Financial Sentiment Lexicon in Spanish for the Argentinian Markets (BYMA)

Abstract

During the last decade studies have shown that lexicon-based Sentiment Analysis of tweets combined with Machine Learning techniques can be used to enhance Algorithmic Trading strategies. The aim of the present work is to show how a specific domain lexicon in finance for the Argentinian Markets (FSAL) provides a better outcome than a generic lexicon (SDAL). First, we introduce a finance tailor-made lexicon. Secondly, we experimentally show that our lexicon outperforms a general purpose one on a corpus of tweets previously classified collaboratively by

specialists in finance. Then, we compare the lexicons applying three different Machine Learning algorithms. Finally, we introduce some preliminary results and conclusions.

KEYWORD: SENTIMENT ANALYSIS, FINANCIAL LEXICON, ALGORITHMIC TRADING, MACHINE LEARNING

1. Introducción

La forma en que las noticias afectan los mercados, esto es, cómo impactan en los volúmenes negociados, en los retornos, en la volatilidad de los precios, en la toma de decisiones relacionadas con la compra-venta de los instrumentos financieros, entre otros, ha sido el centro de estudio de investigaciones durante los últimos años en la comunidad financiera [1, 3, 5, 6, 8, 11, 23, 24, 25, 26, 27, 34, 35, 38, 39, 40, 43, 52, 53]. El contenido "afectivo" de las noticias financieras se ha transformado en el aspecto focal de análisis con la finalidad de comprender, explicar e interpretar la relación entre el discurso pronunciado y la reacción de los mercados.

Diremos que todo discurso, cuya materialización se produce ya sea a través de comentarios en las redes sociales, en informes, titulares, blogs, contiene una dimensión no factual de significado que opera en la transmisión de una opinión o un "tono", un "estado" y que a su vez es capaz invocar una perspectiva matizada de un contenido factual. De esto se desprende una nueva forma de pensar el discurso económico-financiero en todas sus manifestaciones posibles, un interrogante sobre las dimensiones y cuestiones afectivas o juicios valorativos anclados en el discurso y los mecanismos, procedimiento o técnicas para analizarlo.

Así, por una parte, estudios sobre el discurso periodístico económico-financiero [23, 25, 26, 39, 40, 43, 44, 48, 55] han mostrado que las noticias e informes poseen un significado evaluativo valioso con un claro potencial capaz de influir a los actores que participan en la interacción comunicativa. Por otra parte, surge una nueva concepción de los actores o agentes en un mercado financiero concebido como un constructo social donde las acciones están motivadas por comportamientos que se ven afectados por una dinámica que gira en torno al discurso impregnado de un léxico evaluativo.

El léxico evaluativo y la interacción de los actores han motivado una aproximación teórico-metodológica no sólo en lo que refiere a los mecanismos o técnicas de análisis como observaremos en el presente trabajo sino también en cuanto al marco en relación a la concepción terminológica que da lugar a la creación o diseño de diccionarios o lexicones, sean genéricos o de dominio específico, que intervienen en el procesamiento de lenguaje natural y el desarrollo de algoritmos capaces de interpretar diferentes tipos de discursos, en especial el económico-financiero y los enunciados "subjetivos" que lo componen.

En este sentido y en estrecha relación con lo expresado, el gran volumen de datos, que compone nuestro corpus extraído de cuentas de comunidades de confianza de contenido de noticias de Twitter que refleja la situación y el impacto en los mercados financieros argentinos (BYMA¹), requieren ser procesados, es decir, estos discursos y su descomposición en unidades o enunciados² en niveles de análisis más profundos, por lo que precisan de técnicas que puedan resolver

¹ <https://www.byma.com.ar/que-es-byma/sobre-byma/>

² Entendemos por enunciado la unidad de la comunicación discursiva se determina por el cambio de los sujetos discursivos, que se realiza mediante una oración o expresión sintáctica más pequeña.

esta tarea de manera eficiente y con la velocidad que demandan los actores en los mercados financieros, esto es categorizar un enunciado, que contiene una valoración u opinión subjetiva, en positivo o negativo por una parte, y mostrar cómo un diccionario o lexicón de dominio específico realiza aportes significativos y mejora no sólo la categorización del enunciado sino también la interpretación dependiente del dominio donde se genera el enunciado.

Recordemos que todo discurso es contextualmente dependiente y, por lo tanto, no es posible realizar un análisis o desarrollar un modelo predictivo que no contemple nivel pragmático. De la misma manera, y en relación con este último aspecto, aplicar un diccionario o lexicón genérico a un corpus construido con enunciados de un dominio específico (financiero) arrojaría resultados inferiores ya que si bien el lexicón genérico se compone de una gran cantidad de términos variados, existe en algunos casos una diferencia significativa al usar un mismo término en un lexicón genérico y en uno de dominio específico. En otras palabras, supongamos que tomamos el término "bolsa", en un lexicón genérico tendrá tantas acepciones como contextos de aplicación, sin embargo si delimitamos el contexto a un dominio específico, habrá una única relación entre el término y la definición o acepción y su correlación con la opinión expresada en el enunciado.

Ahora bien, todo lo mencionado se re-significa cuando es posible contribuir al desarrollo de una solución en tiempo real para los inversores –los actores del ecosistema financiero– que realizan operaciones de *trading* en base a cómo la información impacta en los mercados. La habilidad de predecir este impacto desde el momento en que se postea o se publica la noticia o comentario en milisegundos es fundamental en la aplicación de lo que se conoce como *trading algorítmico*. Así, el hecho de automatizar ciertas decisiones de trading, la carga de órdenes y la administración de las mismas mediante algoritmos computacionales ha adquirido relevancia en las últimas décadas.

Nuestro trabajo se centra en el Análisis de Sentimiento desde una *perspectiva híbrida*, combinando el uso de un lexicón a medida con el uso de una estrategia de aprendizaje supervisado. En este método se utilizan diccionarios que describen la orientación semántica o polaridad (positiva o negativa) de la palabra. Sobre este diccionario se realizan cálculos mediante algoritmos computacionales de clasificación utilizando técnicas de *Machine Learning* para obtener la orientación del texto completo en base a la polaridad de las palabras incluidas en el mencionado diccionario [2, 19, 22, 24, 29, 32, 33, 36, 37, 39, 54].

El propósito aquí es presentar un lexicón diseñado dentro del dominio específico de las finanzas para Análisis de Sentimiento en español, a la vez que focalizarnos a nivel experimental y problematizar que un lexicón de terminología específica (FSAL) obtiene mejores resultados al compararlo con un lexicón de términos generales en español (SDAL) cuando son aplicados a tweets referidos al mercado financiero en Argentina tomando como metodología la perspectiva híbrida. En primer lugar, en este estudio, nos referiremos a una aproximación general sobre Análisis de Sentimiento y sus aplicaciones en diferentes áreas, principalmente discutiremos su rol en relación al Trading Algorítmico. En segundo lugar, proponemos y evaluamos un lexicón diseñado a medida en español rioplatense basado en noticias financieras extraídas de comentarios de Twitter referidos a los movimientos en el mercado financiero argentino (BYMA), donde explicamos la importancia de diseñar e implementar un lexicón a medida, creado para el dominio de las finanzas circunscripto al contexto económico de la Argentina. Por otra parte, nos referiremos de manera sucinta al enfoque teórico-metodológico que se basa en la extracción y clasificación manual de tweets determinando su polaridad positiva o negativa, así como la descripción general

de la composición semántica del lexicón de dominio específico combinadas con las técnicas de *Machine Learning*. Finalmente, mostraremos y discutiremos los resultados obtenidos de nuestro experimento basado en el estudio comparado de ambos lexicones (SDAL y FSAL).

2. Trabajos relacionados

2.1. Análisis de Sentimiento. Una aproximación general

En esta sección comenzaremos a delinear algunos conceptos fundamentales que constituyen la dinámica del presente trabajo.

En primer lugar, nos enfocaremos en Análisis de Sentimiento y posteriormente haremos hincapié en su relación con los mercados bursátiles así como los aspectos centrales que definen al Trading Algorítmico.

Dado que en sus orígenes, Análisis de Sentimiento, también conocido como Minería de Opinión, ha sido considerado una disciplina perteneciente a las ciencias de la computación y no a la lingüística, ha suscitado debate en cuanto a la definición misma así como el hecho de utilizar ambos términos como sinónimos. Por otra parte, mientras que en la industria Análisis de Sentimiento (de ahora en adelante SA) parece ser el término más apropiado y más utilizado, en la investigación académica se emplean ambos.

De acuerdo con la literatura revisada, una primera aproximación a la definición del término análisis de sentimiento fue realizada por Nashukawa y Yi [34]:

We define the task of our sentiment analysis as to find sentiment expressions for a given subject and determine the polarity of the sentiments. In other words, it is to identify text fragments that denote a sentiment about a subject within documents rather than classifying each document as positive or negative towards the subject.

mientras que *minería de opinión* apareció por primera vez en Dave *et al.*[9]:

the ideal opinion-mining tool would –process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good) En Pang y Lee [35] encontramos que Análisis de Sentimiento y minería de opinión son conceptos relacionados y afirman que si consideramos una interpretación más amplia ambos denotan el mismo campo de estudio, el cual puede ser considerado una sub-área del análisis de la subjetividad. Así, para estos autores análisis de sentimiento es: “the body of work which deals with the computational treatment of opinion, sentiment, and subjectivity in text. Such work has come to be known as opinion mining, sentiment analysis and/or subjectivity analysis”.

Por su parte, Bloom [2] en su tesis doctoral sostiene que “Sentiment Analysis is the task of having computers automatically extract and understand the opinions in a text”. En una conceptualización más reciente, Liu [28] define análisis de sentimiento como:

the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics,

and their attributes [...] From an application point of view, we naturally want to mine people's opinions and feelings towards any subject matter of interest, which is the task of sentiment analysis. More precisely, sentiment analysis is a field of study that aims to extract opinions and sentiments from natural language texts using computational methods.

Ahora bien, aún cuando hemos visto al SA, en sus inicios, como una disciplina dentro de las ciencias de la computación, en la actualidad, y dado la relevancia que han adquirido las opiniones en las redes sociales, donde los comentarios o posts reflejan la naturaleza y preferencias de los usuarios, se ha expandido al área de la ciencias sociales, políticas y económicas. En este sentido podemos decir que el SA se ha transformado en un caso especial del Procesamiento de Lenguaje Natural (NLP) estrechamente relacionado con el análisis semántico, semántica léxica, desambiguación de palabras, extracción de información y análisis del discurso [1, 4, 28, 55].

El creciente interés por las redes sociales ha marcado una nueva tendencia al momento de extraer y analizar contenidos potencialmente relevantes para la toma de decisiones en diferentes ámbitos o dominios, desde la colocación de productos en el mercado basado en opinión de consumidores, los comentarios sobre servicios, la medición de la imagen de celebridades, o de candidatos en política e incluso en finanzas para predecir la compra o venta de títulos. Así, la búsqueda de información que refleja la opinión pública y su impacto en la esfera política, social y económica, tanto organizaciones gubernamentales, entidades privadas, financieras como el área de negocios se han centrado en las aplicaciones de Análisis de Sentimiento.

En consecuencia, identificar el tono "positivo" o "negativo" de un texto (comentarios en blogs, noticias, posts en Twitter, reseñas, resúmenes, etc.) mediante un algoritmo se ha convertido en un tema crucial no sólo en el área de investigación sino también ha contribuido a rediseñar los negocios o modificar las tendencias en la vida social y política.

En los últimos años hubo una marcada tendencia a "minar" las opiniones publicadas por los usuarios de *Twitter* sobre candidatos presidenciales, con el propósito de predecir el resultado de las elecciones utilizando diferentes software de análisis de textos desarrollados para evaluar los componentes emocionales, cognitivos y estructurales de muestras textuales. Un claro ejemplo son las mediciones de intención de voto en EE.UU. y Alemania las cuales analizando el sentimiento sobre la opinión de los candidatos por parte de los usuarios de las redes sociales pudieron predecir con buenos niveles de certeza el resultado de las elecciones, tal es el caso de la elección presidencial de Barack Obama en 2008, donde en el programa televisivo "Hack the debate" transmitido en el mismo momento en que se producía el debate entre Barack Obama y John McCain, se invitaba a la audiencia para que participe posteando comentarios en Twitter [45].

Por otra parte, empresas, start-ups y corporaciones como Google³, Amazon⁴, IBM Watson⁵, Thomson Reuters⁶ y Bloomberg⁷, por nombrar algunas, han desarrollado e implementado plataformas, programas y aplicaciones para monitorear la opinión pública con la finalidad de detectar y analizar el sentimiento.

3 <https://cloud.google.com/natural-language/>

4 <https://aws.amazon.com/comprehend/>

5 <https://www.cognitiva.la/ibm-watson/>

6 <https://www.thomsonreuters.com/en/press-releases/2014/thomson-reuters-adds-unique-twitter-and-news-sentiment-analysis-to-thomson-reuters-eikon.html>

7 <https://www.bloomberg.com/professional/blog/can-get-edge-trading-news-sentiment-data/>

Una aplicación que nos interesa, en especial, y que ha motivado nuestra investigación es la que refiere a los mercados financieros que utilizan técnicas de *Machine Learning* (de ahora en adelante, ML) y Análisis de Sentimiento para generar modelos predictivos sobre los movimientos bursátiles, así como para identificar tópicos relacionados a las finanzas. En referencia a este tema, la Bolsa de Alemania desarrolló un modelo basado en indicadores de Análisis de Sentimiento relacionando noticias que generan un impacto directo sobre el precio de las acciones y aquellas que no lo hacen [15].

Metodologías similares se observan en mercados mucho más pequeños como por ejemplo en el de Croacia, donde se utilizan algoritmos basados en ML para clasificar noticias en "positivas" o "negativas" [51]. Estas noticias también son analizadas desde Twitter, estudiando detalladamente cuáles son las palabras más relevantes de los tweets seleccionados que tienen una correlación con el movimiento del precio de las acciones [11, 41]. Otros trabajos relacionados, recopilan el caso de la identificación de opiniones negativas y positivas en Twitter para predecir el movimiento de los índices en Wall Street, como Dow Jones, S&P 500 y NASDAQ [6, 14, 52, 53].

2.2. Análisis de Sentimiento, el Trading Algorítmico y la situación en Argentina

En esta sección revisaremos las relaciones entre Análisis de Sentimiento y Trading Algorítmico y su situación actual en los mercados financieros argentinos.

Consideramos Trading Algorítmico como la carga de órdenes de compra/venta de títulos financieros en los mercados de manera automática mediante algoritmos computacionales. El Trading Algorítmico representa un alto volumen del trading diario, calculándose en aproximadamente el 75% en los mercados más importantes del mundo. La velocidad con la que un algoritmo reacciona frente a un indicador técnico, ante un patrón encontrado, y el gran volumen de datos y cálculos, que este tipo de programas son capaces de procesar en tiempo real, superan de manera significativa la capacidad de reacción de un humano frente a estas variables. Una de las ventajas más notables se presenta cuando un algoritmo es capaz de detectar e interpretar una noticia o acontecimiento en el mismo instante que se está produciendo y así conocer su correlación con los movimientos de precios con el mercado de valores⁸.

Un claro ejemplo del potencial de Análisis de Sentimiento y Trading Algorítmico se presentó cuando un grupo de piratas informáticos se apoderó de la cuenta de Twitter de Associated Press (@AP) donde se informaba de la explosión de dos bombas en la Casa Blanca. En ese preciso momento diversos algoritmos pudieron interpretar la semántica de dicho enunciado y comenzaron a vender diferentes títulos financieros provocando pérdidas considerables⁹.

Ahora bien, para que esto suceda, los algoritmos basados en análisis de sentimiento deben poder monitorear noticias en tiempo real, determinar si tienen una polaridad positiva o negativa y establecer cómo repercuten las mismas en los mercados financieros argentinos. Con toda esa información deben tener la capacidad de tomar una decisión de compra/venta de un título y enviarla automáticamente al mercado.

⁸ Utilizamos el término mercado de valores en el mismo sentido que mercados financieros.

⁹ H. Moore and D. Roberts, *AP Twitterhack causes panic in Wall Street and sends Dow plunging*. The Guardian Online, UK, 2013. Disponible en: <http://www.theguardian.com/business/2013/apr/23/ap-tweet-hack-wall-street-freefall>. Consultado: 2015/11/1.

En relación con esta temática, en Argentina ya se ha comenzado a operar con la tecnología de Trading Algorítmico con éxito y con un volumen de operaciones significativo, particularmente en BYMA (Bolsas y Mercados Argentinos) o el Mercado de Futuros (ROFEX). Razón por la cual, resulta fundamental combinar estas estrategias para el análisis de noticias financieras utilizando, como se ha mencionado técnicas de Análisis de Sentimiento para poder predecir las reacciones de las opiniones de los usuarios de las cuentas de Twitter y su impacto en el Mercado, es decir, en un contexto situacional específico.

3. Nuestra aproximación

Basados en las consideraciones anteriores y el marco contextual en el que nos circunscribimos, es preciso mencionar que, todos estos tópicos y aplicaciones han sido la fuente que ha motivado nuestra investigación en lo que refiere a los mercados financieros en Argentina, específicamente al estudio de BYMA.

Asimismo, tomando como referencia los trabajos realizados por diversos autores, por razones metodológicas que hacen al presente estudio, es fundamental resaltar que existen dos grandes enfoques de Análisis de Sentimiento. Por una parte, aquellos basados en lexicones que calculan la orientación semántica de un texto a partir de palabras o frases presentes en el mismo, las cuales han sido clasificadas de antemano. Por otra parte, los centrados en aprendizaje supervisado que utilizan clasificadores entrenados a partir de textos previamente clasificados [24, 29, 32, 33, 36, 37, 41, 54]. De modo que, a efectos de desarrollar nuestro experimento consideramos fundamental integrar los enfoques anteriores en lo que se denomina *estrategia híbrida*. Así, hemos adoptado la aproximación de combinar un lexicón a medida, dependiente del dominio de las finanzas, cuyas palabras son indicadores de sentimiento extraídas de las opiniones de usuarios en las redes sociales, en este caso Twitter, y que su vez expresan la polaridad (positiva/negativa) contextual de estos enunciados, y técnicas de *Machine Learning* con algoritmos de clasificación para crear un modelo supervisado en el dominio específico.

Con respecto a los lexicones, cabe destacar que si bien existen diferentes propuestas en inglés, como por ejemplo SentiWordNet [13] y LIWC [42] considerados como los más populares, en el caso de otros idiomas -el español en particular- existen avances con variantes en opciones de lexicones genéricos¹⁰ [30, 33]. Un ejemplo de ello es el Spanish DAL [10], un lexicón de propósito general para el análisis de sentimiento de textos en Español, que consiste en la réplica del Modelo de Whissell [46, 47] para la clasificación de palabras y textos en tres dimensiones: nivel de agrado, activación (activo/pasivo), imaginabilidad. Teniendo en cuenta estas observaciones se ha decidido diseñar un lexicón de términos especializados a medida para el área de finanzas aplicado al mercado de valores de Argentina y que constituye el tema central de nuestra investigación.

¹⁰ Cruz et. al. (2008) diseñó un sistema de clasificación no supervisado de documentos en español (véase: F. Cruz, J.A. Troyano, F. Enriquez and J. Ortega. "Clasificación de documentos basada en la opinión: experimento con un corpus de críticas de cine en español", en *Procesamiento del Lenguaje Natural*, (41), 73-80, 2008). También, J. Brooke, M. Tofiloski and M. Taboada, "Cross-linguistic Sentiment Analysis: From English to Spanish", in *Proceedings of RANLP 2009*. Cf. S. Mohamad, F. Bravo-Marquez, M. Salameh and S. Kiritchenko (2018). SemEval-2018 Task 1: Affect in Tweets in *Proceedings of the 12th International Workshop on Semantic Evaluation*, pp. 1-17. Association for Computational Linguistics.

4. Lexicón de dominio específico diseñado para los Mercados Financieros Argentinos

Observemos el siguiente enunciado extraído de distintas fuentes periodísticas nacionales, el cual despertó reacciones negativas en el público consumidor:

(1) YPF aumentó el precio de la nafta en un 4,5%

Al analizar las opiniones o reacciones -posts en Twitter- desde diferentes contextos situacionales, se pudo detectar que imprimían un impacto diferente en los resultados. Por un lado, consideradas desde un contexto político y socio-cultural más amplio, se observó un alto índice de sentimiento negativo, el cual surgió del estudio de la composición semántica referida al aumento de precio al público consumidor, aun cuando el valor de las acciones de YPF ese mismo día subió considerablemente, representando un efecto positivo para los inversores. Para estas mismas opiniones analizadas desde otro contexto más específico, esto es, desde el mundo de las finanzas y mercados, y de acuerdo con los especialistas en esta materia, otra posible interpretación admitiría que si bien el sentimiento "popular" sobre esa noticia fue negativo, la repercusión en el mercado de valores fue altamente positiva ya que si aumenta el precio de la nafta, se espera que YPF aumente sus ingresos, por lo cual las acciones que cotizan en el mercado aumentan su valor.

Otro ejemplo interesante, se manifiesta en el enunciado descrito a continuación:

(2) Espectacular éxito del Gobierno en el allanamiento de las cuevas financieras para bajar el dólar Blue

Claramente, este enunciado tiene una estructura semántica positiva fácilmente detectable por cualquier algoritmo, aunque la repercusión en el mercado fue sensiblemente negativa. Aquí operaría el mismo principio de análisis mencionado con anterioridad, por lo cual podemos decir que la tendencia "positiva" o "negativa" de un enunciado además de depender del estudio semántico es contextualmente dependiente.

Dada entonces las características en los enunciados, los contextos en los que la noticia y comentarios se generan, la repercusión sea esta positiva o negativa, entre otros factores, y debido a que el mundo de las finanzas posee un lenguaje técnico muy específico y, tanto la interpretación como la clasificación de las noticias en positivas o negativas debe ser analizada por expertos en la materia, el Trading Algorítmico basado en Análisis de Sentimiento en nuestro modelo sienta sus bases en un lexicón compuesto por un diccionario de palabras financieras específicas para el Mercado Argentino diseñado por especialistas en Lingüística y Análisis del Discurso y un corpus de noticias clasificadas por analistas financieros, economistas, expertos en política económica, entre otros.

4.1. Creación del Lexicón FSAL. Características

Dada la filosofía colaborativa y multidisciplinaria de nuestro proyecto de investigación, desde en el enfoque teórico-metodológico lingüístico se ha diseñado un lexicón basado en semántica especializada en finanzas que reflejan una *apreciación o juicio valorativo*. El mismo incluye una selección de las palabras más críticas en este contexto para determinar la carga de sentimiento de un texto (enunciado) con contenido económico-financiero. Cabe mencionar que, el lexicón compuesto hasta el momento, de 5320 palabras, se encuentra en una etapa preliminar, el cual será ampliado y modificado de acuerdo con los requerimientos y evolución del proyecto.

Fase 1. Extracción de unidades léxicas

Como primer paso en la creación de nuestro lexicón, se extraen del corpus las palabras (unidades léxicas) más relevantes para el estudio de Análisis de Sentimiento. Este corpus está compuesto, a su vez, por una selección representativa de *tweets*. Luego, a cada una de esas palabras se les asigna un peso que refleja su polaridad positiva o negativa. Dicho peso se encuentra en un rango discreto entre [-5,5], siendo [-5] la carga más negativa y [5] la carga más positiva.

Fase 2. Selección y categorización de las unidades léxicas

El lexicón está conformado por palabras "llenas" o "léxicas" [17, 18], aquellas que poseen un contenido descriptivo inherente, es decir, que expresan una idea o conceptos y pueden describirse mediante rasgos semánticos, donde hemos excluido palabras "vacías" o "funcionales" [17, 18] las cuales no expresan una idea y su significado es fundamentalmente gramatical que se precisa en relación con otras palabras que transmiten información categorial o entre las que establecen relaciones sintácticas. Por otra parte, consideramos aquellas palabras o unidades léxicas que se "activan" en una situación comunicativa especializada, en este caso la interacción discursiva en los mercados financieros. Así, la información semántica que compone el significado especializado se distingue del no especializado por su configuración. En este sentido, observaremos que una de las propiedades centrales del término especializado refiere a la mayor especificidad referencial que se desprende de la cantidad y tipo de información activada. Asimismo, consideraremos que la configuración semántica de estos términos es más densa y más compleja ya que presentan varios niveles de relación de imbricación entre las predicaciones activadas [7].

Merece aquí una especial mención el tratamiento de las palabras o unidades léxicas de la clase sustantivos. Desde nuestra aproximación, clasificamos los sustantivos en tres categorías diferentes:

- » (a) aquellos que no poseen carga semántica "afectiva"¹¹. Por ejemplo: casa, árbol, perro. Estos sustantivos no forman parte de nuestro lexicón;
- » (b) aquellos que claramente poseen una carga semántica afectiva. Por ejemplo: abundancia, crisis. Estos sustantivos se incorporan a nuestro lexicón con un peso apenas superior al neutro; y
- » (c) aquellos sustantivos que representan "Drivers de Mercado". Por ejemplo: default, China, corralito, Dólar Blue, petróleo. Estos términos no forman parte del lexicón de finanzas (FSAL) si no que son tratados de manera independiente y se constituyen en componentes esenciales para la creación de un segundo lexicón, denominado Lexicón de Drivers (LDR)¹², considerando los modificadores que cada driver posee. Diremos que un "driver de mercado, es aquel factor que ejerce una influencia en los movimientos bursátiles de manera directa en un momento determinado y es contextualmente dependiente". En los ejemplos a continuación:
 - » (3) La crisis en China afectó los mercados de Wall Street
 - » (4) La baja de los precios del petróleo derribo al índice Dow Jones

donde "China" y "petróleo" constituyen estas palabras o unidades léxicas que se activan en una situación comunicativa especializada, esto es, el contexto de situación en el que se originan y en el que, a su vez, generan un impacto o ejercen algún tipo de influencia.

¹¹ El marco de la Teoría de la Valoración se orienta hacia un recorrido de los dominios semánticos y cómo estos operan en el discurso. Las categorizaciones agrupan estructuras gramaticales bajo un denominador discursivo semántico. Afecto: su valor se construye como cualidades (adjetivos: Estoy contento...), procesos (verbos: Me agrada...), adjuntos (felizmente) y en algunos casos con entidades virtuales (sustantivos) mediante nominalizaciones (felicidad) (Véase. P.R.R. White, "Appraisal", en Verschueren, J. and Östman (eds.), Handbook of Pragmatics Highlights, Amsterdam: John Benjamins Publishing Company, 2011.)

¹² El Lexicón de Drivers es parte de un trabajo de investigación en curso.

Fase 3. Contenido semántico evaluativo

Además de la categorización de la Fase 2, decidimos trabajar en un segundo nivel de análisis, que se corresponde con el contenido semántico evaluativo que las palabras pueden tener. Sostenemos que las palabras tienen un sesgo implícito de valoración o modalización y de acuerdo a su ordenamiento y selección pueden brindar perspectivas o maneras de valorar lo que se dice, esto es, asignarle un valor al enunciado. En este nivel tomamos como punto de partida los postulados de la Teoría de la Valoración [30, 31, 48, 49], desarrollada dentro del enfoque de la Lingüística Sistemática Funcional [12, 20], que se ocupa particularmente del lenguaje (la expresión lingüística) de la valoración, la actitud y la emoción, y del conjunto de recursos que explícitamente posicionan de manera interpersonal las propuestas y proposiciones textuales.

La Teoría de la Valoración brinda un análisis de aquellos significados mediante los cuales un texto o un enunciado (en nuestro caso comentarios sobre noticias financieras de los mercados financieros argentinos extraídos de Twitter) transmiten una valoración o evaluación positiva o negativa. Estos "recursos" que crean significados se agrupan en lo que se denomina "lenguaje evaluativo" sobre la base de que todos ellos constituyen un medio por el cual el actor, esto es el hablante/productor del mensaje o comentario, transmite la opinión o posición que adopta respecto de un fenómeno dado, es decir una entidad, un suceso, o estado de las cosas [49].

Resumiendo, de acuerdo a lo mencionado, nuestro lexicón se compone de términos especializados en economía y finanzas en su acepción más completa "lexicón económico-financiero relacionado con terminología del discurso político y jurídico" con los que se puede determinar la actitud, es decir, significados por los cuales los textos/hablantes atribuyen un valor o una evaluación intersubjetiva a los participantes y los procesos, ya sea relacionados con respuestas emocionales o con sistemas de valores culturalmente determinados (afecto, juicio, apreciación).

4.2. Sistema de filtrado

A fin de eliminar aquellas palabras que no aportan un significado contextual específico al dominio en el que estamos trabajando, hemos decidido aplicar un sistema de exclusión-inclusión de términos (filtrado), donde en primer lugar, se eliminaron aquellos términos que no se corresponden con la temática principal por sus características de generalidad, y en segundo lugar, aquellos que tienen un significado gramatical, que se precisa solamente en relación con otras palabras de las que transmiten información categorial o entre las que establecen relaciones sintácticas.

Consideramos, entonces, excluir sustantivos comunes que no tienen un significado semántico con una valoración intrínseca, esto es, aquellos que consideramos "neutros", por ejemplo: gente, nombre, uso, etc.; así como también sustantivos propios como nombres de personas, países, lugares geográficos (*v. supra* 3.1.), además de gentilicios, siglas, adverbios de lugar y tiempo, el verbo "haber" en todas sus formas (desinencias verbales) y los que denominamos verbos de estado (ser, estar). Por otra parte, se excluyen determinantes (artículos, demostrativos y posesivos) y cuantificadores, conjunciones coordinantes y subordinantes, preposiciones (plenas, semiplenas, y gramaticalizadas).

En el caso del lexicón de términos¹³ especializados que se presenta en nuestro estudio, para evitar la repetición, omisión e incorporación de un término que no se corresponda con las categorías

¹³ Utilizaremos "palabra, unidad léxica o término" de manera intercambiable.

gramaticales o no contengan en sí mismo una carga valorativa, se ha desarrollado un primer prototipo de asistencia para el curado del lexicón al que denominamos "Lexicon Checker" que nos permite revisar aquellos términos que se encuentran en los tweets relacionado con noticias o comentarios en finanzas que componen el corpus y agregar al listado los términos que fueron omitidos en una primera selección manual, además de asignarle el peso correspondiente de acuerdo a la escala diseñada especialmente, cuyo rango en valores positivos es de [1] a [5] y en valores negativos de [-1] a [-5].

4.3. El Corpus

Con el fin de construir nuestro corpus de análisis hemos recolectado *tweets* cuyo contenido es significativamente relevante para los mercados financieros argentinos. Estos *tweets* provienen de una selección de cuentas de Twitter de comunidades de confianza conformada por expertos de reconocida trayectoria en el ámbito económico-financiero. Además, dichos *tweets* fueron clasificados de manera colaborativa por especialistas del dominio de las finanzas. La plataforma diseñada para este propósito, consta de una interfaz web, donde se cuenta con un usuario y contraseña para ingresar. Con esta herramienta, los expertos tienen la oportunidad de ver cada uno de los enunciados y determinar si poseen una carga positiva o negativa en relación al impacto que tiene dicho enunciado en el Mercado. Por otra parte, hemos desechado aquellos *tweets* cuyo sentimiento no era posible determinar o eran de carácter neutro. Como resultado, se obtuvo un "corpus equilibrado" (*balanced-corpus*) constituido por 500 *tweets* (250 con carga positiva y 250 con carga negativa).

4.4. Algoritmos de clasificación

A efectos de evaluar el desempeño de nuestros algoritmos en cuanto al grado de certeza en la clasificación de *tweets* utilizando tanto el lexicón genérico como nuestro lexicón de términos especializados aplicados a un dominio específico, hemos recurrido a diferentes instancias.

Los algoritmos de clasificación que hemos utilizado son: *Naive Bayes*, *Random Forest* y *Support Vector Machines*. Tanto el procesamiento de datos y procesamiento de textos, como la implementación de los algoritmos, fueron realizados con la plataforma de análisis estadístico R¹⁴.

5. Discusión y resultados

Sostenemos como hipótesis central de nuestro trabajo que *un lexicón hecho a medida para el mercado financiero de Buenos Aires (FSAL) permite obtener mejores resultados de clasificación que si se utiliza un lexicón de propósito general (SDAL)*. En este sentido, hemos llevado a cabo un estudio comparativo para evaluar el rendimiento de nuestro lexicón de dominio específico en español rioplatense diseñado para el ecosistema financiero de Argentina, lo cual permitiría optimizar las estrategias de Trading Algorítmico.

Así, el procedimiento consiste en, dado un corpus de *tweets* determinado, previamente clasificado por expertos y manteniendo como constante nuestros algoritmos de aprendizaje supervisado basados en Machine Learning, variar el lexicón, esto es FSAL y SDAL, lo cual nos permitirá a nivel comparativo determinar la efectividad de lexicones así como el grado de certeza en la clasificación de nuestro lexicón (FSAL) al contrastarlo con el de propósito general (SDAL¹⁵) basado en la réplica del Modelo de Whissell [46, 47].

¹⁴Disponible en: <http://r-project.org>

¹⁵ La propuesta realizada por la UBA (Dell' Amerlina Ríos y Gravano, 2013), se centra en la réplica del Modelo de Whissell (1986, 1989)

Con el propósito de evaluar la efectividad del análisis se aplica la técnica de validación cruzada (X-Validation, con $X = 10$) comúnmente utilizada en pruebas en el dominio del aprendizaje supervisado. Se inicia con un corpus de textos (*tweets*) que han sido previamente clasificados en positivos o negativos. A diferencia de otros enfoques que también utilizan aprendizaje supervisado, en nuestro caso, para el armado del vector de características (feature vector) sólo se tienen en cuenta las palabras existentes en el lexicón. El clasificador resultante se emplea para predecir el sentimiento de los textos en la partición de prueba. Finalmente, se reportan los aciertos y errores tanto para los textos negativos como los positivos.

La Tabla I muestra los resultados que se corresponden a la ejecución de validación cruzada. Para ello se utiliza la implementación de cross-validation provista por la librería Caret de R¹⁶. Cada ejecución utiliza una partición diferente del corpus, siempre en una relación 90% para entrenamiento, 10% para prueba.

	SDAL	FSAL
Naive Bayes	65,6%	75,5%
Random Forest	71,5%	80,5%
SVM	64%	77%

Tabla I: Resultados

Así la Tabla I, muestra que, para todos los algoritmos utilizados, nuestro lexicón FSAL resultó en una mejor clasificación que el SDAL utilizando como medida de rendimiento en la matriz de confusión la precisión total (overall accuracy).

Dado que nuestra estrategia de clasificación supervisada sólo considera las palabras presentes en el lexicón sería posible inferir que la diferencia en resultados se deba a una diferencia en el número de palabras de cada lexicón que son relevantes al corpus. Para descartar esta posibilidad, hemos analizado dicha variable, obteniendo un valor de 362 palabras del SDAL presentes en el corpus, contra 391 del FSAL. Es decir que 12.87% de las palabras del SDAL están presentes en el corpus de estudio, mientras que para el FSAL el porcentaje de palabras es de 15.95%. Si bien existe una diferencia en el número de palabras, entendemos que no es suficiente para explicar la diferencia en los resultados. Complementariamente, la Tabla II ofrece un resumen de la presencia de palabras de cada lexicón en los tweets del corpus. Se observa que 31 tweets del corpus no tienen ninguna palabra (columna 0) del SDAL, mientras que sólo 15 no tienen ninguna palabra del FSAL.

Palabras	0	1	2	3	4	5	6	7+
SDAL	31	45	55	39	29	25	8	4
FSAL	15	62	82	40	21	11	3	2

Tabla II: Presencia de los lexicones en los tweets

basado en la clasificación de palabras y textos en inglés en tres dimensiones: nivel de agrado, activación (activo/pasivo), imaginabilidad (pleasantness, activation, imagination). Para conocer más sobre este experimento, véase: Dell' Amerlina Ríos, Matías y Agustín, Gravano (2013): "Spanish DAL: A Spanish Dictionary of Affect in Language". Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 21–28, Atlanta, Georgia, Association for Computational Linguistics.

¹⁶ Librería Caret R (Classification and Regression Training). Disponible en: <https://cran.r-project.org/web/packages/caret/index.html>

La Tabla III, a continuación, muestra una lista de las palabras de cada lexicón que aparecen al menos 10 veces en el corpus. Puede notarse que son menos (y más específicas a finanzas) las palabras del FSAL cuya frecuencia impacta en el resultado. Asimismo, puede verse que el SDAL (por su carácter de lexicón de propósito general) asigna una carga valorativa a palabras que no la tienen necesariamente en el dominio de finanzas como por ejemplo: una, pero, vía.

FSAL		SDAL	
Palabra	Frecuencia	Palabra	Frecuencia
no	87	dólar	43
más	33	más	33
inflación	31	bolsa	30
baja	29	una	30
default	14	gobierno	25
deuda	14	peso	18
problema	14	mercado	16
sube	13	deuda	14
caída	11	país	14
		problema	14
		economía	12
		vía	12
		caída	11
		pago	11
		año	10
		emisión	10
		hasta	10
		ley	10

Tabla III: Palabras con frecuencia 10 o más

7. Conclusiones

En el presente trabajo se han mostrado las relaciones que existen entre Análisis de Sentimiento y Trading Algorítmico y cómo un lexicón diseñado a medida de terminología específica (lenguaje especializado) utilizado en los mercados argentinos mejora los resultados si lo comparamos con uno de propósito general para la clasificación de comentarios de noticias en positivas y negativas.

Hasta el momento hemos logrado construir el lexicón a medida (FSAL), y desarrollado los algoritmos que clasifican con un nivel de certeza aceptable los enunciados en positivos y negativos.

Cabe señalar que este trabajo representa uno de los aspectos centrales de nuestro proyecto de investigación. Así, es imprescindible enfatizar en lo que respecta al análisis de textos “entiéndase por ello el discurso económico-financiero que incluye noticias, opiniones, comentarios en las redes sociales, más específicamente en Twitter” que continuamos perfeccionando y entrenando los algoritmos a nivel semántico.

Finalmente, es preciso señalar que en trabajos futuros nos enfocaremos por una parte, en el tratamiento y comportamiento semántico de los términos en relación con la *interfaz sintáctica* y por otra parte, estaremos evaluando el rendimiento de modelos más modernos y robustos basados en Aprendizaje Profundo (*Deep Learning*), tales como Redes Neuronales Convolucionales (*Convolutional Neural Networks*) y Aprendizaje por Refuerzo (*Reinforcement Learning*).

Referencias

- » S. Aghaian and P. Kolm (2017). Financial Sentiment Analysis Using Machine Learning Techniques in *International Journal of Investment Management and Financial Innovations*, 3(1):1-9.
- » M. Z. Asghar, Khan, A., Ahmad, S., Qasim, M. and I. A. Khan (2017). Lexicon-enhanced sentiment analysis framework using rule-based classification scheme in *PLOS One*, 12 (2).
- » J. Bollen, H. Mao, and X. Zeng (2011). “Twitter mood predicts the stock market” in *Journal of Computational Science*, 2(1):1-8.
- » Bloom, K. (2011). *Sentiment Analysis based on appraisal theory and functional local grammars* (Ph. D. Thesis). Graduate College, Illinois Institute of Technology, Chicago. Recuperada de: https://www.scss.tcd.ie/Khurshid.Ahmad/Research/bloom_dissertation.pdf
- » E. Brown (2012). “Will twitter make you a better investor? A look at sentiment, user reputation and their effect on the stock market”, in *SAIS Proceedings*.
- » R. Chen and M. Lazer (2011) “Sentiment analysis of twitter feeds for the prediction of stock market movement”.
- » E. Ciapuscio, A. Adelstein, L. Brandani, L. Ferrari, S. Gallardo, S. Kornfeld, I. Kuguel y G. Resnik (2009). *De la palabra al texto. Estudios lingüísticos del español*. Buenos Aires: Eudeba.
- » N. P. Cruz, Taboada, M., Mitkov, R. (2016). A machine-learning approach to negation and speculation detection for sentiment analysis. *J. Assoc. Inf. Sci. Technol.* **67**, 2118–2136.
- » K. Dave, S. Lawrence and D. M. Pennock. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pp. 519–528, New York, NY, USA. ACM.
- » M. G. Dell’ Amerlina Ríos and M. Gravano (2013). “Spanish DAL: A Spanish Dictionary of Affect in Language”, in *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Atlanta, Georgia, Association for Computational Linguistics, pp. 21–28.
- » A. Devitt and K. Ahmad (2007). “Sentiment Polarity Identification in Financial News: A Cohesion-based Approach”, Prague, Czech Republic: Association for Computational Linguistics.
- » S. Eiggins (2002). *Introduction to Systemic Functional Linguistics*. London and New York: Pinter Publishers.
- » A. Esuli and F. Sebastiani (2006). “SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining,” in *Proceedings of the 5th Conference of Language Resource Evaluation*, pp. 417–422.
- » R. Feldman, B. Rosenfeld, R. Bar-Haim and M. Fresko (2011). “The stock sonar—sentiment analysis of stocks based on a hybrid approach”, in *Twenty-Third IAAI Conference*.

- » P. Finter, A. Niessen-Ruenzi and S. Ruenzi (2012) "The Impact of Investor Sentiment on the German Stock Market", *Springer*, vol. 82, no. 2, pp.133-163.
- » J.A. García Díaz, Salas-Zárate, M.P., Hernández-Alcaráz, M.L., Valencia-García, R. y J. M. Gómez-Berbíz. (2018). "Machine Learning Based Sentiment Analysis on Spanish Financial Tweets". In: Rocha Á., Adeli H., Reis L.P., Costanzo S. (eds) Trends and Advances in Information Systems and Technologies. WorldCIST'18 2018. Advances in Intelligent Systems and Computing, vol 745. Springer, Cham.
- » M. Giammatteo y H. Albano (2006). *¿Cómo se clasifican las palabras?* Buenos Aires: Littera Ediciones.
- » M. Giammatteo y H. Albano (2009). *Lengua, Léxico, Gramática y Texto. Un enfoque para su enseñanza basado en estrategias múltiples*. Buenos Aires: Biblos.
- » M. Ghiassi and S. Lee (2018). "A domain transferable lexicon set for Twitter Sentiment Analysis using a supervised Machine Learning Approach" in *Expert System with Applications*, v.106, pp. 197-206. Elsevier.
- » M.A.K. Halliday (1994). *An Introduction to Functional Grammar*. London: Edward Arnold Publishers.
- » T. Hendershott, C.M. Jones and A. Menkveld (2011). "Does Algorithmic Trading Improve Liquidity?" in *The Journal of Finance*, vol. 66, no. 6, pp. 1-34.
- » C. Kaushik and A. Mishra (2014). "A scalable, lexicon based technique for Sentiment Analysis", in the *International Journal in Foundations of Computer Science & Technology (IJFCST)*, vol.4, no.5.
- » C.S.G. Khoo, A. Nourbakhsh and J.C. Na (2014). "Sentiment analysis of online news text: A case study of appraisal theory", in *Online Information Review*, 36(6).
- » O. Kolchyna, Souza, T.P., Treleaven, P, and Aste, T. (2016) Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination in the *Handbook of Sentiment Analysis in Finance*. Mitra, G and Yu, X (Eds.). ISBN 1910571571.
- » Kordonis, J., Symeonidis, S., Arampatzis, A. (2016). Stock price forecasting via sentiment analysis on Twitter. In: Proceedings of the 20th Pan-Hellenic Conference on Informatics - PCI 2016. pp. 1–6. ACM Press, New York.
- » S. Krishnamoorthy (2017). "Sentiment Analysis of Financial News Articles using performance indicators" in *Knowledge and Information System*. Springer International Publishing, 1-22.
- » A. Kumar, Sethi, A., Akhtar, S., Ekbal, A., Biemann, C., Bhattacharyya, P., (2017). "IITPB at SemEval-2017 Task 5: Sentiment Prediction in Financial Text" in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017).
- » B. Liu (2015). *Sentiment Analysis. Mining Opinions, Sentiments and Emotions*. New York: Cambridge University Press.
- » H. Mao, P. Gao, Y. Wang and J. Bollen (2014). "Automatic Construction of Financial Semantic Orientation Lexicon from Large-Scale Chinese News Corpus", in *7th Financial Risks International Forum*, vol. 20, no.20, pp. 1-18.
- » J. R. Martin and P. R. R. White (2015). *The Language of Evaluation, Appraisal in English*, London and New York: Palgrave Macmillan.
- » J. R. Martin and D. Rose (2003). "Appraisal", in *Working with Discourse. Meaning beyond the clause*. London: Continuum.
- » A. Moreno-Ortiz y C. Pérez Hernández (2013). "Análisis de Sentimiento basado en lexicones de mensajes de Twitter en Español, en *Procesamiento del Lenguaje Natural*, Revista nº 50, pp. 93-100.
- » A. Muhammad, N. Wiratunga, R. Lothian and R. Glassey (2013). Domain-Based Lexicon Enhancement for Sentiment Analysis, in *Proceedings of the BCS SGAI Workshop on Social Media*, Cambridge, UK, December 10.
- » T. Nashukawa and J. Yi (2003). Sentiment Analysis Capturing Favourability Using Natural Language Processing. *K-CAP'03*, October 23–25, 2003, Sanibel Island, Florida, USA.

- » B. Pang and L. Lee (2008). "Opinion mining and sentiment analysis", in *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp.1–135.
- » V. Perez-Rosas, C. Banea and M. Rada (2012). "Learning Sentiment Lexicons in Spanish", in *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, vol.12, pp. 3077-3081. European Language Resources Association (ELRA).
- » C. Potts (2010). "Sentiment lexicons", in *Linguist 287 / CS 424P: Extracting Social Meaning and Sentiment*.
- » T. Rao and S. Srivastava (2012). "Analyzing stock market movements using twitter sentiment analysis", in *Proceedings of ASONAM 2012*, pp. 119-123.
- » Saleiro, P., Mendes Rodrigues, E. , Soares, C. y E. Oliveira (2017). FEUP at SemEval-2017 Task 5: Predicting Sentiment Polarity and Intensity with Financial Word Embeddings, en *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, pages 904–908.
- » H. K. Sul, A.R. Dennis and L. Yuan (2014). "Trading on Twitter: The Financial Information Content of Emotion in Social Media", in *System Sciences (HICSS), 47th Hawaii International Conference*.
- » M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede (2011) "Lexicon-Based Methods for Sentiment Analysis", in *Association for Computational Linguistics*, vol. 37, no. 2, pp. 267-307.
- » Y. R. Tausczik and J. W. (2010). "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods", in *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54.
- » P. Tedlock (2007). "Giving content to investor sentiment: The role of media in the stock market", in *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168.
- » P. Tedlock, C. Maytal Saar-Tsechansky, and S. Macskassy (2008). "More than words: quantifying language to measure firms' fundamentals", in *The Journal of Finance*, 63(3):1437–1467.
- » A. Tumasjan, T.O. Sprenger, P. G. Sandner and I. M. Welp (2010). "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, vol.10, pp. 178-185.
- » C. Whissell, M. Fournier, R. Pelland, D. Weir, and K. Makarec (1986). "A dictionary of affect in language" in *IV. Reliability, validity, and applications. Perceptual and Motor Skills*, vol. 62, no. 3, pp. 875–888.
- » C. Whissell (1989). "The dictionary of affect in language", in Robert Plutchik and Henry Kellerman (ed.) *Emotion: Theory, research, and experience*, pp.113–131.
- » P.R.R. White (2006). "Evaluative Semantics and Ideological Positioning in Journalistic Discourse", in *Image and Ideology in the Mass Media*, Lassen, I. (ed.), Amsterdam and Philadelphia: John Benjamins, pp. 45- 73.
- » P.R.R. White (2015). "Appraisal Theory", in *The International Encyclopedia of Language and Social Interaction*, Tracy, K. (ed.), Ilie, C. and Sandel, T. (Associated Editors), London: John Wiley and Son, Inc.
- » M. Zaki and M. Wagner (2014). *Data Mining and Analysis. Fundamental Concepts and Algorithms*. U.S.A: Cambridge University Press.
- » A. Željko, N. Ljubešić and M. Tadić (2010). "Towards Sentiment Analysis of Financial Texts in Croatian", in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. European Language Resources Association.
- » L. Zhang (2013). "Sentiment Analysis on Twitter with stock price and significant keyword correlation", Ph. D. dissertation.
- » W. Zhang, and S. Skiena (2010). "Trading strategies to exploit blog and news sentiment", in *The 4th International AAAI Conference on Weblogs and Social Media*.
- » L. Zhang, R. Ghosh, M. Dekhil, M. Hsu and B. Liu (2011). "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis", HP Laboratories, HPL-2011-89.

- » X. Zhang (2016). "Positive Discourse Analysis of Attitude System in Economic News", in *International Journal of Arts and Commerce*.

Sobre los autores

Juan Pablo Braña. Es coordinador y docente de la Diplomatura en Análisis de Datos para Negocios, Finanzas e Investigación de Mercados de la Universidad Abierta Interamericana. Es investigador del Centro de Altos Estudios Tecnológicos e Informáticos (CAETI). Su foco actual de investigación es el Trading Algorítmico.

Alejandra M. J. Litterio. Es Profesora de Inglés, Traductora Literario-Técnico Científica en Inglés y maestrando en la Maestría en Análisis del Discurso en Facultad de Filosofía y Letras de la Universidad de Buenos Aires. Profesora de Procesamiento de Lenguaje Natural y Sentiment Analysis de la Diplomatura en Análisis de Datos para Negocios, Finanzas e Investigación de Mercados de la Universidad Abierta Interamericana. Es investigadora del Centro de Altos Estudios Tecnológicos e Informáticos (CAETI). Su foco actual de trabajo es desarrollo del lexicón en finanzas en español y lexicón de driver de mercados, y aplicación de la Teoría de la Valoración para Análisis de Sentimiento.

Alejandro Fernández. Es Licenciado en Informática de la Universidad Nacional de La Plata, y Dr. en Ciencias (Dr. rer. nat.) de la FernUniversität Hagen, de Alemania. Es profesor adjunto en la Facultad de Informática de la Universidad Nacional de La Plata y en la Universidad Abierta Interamericana. Es investigador asistente de la Comisión de Investigaciones Científicas de la Provincia de Buenos, en el Centro Asociado LIFIA. Su foco actual de trabajo el desarrollo de tecnologías para la gestión del conocimiento en red.

