



Facultad de Ciencias Económicas y Empresariales
Grado en Administración y Dirección de
Empresas

Trabajo Fin de Grado

*¿Se puede medir el impacto del Brexit
sobre la cotización del Banco Santander
a través de los sentimientos de Twitter?*

Análisis desde una perspectiva económica y analítica

Alumna: **Blanca Castro Madruga**

Director: María Coronado Vaca

Madrid, junio 2020

Resumen

El Banco Santander lleva operando en el Reino Unido desde 2004. Su presencia en el país le ha hecho vulnerable ante acontecimientos como el *Brexit*. Por ello, este trabajo ha abordado la posibilidad de calcular el impacto real del *Brexit* sobre el Banco Santander a través de dos variables: la cotización y los sentimientos de Twitter.

Diversos estudios han probado que el precio bursátil recoge tanto la influencia de cualquier proceso histórico como el *Brexit*, como de las opiniones que se comparten en redes sociales. Con este propósito se han utilizado técnicas de *data analytics* a través de lenguajes de programación como R y Python para la extracción de datos y medición de los sentimientos. Tras un análisis empírico, exploratorio y de sentimientos se ha concluido que no existe ningún tipo de correlación entre el precio de mercado del Banco Santander y los sentimientos de Twitter respecto al *Brexit*. Durante el estudio, se han identificado distintas limitaciones relacionadas con la ética en el uso de datos y la exactitud en el análisis de sentimientos con el propósito de facilitar el trabajo de futuras investigaciones.

Palabras clave

Banco Santander, *Brexit*, cambios regulatorios, cotización, noticias, Twitter, análisis de sentimientos, correlación

Abstract

Banco Santander has been operating in the UK since 2004. Its presence in the country has made it vulnerable to events such as Brexit. This paper therefore looks at the possibility of calculating the real impact of Brexit on Banco Santander via two variables: the share price and the sentiment of Twitter.

Several studies have proved that the stock price includes both the influence of any historical process such as Brexit, and the opinions shared in social networks. For this purpose, data analytics techniques have been used through programming languages such as R and Python for data extraction and sentiment measurement. After an empirical, exploratory and sentimental analysis, it was concluded that there is no correlation between the market price of Banco Santander and Twitter's sentiment on Brexit. During the study, several limitations related to the ethics of data use and the accuracy of sentiment analysis have been identified in order to facilitate the work of future research.

Key words

Banco Santander, Brexit, regulatory changes, quote, news, Twitter, sentiment analysis, correlation

ÍNDICE DE ILUSTRACIONES

<i>Ilustración 1: Balanza Comercial del Reino Unido en 2018</i>	15
<i>Ilustración 2: Precio de cierre ajustado de la cotización del Banco Santander entre 2012 y 2015</i>	20
<i>Ilustración 3: Extracto del código usado en Python para la descarga de tweets</i>	36
<i>Ilustración 4: Extracto del código de Rstudio de limpieza de texto</i>	40
<i>Ilustración 5: Características de la matriz de términos</i>	46
<i>Ilustración 6: Nube de palabras del sentimiento en Twitter sobre el Banco Santander y el Brexit</i>	47
<i>Ilustración 7: Precio de cierre ajustado de las acciones del Banco Santander entre 2016 y 2020</i>	48
<i>Ilustración 8: Volumen de Tweets por semana</i>	50
<i>Ilustración 9: Diagrama de Barras sobre los usuarios de Twitter más mencionados en el tema sobre el Banco Santander y el Brexit</i>	51
<i>Ilustración 10: Diagrama de barras de las palabras más usadas en el dataset de Twitter durante el periodo 2016-2020</i>	53
<i>Ilustración 11: Contribución de los tweets al sentimiento</i>	54
<i>Ilustración 12: Frecuencia por tipo de sentimiento respecto al Banco Santander y al Brexit</i>	55
<i>Ilustración 13: Frecuencia de Tweets por sentimiento: Positivo vs Negativo</i>	57
<i>Ilustración 14: Correlación entre el sentimiento promedio de Twitter vs Precio de cierre del Banco Santander</i>	59
<i>Ilustración 15: Sentimiento promedio de Twitter vs Precio de cierre Banco Santander (año compra del Banco Popular por el Banco Santander)</i>	60
<i>Ilustración 16: Diagrama de dispersión del sentimiento respecto al precio</i>	61
<i>Ilustración 17: Volumen de Tweets vs Precio de cotización (de cierre ajustado) del Banco Santander</i>	62

ÍNDICE DE TABLAS

<i>Tabla 1: Aportación de los ingresos por servicios financieros del Reino Unido a la UE, por sector (2015, miles de millones de £)</i>	<i>14</i>
<i>Tabla 2: Indicadores de Desarrollo del Reino Unido entre 2013 y 2016</i>	<i>17</i>
<i>Tabla 3: Medición del posicionamiento ideológico de los lectores de prensa</i>	<i>29</i>
<i>Tabla 4: Patrones de comportamiento identificados como Bots</i>	<i>41</i>

LISTA DE ACRÓNIMOS

API	<i>Application programming interface</i>
BCE	Banco Central Europeo
CEO	<i>Chief Executive Officer</i>
CIS	Centro de investigaciones sociológicas
CRM	<i>Customer relationship management</i>
ESMA	Autoridad Europea de valores y mercados
HTML	<i>HyperText Markup Language</i>
MiFID	Directiva sobre Mercados de Instrumentos Financieros
OCDE	Organización de Cooperación y Desarrollo Económico
OMC	Organización Mundial del Comercio
PYME	Pequeña y mediana empresa
RU	Reino Unido
UE	Unión Europea
URL	Localizador de recursos uniforme

ÍNDICE DE CONTENIDOS

1. Introducción	9
1.1. Objetivo	9
1.2. Justificación del tema objeto de estudio	9
1.3. Metodología y Estructura	12
2. Análisis económico-financiero anterior y posterior al referéndum del <i>Brexit</i>	13
2.1. Importancia económica de la UE en el Reino Unido	13
2.2. Análisis económico pre-referéndum de 2016	15
2.2.1. Análisis de la economía británica en 2015	16
2.2.2. Banco Santander en el Reino Unido: Características	18
2.3. Análisis económico en el contexto actual	20
2.3.1. Breve cronología del <i>Brexit</i>	20
2.3.2. Panorama económico del Reino Unido	21
2.3.3. Implicaciones del <i>Brexit</i> en el sector bancario del RU	22
2.3.3.1. Sector bancario: Banco Santander	23
3. Marco teórico y revisión de la literatura	26
3.1. Importancia de Twitter como medio de información	26
3.2. Estudios diversos	28
3.3. Análisis de sentimientos	31
3.3.1. Limitaciones en Twitter	32
4. Metodología	34
4.1. Descarga de datos	34
4.1.1. Datos financieros del Banco Santander	34
4.1.2. Datos de Twitter	35
4.1.2.1. Curiosidades en la descarga	37
4.2. Tratamiento de los datos	39
4.2.1. Suciedad de los datos	39
4.2.2. Identificación de <i>bots</i>	40
4.3. Análisis exploratorio	43
4.4. Análisis de sentimientos	43
4.4.1. Análisis léxico	43
4.4.2. Análisis semántico	44

4.4.3.	Método alternativo al enfoque semántico	45
5.	Análisis de Resultados	48
5.1.	Análisis del precio de cotización del Banco Santander	48
5.2.	Análisis exploratorio	49
5.3.	Sentimientos	54
5.4.	Correlación entre dos variables: sentimiento vs precio	58
5.5.	Correlación entre dos variables: volumen vs precio	62
6.	Conclusiones	64
6.1.	Limitaciones del estudio	65
6.1.1.	Problemas éticos en la gestión de los datos	65
6.1.2.	Falta de datos	65
6.1.3.	Falta de exactitud en el análisis de sentimientos	66
6.1.4.	Falta de una ponderación justa de las opiniones y las noticias	66
6.1.5.	Variación del precio de cotización en función del índice bursátil a analizar	67
6.1.6.	Otros factores involucrados	67
7.	Bibliografía	68

“Los mercados pueden mantener su irracionalidad más tiempo del que tú puedes mantener tu solvencia”

John Maynard Keynes

1. Introducción

1.1. Objetivo

El objetivo del presente trabajo de investigación presenta una triple vertiente. Por un lado, se expondrá el impacto que el *Brexit* ha tenido en la cotización del Banco Santander desde el 23 de junio de 2016¹ hasta el 31 de enero de 2020. Por otro lado, y mediante metodologías de *data analytics*, analizar la percepción o el sentimiento de Twitter que genera el *Brexit* con respecto al Banco Santander. Por último, estudiar si ambas variables (cotización y sentimiento) muestran algún tipo de correlación.

1.2. Justificación del tema objeto de estudio

En 2019 los españoles han utilizado Internet aproximadamente 161 minutos al día (Statista, 2019). Este es un tiempo suficiente para compartir todo tipo de información, leer noticias y formular opiniones acerca de ellas. La Asociación de Medios de Información ha llevado a cabo un estudio sobre una muestra de más de 2000 personas y ha desvelado que el 57,8% de lectores de noticias en Internet no retiene más allá del titular de un periódico (Asociación de Medios de Información, 2017).

Los titulares resaltan información que ha ocurrido en el mundo, pero están llenos de intenciones, valores y sentimientos que se manifiestan en el acto de ser comunicado. Un filósofo de la ciencia Norwood R. Hanson expresó con claridad que “los hechos están cargados de teoría”, es decir, que el sujeto que escribe un titular de una noticia cualquiera “está enmarcado en un marco conceptual dentro del cual realiza una determinada elección retórico-lingüística intencionada” (Veiga, 2016). Los titulares, ya sean de periódicos o de artículos de opinión, en su mayoría, están politizados, “moldean” la opinión pública y nos previene sobre qué ideología están “asimilando”.

¹ Fecha del Referéndum en Gran Bretaña acerca de la permanencia o no del Reino Unido en la Unión Europea.

Con la llegada del nuevo siglo, la globalización y la masificación de información en medios digitales, la información que resaltan los medios de comunicación puede influir de muchas maneras en cualquier lector y además de forma muy distinta. Las noticias generan impacto y crean patrones de comportamiento en los usuarios como está ocurriendo con el coronavirus. No fue una excepción la noticia del referéndum acerca de la salida de la UE que se produjo en 2016 en el Reino Unido, donde la mayoría de los bancos europeos mostraron temor ante la incertidumbre que acontecía. Uno de ellos, fue el Banco Santander, que opera desde 2004 en el Reino Unido a través de una filial.

En un banco, el primer impacto de cualquier acontecimiento mundial se refleja en su valor de cotización en el mercado bursátil. En el caso del Banco Santander, en diversos índices bursátiles². Sin embargo, en los medios de comunicación como Twitter, el efecto de un hecho como el *Brexit*, además de bombardear de noticias la red social, también incide en la formulación de distintas interpretaciones o sentimientos por personas de muy diversas ideologías de lo que podría pasar en un futuro en relación con el Banco Santander. La información que se comparte en Twitter es dimensional y gracias a técnicas de *Business Analytics*, como el procesamiento de lenguaje natural, se pueden analizar los sentimientos de sus usuarios. Este tipo de herramientas permiten el tratamiento de un gran volumen de datos y extraer unas conclusiones muy válidas para cualquier sector que de otra manera no se hubieran podido percibir.

Ante estas evidencias, el tema objeto de estudio de este trabajo se ha elegido con el propósito de poner en la práctica las distintas técnicas de *Business Intelligence y Analytics* que hoy en día se consideran herramientas tan fundamentales en el mundo corporativo.

A partir de esta primera intención, se ha buscado un tema que trate de combinar el funcionamiento del sector bancario y su relación con las políticas económicas e internacionales. Los riesgos económicos del *Brexit* en la economía británica es un debate que hoy en día es discutible y existe una gran incertidumbre sobre los efectos que generará en el sector bancario europeo. Se han materializado algunas consecuencias de éste en distintas acciones llevadas a cabo por diversos bancos; no obstante, existe un gran

² Aunque el Banco Santander cotiza en distintas bolsas mundiales como en el IBEX-35, Bloomberg World Banks o Euro Stoxx Banks, el estudio se centra exclusivamente en el NYSE (New York Stock Exchange), el mayor mercado de valores del mundo en volumen monetario.

desconocimiento acerca de la verdadera percepción existente del impacto del *Brexit* en este sector. Una de las entidades financieras más afectadas, como se ha mencionado, es el Banco Santander, también reconocido desde 2018 como el primer banco europeo por capitalización bursátil con más de 60.000 millones de euros (Statista, 2020).

La entrada del Banco Santander en el mercado británico se concretó en noviembre de 2004 y, desde ese momento, su estructura se ha ido transformando hasta componer tres segmentos de negocio de clientes, apoyados por el Centro Corporativo (Banco Santander, s.f.). Sin embargo, en 2019 el Banco Santander se ha visto afectado por las exigencias de la Autoridad de Regulación Prudencial³, que ha obligado a que, a partir del 1 de enero de 2019, las entidades con más de 25.000 millones de libras separen las operaciones minoristas de la banca de inversión, lo que sin ser su objetivo prioritario ha creado más competencia en el mercado del país. Esta normativa ha provocado, según Ana Botín⁴, la duplicación de algunas funciones y, en consecuencia, una disminución de la capacidad de generar beneficios en el banco debido al incremento de los costes (Banco Santander SA, 2019). De hecho, el banco anunció en septiembre de 2019, unos recortes en sus resultados de unos 1.500 millones de euros a causa del deterioro en mercado británico.

No obstante, el Banco Santander no sólo se ha visto afectado por la nueva regulación. El *Brexit* es una realidad y está provocando una desaceleración de la economía británica. A pesar de la incertidumbre que genera, dada la nula definición de lo que será y la falta de acuerdos en el periodo transitorio, las empresas españolas que operan en el Reino Unido se han venido preparando hace tiempo, y el Banco Santander no es una excepción. La presidenta del Banco, entre otras, ha establecido algunas hipótesis de lo que podría pasar, pero se requiere de un análisis más profundo de la información que señalen el impacto real en su valor de cotización en el mercado bursátil, a lo que ayudará este estudio con el análisis del sentimiento percibido en las redes sociales.

³ Organismo impulsado por el ministro de economía británico en 2013. Forma parte del Banco de Inglaterra y constituye una reforma del organismo de regulación del sistema financiero.

⁴ Presidenta del Banco Santander desde 2014.

1.3. Metodología y Estructura

La metodología utilizada en este estudio se ha dividido en dos partes: por un lado, una metodología general, centrada en el uso de recursos de información procedentes de organismos como la OCDE y el Banco Mundial, bases de datos como *EbscoHost* y *Google Scholar*, un buscador especializado en la búsqueda de contenido científico-académico; por otro lado, una metodología concreta enfocada en la parte analítica-empírica de esta investigación. Esta última se desarrolla posteriormente en el capítulo 4.

La estructura del presente trabajo de investigación se articula en cinco partes, además de esta introducción. En primer lugar, en el capítulo 2, se presenta una contextualización económica, financiera y regulatoria del Reino Unido, la UE, el Banco Santander y el *Brexit*, en la que se recoge un amplio recorrido de la relación entre la UE y el Reino Unido y su implicación con el sector bancario en los últimos cinco años, dando especial importancia al *Brexit* y a la posición que ocupa el Banco Santander. Seguidamente, en el capítulo 3, se establece el marco teórico centrado en la importancia de Twitter como medio de información, realizando a su vez la revisión de literatura relevante para nuestro análisis y en una breve introducción al análisis de sentimientos. El capítulo 4, de metodología, está enfocado en la descarga de datos y su posterior tratamiento, así como en las aplicaciones y recursos del análisis de sentimientos. En la sección 5 se presentan los resultados obtenidos usando las técnicas explicadas en el capítulo anterior, dando lugar al sentimiento promedio en Twitter respecto al tema que nos concierne y comparando este sentimiento con el precio de cierre ajustado del Banco Santander. Finalmente, el último capítulo, el seis, expone las conclusiones y cierra el trabajo con una serie de limitaciones a tener en cuenta para posteriores investigaciones.

2. Análisis económico-financiero anterior y posterior al referéndum del *Brexit*

En este capítulo, se realiza un recorrido de la relación existente entre el Reino Unido y la Unión Europea en términos financieros para un mayor entendimiento del tema a tratar definido en los objetivos. A continuación, se analiza la situación financiera del Reino Unido antes del referéndum de 2016 y después de éste, para deducir qué cambios se han producido y qué medidas se han implementado. También conocer el funcionamiento del Banco Santander y su presencia en el Reino Unido resulta clave ante la presente investigación. En este sentido, se analiza el Banco Santander antes del *Brexit*, es decir, en 2015, y en el periodo pos-referéndum hasta la actualidad.

Para ello, este apartado se ha basado en informes y artículos académicos de gran prestigio de diversas instituciones mundiales y organizaciones internacionales con el propósito de aportar un mayor enfoque económico al presente trabajo de investigación. La OCDE, el Banco Mundial, el Grupo Banco Santander SA son ejemplos de algunas de estas instituciones de las que se han obtenido una gran variedad de datos e información.

2.1. Importancia económica de la UE en el Reino Unido

Según la firma internacional *Ashurst*, el Reino Unido es el principal centro internacional de Europa y lidera la mayoría de las áreas de servicios financieros (Ashurst, 2016). También considera que tiene una gran presencia en el mercado mundial de divisas, con acceso a otros mercados de todo el mundo gracias a los acuerdos de libre comercio concertados por la UE con otros países (Ashurst, 2016).

En 2015, el Banco de Inglaterra señaló la influencia de la Unión Europea en la economía británica a través de tres cuestiones.

En primer lugar, argumentó que la adhesión a la UE ha reforzado el dinamismo de la economía del Reino Unido, en la medida en que aumenta su apertura económica y financiera (Bank of England, 2015), la cual promueve la innovación y la adopción de nuevas tecnologías mediante la libre circulación del capital y la mano de obra, es decir,

permite a las empresas explotar economías de escala y especializarse en una gama más reducida de productos. Esto refleja una ventaja competitiva para el Reino Unido, dado que la economía más dinámica está preparada para distintos imprevistos y puede crecer más rápidamente sin una posible desestabilización financiera.

En segundo lugar, como resultado de una integración más estrecha con la Unión Europea y, más recientemente, con la zona del euro, el Banco de Inglaterra expuso también que han aumentado los desafíos, lo que ha supuesto una mayor preparación de Gran Bretaña ante conmociones económicas y financieras procedentes del extranjero (Bank of England, 2015). En tercer y último lugar, los reglamentos y directivas de la Unión Europea han definido el sector bancario de la región, lo que hace que el Reino Unido esté sujeto a cualquier cambio de regulación bancaria europea.

Sin embargo, el Reino Unido no es miembro de la eurozona y, por lo tanto, no tiene ningún papel dentro del Banco Central Europeo (BCE). Por el contrario, el Banco de Inglaterra sí ejerce una influencia importante tanto a nivel europeo como mundial, al ser considerado un gran aportador de la reglamentación europea en el ámbito bancario y empresarial de servicios financieros. (Bank of England, 2015).

El sector bancario, sin duda, ha tenido un peso considerable dentro de la Unión Europea en los últimos años. Según una de las consultoras más influyentes, Oliver Wyman, es el sector más grande, representando un 54% de los ingresos por servicios financieros de la organización supranacional, como se puede observar en la Tabla 1 (Armour, 2017).

Tabla 1: Aportación de los ingresos por servicios financieros del Reino Unido a la UE, por sector (2015, miles de millones de £)

Sector	Banca	Gestión de activos	Cobertura	Infraestructura comercial	Total
Ingresos intracomunitarios	27	6	5	12	50
Ingresos totales	117	23	42	26	208
Fracción de la UE de los ingresos totales del sector	0,23	0,26	0,12	0,46	0,24
Fracción sectorial de los ingresos totales de la UE	0,54	0,12	0,1	0,24	1

Fuente: Los datos de ingresos son de un estudio de Oliver Wyman (2016, p.6). Los ingresos intracomunitarios comprenden los servicios financieros del Reino Unido procedentes de servicios de negocios internacionales y mayoristas relacionados con la UE. (Armour, 2017)

En cuanto a los flujos comerciales con la Unión Europea, éstos constituyen un factor esencial para el Reino Unido. De hecho, la balanza comercial del Reino Unido presenta un déficit excepcional que ha venido empeorando desde 2016, especialmente por la incertidumbre generada por el *Brexit*. El déficit es de 196.903,5 millones de euros en 2019, lo que supone un 7,8% de su PIB (Macro, 2019). Su principal motivo es que la Unión Europea se define como su gran socio comercial. Más de la mitad de sus importaciones de bienes y donde se dirigen casi la mitad de sus exportaciones proceden de la Unión Europea. En la Ilustración 1 se puede observar esta importancia a través del porcentaje de exportaciones e importaciones en 2018 extraído de la OCDE. No obstante, el Reino Unido también supone una gran aportación a la UE a nivel comercial, dado que le permite mantener un vínculo con los países de la *CommonWealth* y mantener una visión más global.

Ilustración 1: Balanza Comercial del Reino Unido en 2018



Fuente: *Elaboración propia. Datos extraídos de la OCDE (2018).*

2.2. Análisis económico pre-referéndum de 2016

A partir de este primer entendimiento sobre la relación económica existente entre la UE y el RU se va a analizar la economía británica antes del referéndum de 2016. Este análisis se ha extraído especialmente del informe de la OCDE sobre las perspectivas económicas mundiales de 2015 y el estudio se ha centrado en la sección referente al Reino Unido.

2.2.1. Análisis de la economía británica en 2015

Según el informe de la OCDE, en 2015 se prevé que la economía británica destaque por su gran crecimiento, especialmente impulsado por la demanda interna. Sin embargo, se estima que el crecimiento del PIB se desacelere en 2016, siendo el consumo y la inversión privados los principales motores de éste. En cuanto a la balanza por cuenta corriente, existen cálculos de que alcance en el futuro una mejora considerable debido a la mayor rentabilidad de las inversiones en el extranjero, gracias especialmente a la recuperación de la zona euro. En estos momentos, su déficit por cuenta corriente se ha incrementado al importar en una mayor cantidad de la Unión Europea tras la crisis. La primera fila de la Tabla 2 refleja este incremento del déficit entre 2013 y 2016.

Durante esta época, la productividad laboral, los salarios y la competitividad siguen estancados tras la crisis y hay probabilidades de que este bajo rendimiento se deba a factores estructurales más que cíclicos. La propuesta de la OCDE en este aspecto pasa por demandar una inversión más fuerte a través de un mayor acceso de las empresas a la financiación, así como una mejora de las infraestructuras, puesto que la calidad percibida de éstas es inferior a la de otros países del G7. Otra solución que propone la OCDE es el fortalecimiento de la estrategia y la planificación gubernamentales a largo plazo con el propósito de atraer a inversores privados y aprovechar su inversión junto con los recursos públicos.

De otro modo, la industria del Reino Unido está muy desarrollada y contribuye con un 18,14% al PIB. El sector servicios, enfocado en el sector bancario, por otro lado, representa un 70,4% de la economía del país, a pesar de tener una moneda distinta al euro: la libra esterlina. En las últimas filas de la Tabla 2 se puede apreciar el porcentaje en que contribuye cada sector a la economía británica en 2015.

Tabla 2: Indicadores de Desarrollo del Reino Unido entre 2013 y 2016

Country Name	Series Name	2013	2014	2015	2016
United Kingdom	Current account balance (% of GDP)	-4,76891	-4,72067	-4,90301	-5,26389
United Kingdom	Industry (including construction), value	18,81703	18,35948	18,14162	17,58114
United Kingdom	Agriculture, forestry, and fishing, value :	0,672181	0,666984	0,621473	0,604567
United Kingdom	Services, value added (% of GDP)	69,76285	70,13686	70,40841	70,92242

Data from database: World Development Indicators

Last Updated: 02/27/2020

Fuente: World Bank, 2020

No obstante, el sector bancario en el Reino Unido está muy afectado por la crisis. En el modelo precrisis los bancos destacaban por ofrecer una financiación a corto plazo y una regulación y supervisión relativamente indulgente (Armour, 2017). En consecuencia, el sector bancario presentaba problemas de liquidez por la gran dimensionalidad del apalancamiento. Este fue el motivo por el que el Tesoro Público del Reino Unido tuvo que intervenir a través de distintas medidas de recapitalización. Las medidas que adoptó se recogen en el estudio económico del Reino Unido de la OCDE. Entre ellas destacan:

- Recapitalización del *Lloyds Banking Group* y del *Royal Bank* de Escocia a través de una serie de transacciones donde el Tesoro Público adquirió el 41% del primero y el 83% del segundo.
- Prestación de dinero a través de un Plan de Compensación de Servicios Financieros para poder garantizar los depósitos de los clientes de hasta 50000 libras esterlinas.
- Prestación de dinero directamente a bancos insolventes para que puedan devolver los depósitos de clientes de más de 50000 libras esterlinas, incluidos el Banco Escocés de Londres, la Sociedad de Construcción Dunfermline y varios bancos islandeses.
- Dirigir el Plan de garantía de crédito, entre octubre de 2008 y enero de 2012, para ayudar a restablecer la confianza de los inversores en la financiación mayorista de los bancos, garantizando ciertas deudas no garantizadas a cambio de una comisión.

En el modelo poscrisis, los créditos se han reducido, especialmente, aquellos que se conceden al sector corporativo, lo que ha provocado que las PYMEs del Reino Unido hayan presentado dificultades en la financiación, al no tener alternativas al sector bancario. El gobierno británico ha tenido que impulsar una serie de programas para estimular los préstamos a estas pequeñas y medianas empresas. Del mismo modo, el sector bancario del Reino Unido debe continuar su transformación para centrarse más en las necesidades genuinas de los hogares y de los clientes de las pequeñas y medianas empresas y para ganarse su confianza y la de los reguladores, los responsables de la formulación de políticas, los inversores y el público en general.

En cuanto a la regulación del sector financiero del Reino Unido, el marco institucional para la estabilidad financiera se ha reformado a causa de la crisis, con la creación del Comité de Política Financiera y la Autoridad de Regulación Prudencial del Banco de Inglaterra lo que ha permitido que el Reino Unido desarrolle su rol como centro financiero internacional. Esta cuestión es fundamental dado que en el Reino Unido operan más bancos extranjeros que en otro país de la Unión Europea y alrededor de la mitad de las empresas financieras más grandes del mundo tiene su sede allí.

Ante estas evidencias, pasaremos a analizar el banco que nos concierne en este estudio en el Reino Unido antes de la proclamación del *Brexit*.

2.2.2. Banco Santander en el Reino Unido: Características

El grupo Santander UK Holdings PLC es la filial de la empresa española “Banco Santander SA” que opera en el Reino Unido. Esta división es la tercera unidad más grande del grupo, al ser una de las principales compañías de servicios financieros personales del Reino Unido, así como uno de los mayores proveedores de hipotecas y ahorros de éste (Banco Santander, s.f.). Su entrada en el mercado británico se concretó en noviembre de 2004 y, desde ese momento, se ha transformado, pasando de tres antiguas sociedades en construcción a un banco comercial y minorista de servicio completo. En este sentido, su negocio se gestiona a través de tres segmentos de negocio de clientes, apoyados por el Centro Corporativo: banca minorista, banca comercial y banca corporativa global (Banco Santander, s.f.).

Según un informe publicado por la filial británica del Banco Santander en 2015, la entidad sigue su propio camino, iniciado por Ana Botín durante su época como CEO, para convertirse en un banco que es simple, personal y justo y que se gana la lealtad duradera de sus clientes, personas, accionistas y comunidades. De hecho, fue votado por los consumidores del Reino Unido como el “Banco de Mayor Confianza” (Santander UK plc, 2015).

En el mercado de las PYMES, ha crecido de manera constante, pero aun así sólo representa una pequeña proporción del mercado de préstamos a PYMES, mientras que, los cuatro grandes bancos del Reino Unido de forma conjunta, se hacen cargo del 85% del volumen de préstamos pendientes (Santander UK plc, 2015).

Por otro lado, el grupo Santander en el Reino Unido ha utilizado su posición como "*scale challenger*" para lanzar productos innovadores que requieren una inversión sustancial y la capacidad de llegar a un mercado nacional. Entre ellos se incluye la Cuenta Corriente 123, con más de cuatro millones de clientes, que ha provocado cambios en otras partes del sector a medida que aumentan sus usuarios. También ha desarrollado una serie de avances digitales para aumentar la fidelidad de sus consumidores en el segmento de banca minorista. Entre ellos destacan una nueva plataforma de gestión de patrimonio y el establecimiento de una red wifi y tabletas inteligentes desplegadas en las sucursales para mejorar el servicio al cliente. Por último, se ha desarrollado un paquete de análisis con capacidad para analizar y controlar el rendimiento en todas las plataformas de la banca online, así como una nueva funcionalidad de CRM para mejorar la experiencia del cliente existente (Santander UK Group Holdings plc Q2, 2015).

Gracias a ello el Banco Santander en el Reino Unido se ha ido recuperando en el mercado desde el impacto sufrido por la crisis de 2007-2008 y ha ido adquiriendo valor. Se puede apreciar en el valor de la cotización que, desde mediados de 2012 hasta 2015, se ha incrementado considerablemente (véase Ilustración 2).

Ilustración 2: Precio de cierre ajustado de la cotización del Banco Santander entre 2012 y 2015



Fuente: Elaboración propia con la Matplotlib Python. Datos extraídos de Yahoo finance.

2.3. Análisis económico en el contexto actual

2.3.1. Breve cronología del *Brexit*

Hace tres años y medio, concretamente el 23 de junio de 2016, tuvo lugar en el Reino Unido un referéndum acerca de su permanencia o no en la UE. El 52% de los británicos votaron a favor de que abandonase la institución. El referéndum fue convocado por David Cameron ante la amenaza constante del populismo nacionalista liderado por Nigel Farage. David Cameron aseguraba que podía vencer en la consulta; sin embargo, tuvo que dimitir al día siguiente ante los impactantes resultados de dicho proceso político.

Bajo el mandato de Theresa May (2016-2019), se activó el artículo 50 del Tratado de Lisboa de la UE y se estableció un plazo de dos años para negociar la salida, siendo el 29 de marzo de 2019 la fecha de culminación del *Brexit*. Sin embargo, la salida tuvo que prolongarse varias veces tras la falta de un acuerdo con Bruselas y la presión de la oposición. No obstante, con la llegada de Boris Johnson al poder, el *Brexit* se ha convertido en una realidad. El Reino Unido ya no es miembro de la Unión Europea desde el 31 de enero de 2020, sin embargo, se ha establecido un periodo transitorio de un año donde el país seguirá actuando bajo la jurisdicción de la UE, con el propósito de encauzar las negociaciones comerciales. Parece que la pandemia mundial del coronavirus está

afectando en la evolución de estas negociaciones y se han vuelto a pedir nuevas prórrogas. Sin embargo, Boris Johnson no las acepta lo que acrecienta la probabilidad de que el *Brexit* se caracterice por ser caótico y sin acuerdo.

2.3.2. Panorama económico del Reino Unido

En estos momentos, las perspectivas económicas presentan una gran incertidumbre en comparación a 2015, incertidumbre manifestada por el *Brexit* y por el desconcierto que ha provocado la epidemia del Covid-19 en el mundo. En cuanto al *Brexit*, tema que aquí se nos plantea, existen ciertas hipótesis⁵. La primera se relaciona con una transición suave que termine después de 2021, donde se espera que la actividad crezca alrededor del 1%. (Hawksworth & Teow, 2019). Sin embargo, su complejidad seguiría frenando la inversión hasta que hubiese claridad sobre los futuros acuerdos comerciales. Por otro lado, si se produjera una salida perjudicial, el Banco de Inglaterra debería estimular la economía mediante la reducción de los tipos de interés y la compra de bonos, aunque esto sería insuficiente para compensar plenamente la considerable pérdida de producción en tal caso (OCDE, 2019).

No obstante, en general se prevé que la inflación se reduzca por debajo del 2% (OCDE, 2019). El Reino Unido debe impulsar acuerdos que aseguren sus ingresos, ya sea con la Unión Europea o en el extranjero, para no caer en una desaceleración económica continuada y sufrir un deterioro en el comercio e inversión empresarial. Hay signos negativos de que, debido al *Brexit*, la tasa de desempleo aumente también (De, 2020) y que el flujo de capital humano exterior disminuya por la incertidumbre de las condiciones laborales, especialmente el número de migrantes procedentes de la Unión Europea. Además, los riesgos siguen aumentando, acentuados por el brote de coronavirus, el cual ha afectado a la actividad mundial y ha interrumpido todas las cadenas de suministro en la economía.

⁵ Todas ellas formuladas antes del nuevo escenario post Covid-19, que está obligando a reformularlas justo en el momento de entregar este trabajo.

2.3.3. Implicaciones del *Brexit* en el sector bancario del RU

La banca y la industria de servicios financieros son un componente vital de la economía del Reino Unido y como se ha reflejado mantienen una relación muy estrecha con la Unión Europea. La falta de acceso del sistema de servicios financieros del Reino Unido al mercado del Espacio Económico Europeo puede afectar negativamente el desempeño de su economía.

Aunque el Reino Unido posea sus propios organismos reguladores, como la Autoridad de Conducta Financiera, gran parte de los servicios financieros del país están regulados a nivel europeo, a través de la Autoridad Europea de Valores y Mercados (ESMA), la cual supervisa la calificación crediticia de las agencias. (Ashurst, 2016)

Una de las legislaciones más importantes de la UE y del Espacio Económico Europeo desde el punto de vista bancario (además de Basilea III), es la Directiva de Mercados de Instrumentos Financieros (MiFiD) que entró en vigor el 1 de noviembre de 2007. Ésta regula la realización de servicios de inversión en los Estados miembros y permite a los bancos realizar operaciones en toda Europa sin obtener una licencia en cada país, para crear un único mercado de servicios financieros y fomentar la transparencia y protección del consumidor (Bankinter, 2016). Esto es lo que se ha conocido como el derecho de pasaporte. Por su parte, prohíbe que los estados que tienen su sede en otro país puedan hacer uso del sistema de pasaportes a no ser que establezcan una presencia autorizada en la jurisdicción en cuestión (Ashurst, 2016). De hecho, muchos bancos optaron por establecerse en Londres a través de una filial para acceder a los mercados de la UE a través de este sistema.

La MiFiD II, aprobada el 16 de abril de 2014, ha sido más flexible en cuanto a la prestación de servicios, estableciendo dos regímenes separados para terceros países y distinguiendo entre clientes minoristas y clientes profesionales (Bankinter, 2016). En referencia a estos últimos, ha permitido que una entidad de terceros países realice negocios sin la obligación de establecer una sucursal en un estado miembro, pero con la condición de que esté registrado en la ESMA (Condado, 2017). Sin embargo, en relación con los clientes minoristas, se mantiene la obligación de poseer una sucursal en un Estado miembro.

El *Brexit* ha creado una incertidumbre y una gran volatilidad en el mercado que está provocando un menor crecimiento y la posibilidad de una desaceleración significativa en el sector bancario europeo. A pesar de que la banca europea se nutrió de activos líquidos de alta calidad y endureció la regulación después de la crisis, el sector bancario europeo está sufriendo con los cambios regulatorios a causa del *Brexit*. Las implicaciones que se perciben de primera mano si no se alcanza una negociación en el sector bancario serían muy diversas. En primer lugar, el derecho de pasaporte que permite realizar operaciones transfronterizas entre Estados de la Unión Europea podría ser anulado. De hecho, si no se mantiene, muchos bancos ingleses tendrán que llevar a cabo una reestructuración si desean continuar con sus actividades minoristas en el Espacio Económico Europeo. La deslocalización de muchas entidades financieras del Reino Unido podría ser una opción. Además, estos bancos actuarían bajo los requisitos de la OMC que conllevan limitaciones significativas en el comercio transfronterizo en comparación con el sistema de pasaportes (Sowels, 2018). Supondría que todas las exportaciones del Reino Unido a la UE tendrían que cumplir las regulaciones comerciales de la institución, mientras que el Reino Unido, no tendría influencia sobre el desarrollo de éstas (Ashurst, 2016).

Por último, la regulación del Reino Unido al diferir de la UE podría acarrear costes y complejidad. No obstante, como hemos expuesto con anterioridad, algunas reglamentaciones de la UE se originan en los compromisos internacionales del Reino Unido, como es el caso de los requisitos de capital y liquidez de Basilea III y IV, por lo que hace poco probable que el Reino Unido se aleje de la UE en esta área (PwC, 2017). En consecuencia, es poco probable que en un principio el panorama reglamentario cambie de manera significativa.

2.3.3.1. Sector bancario: Banco Santander

En 2017, el Banco Santander compró el Banco Popular al Mecanismo Único de Resolución europeo por un precio mínimo, un euro. Esta medida se tomó debido a los problemas de liquidez del Banco Popular, sin embargo, ha ayudado al Banco Santander para empujar su actividad en España, la cual se había estancado ante el impacto de la filial del Reino Unido a causa del *Brexit*.

A pesar de esta adquisición, según Nathan Bostock⁶, los resultados de 2019 se han visto afectados por la presión competitiva continua, una agenda regulatoria exigente y la inversión que se está realizando para transformar el banco para el futuro (Banco Santander SA, 2019). De hecho, las encuestas de consumidores y empresas siguen siendo débiles a causa de las preocupaciones sobre el *Brexit* y la economía en general.

En la filial del Banco Santander en el Reino Unido, el *Brexit* se ha reflejado en varias acciones llevadas a cabo por ésta. En primer lugar, ha desarrollado un foro para coordinar y aumentar los impactos operacionales y que está diseñado para que las funciones empresariales y de apoyo tomen medidas proactivas cuando sea necesario (Banco Santander SA, 2019). Este foro se ha denominado *Brexit Response Group*.

Por otro lado, la filial presenta altos niveles de liquidez para asegurar una gran preparación ante la incertidumbre del *Brexit*. Aunque haya disminuido en relación con el año anterior, su coeficiente de cobertura de liquidez se encuentra en el 155%⁷, lo que supone que el banco tiene un fondo adecuado de activos líquidos de alta calidad, coherente con las medidas establecidas después de la crisis.

En cuanto a sus fuentes de ingresos, en comparación con el 2018, los ingresos no financieros han disminuido un 11% en gran parte debido al cierre de negocios comerciales después de la implementación de los cambios de *ring-fencing*⁸ (Santander Uk Group Holdings plc, 2019). Como banco del Reino Unido, la filial está sujeta a los requisitos de la Autoridad de Conducta Financiera y de la Autoridad Reguladora Prudencial. Esto ha obligado al grupo a separar su negocio estratégico del resto de actividades. Es por ello por lo que en 2018 el Banco Santander transfirió alrededor de 22.900 millones de activos y 20.700 millones de libras de pasivos de la filial británica a la sucursal de Banco Santander en Londres, para cumplir con esa exigencia regulatoria (Aguado, 2018). No obstante, la caída de ingresos ha sido compensada por una contraprestación adicional de 15 millones de libras recibida en relación con la venta de acciones de *Vocalink*, empresa de sistemas de pago con sede en el Reino Unido.

⁶ CEO del grupo Santander UK Plc.

⁷ Obtenido de Informe Anual del Banco Santander de 2019.

⁸ Barrera establecida por la Ley de Reforma bancaria de 2013 que separa una parte de los activos financieros de un banco del resto.

El cambio regulatorio del Reino Unido después del *Brexit* puede añadir más complejidad. Aunque en un principio se prevé que los cambios no sean significativos, el gobierno del Reino Unido ha publicado más de diez instrumentos reglamentarios en virtud de la Ley de Sanciones y lucha contra el blanqueo de dinero en 2018, que entraría en vigor si el Reino Unido saliera de la UE sin un acuerdo (Santander Uk Group Holdings plc, 2019).

El difícil entorno regulatorio, incluyendo los impactos de la Ley de Reforma Bancaria, y la incertidumbre generada por el *Brexit* ha generado un deterioro contable por importe de aproximadamente 1.500 millones de euros en el fondo de comercio del Banco Santander, así como una menor capacidad para generar beneficio (Santander Board, 2019). La Ley de Reforma Bancaria también ha dado lugar a un aumento de 40.000 millones de euros en los activos de la sucursal española de Londres (Santander Board, 2019).

Por último, la tasa de desempleo del Reino Unido se ha visto perjudicada por el cierre de 140 oficinas en el Reino Unido para adaptar su estructura a la demanda de banca digital (Inversor Ediciones, SL, 2020).

3. Marco teórico y revisión de la literatura

Después de haber revisado en profundidad la situación económica que concierne al Banco Santander y al Reino Unido desde 2015 hasta la actualidad, el capítulo 3 se adentra en el campo de los fundamentos teóricos, los cuales se constituyen una pieza fundamental para comprender el desarrollo del trabajo. Y en él se realiza además la revisión de la literatura relativa a éstos.

3.1. Importancia de Twitter como medio de información

Los periódicos tradicionales están dejando de ser el medio más visitado para obtener información. Hoy nueve de cada diez personas tienen abierto al menos una red social en su ordenador (Instituto Nacional de Estadística, 2019), lo que significa que las redes sociales están ganando terreno en este ámbito y se están convirtiendo en el principal sustituto de los medios de comunicación tradicionales. Según James Hohmann, corresponsal nacional en el periódico *Washington Post*, “*las plataformas de medios sociales continúan emergiendo como herramientas esenciales de periodismo. Estos medios ofrecen excitantes oportunidades para que los reporteros colecten información y para que las organizaciones periodísticas expandan el alcance de su contenido*” (Blanco, 2013). Se podría decir que en la actualidad las redes sociales han cambiado el paradigma del acceso a la información.

Twitter es la última de las redes sociales de uso generalizado y se ha convertido con el paso del tiempo en un medio de comunicación que está dejando a los medios tradicionales en un segundo plano.

Fue creada en 2006 por Jack Dorsey, Evan Williams y Biz Stone como una plataforma social de *microblogging*⁹ (Moreno, 2015) con una seña de identidad clara: el envío de mensajes cortos de menos de 140 caracteres. Hoy Twitter se define a si misma en su pagina web: “Es como recibir un periódico cuyos títulos siempre te resultan interesantes, cambiando títulos por titulares” (Twitter, 2019). Los propios medios de comunicación

⁹ Herramienta de comunicación que permite a sus usuarios enviar y publicar mensajes breves, generalmente solo de texto.

tradicional también poseen sus cuentas y comparten sus titulares en esta red social a través de hipervínculos, creando diversas opiniones al respecto.

Además, Twitter posee una multitud de funciones que permiten desde interactuar entre usuarios con el correspondiente intercambio de información, introducir una noticia completa de un sector a través de un hipervínculo hasta proporcionar los temas más populares del momento. Entre ellas destacan:

- Mencionar a un usuario, retwittear o marcar favorito. Todas ellas se caracterizan por ser responsables de la interacción pública entre usuarios en Twitter y promover hilos de conversación. Si mencionas a un usuario en tu publicación se debe realizar poniendo el signo @ con anterioridad, para que la persona reciba un aviso y ver los tweets en los que le han mencionado.
- Crear listas temáticas a través de los *hashtags*. Éstos son etiquetas muy útiles que permiten organizar el contenido de los tweets respecto a un determinado tema.
- Establecer *Trending topics*. Se corresponde con los temas más populares en Twitter en ese momento, es decir, aquellos más tratados en toda la red.
- Crear hipervínculos a una mención o a un enlace a una web. Esta función permite a los usuarios ahorrar el tiempo a la hora de escribir los tuits. Es muy útil para proporcionar noticias de los periódicos.

La utilidad masiva de Twitter radica en el hecho de que es una red eminentemente informativa. Puede utilizarse bien para estar al día de la actualidad general o bien para formular mensajes y expresar opiniones acerca de cualquier tema. En total en Twitter, existen 330 millones de usuarios activos mensuales y el 71% de ellos ha confirmado que usa la red social para acceder a las noticias (Omnicores, 2020). A pesar de que existe una intencionalidad diferente en función de quien comparta una noticia a través de un titular, la cual está influenciada por sentimientos, las noticias son hechos objetivos y no representan opiniones.

El problema surge cuando “el usuario de redes sociales tiende a calificar como noticia casi cualquier cosa que lee”, así lo comenta David Alejandro Rodríguez, banquero de BBVA Bancomer (Sánchez, 2017). En Twitter, no solo se divulgan noticias, sino que la mayoría de los datos acumulados representan opiniones. Opiniones que, con tan solo 140

caracteres, pueden afectar directamente a los mercados y que vienen influenciadas por sentimientos.

3.2. Estudios diversos

Cuando se anuncian noticias tan impactantes como la salida del Reino Unido de la Unión Europea, los mercados financieros europeos son los primeros afectados. Esto es debido a que el mercado se alimenta de noticias tales como el *Brexit*, que implica a numerosos sectores europeos. Del mismo modo, afecta a las empresas o a los bancos que toman en consideración ciertas medidas ante las posibles implicaciones futuras que se reflejan en la prensa como hemos podido comprobar en el capítulo 2 con el Banco Santander. En otras palabras, las noticias encierran en sí mismas valores y sentimientos.

Por otro lado, los lectores de esas noticias, en función del periódico que elijan, también muestran un perfil ideológico. Un estudio publicado en 2015 por el Centro de Medios, Reputación e Intangibles de la Universidad de Navarra mostró el perfil ideológico de los lectores de prensa. Una parte de la metodología utilizada fueron dos encuestas realizadas en Twitter, en las cuales se debía indicar la valoración ideológica que personalmente se indicaba a cada periódico. El cálculo del posicionamiento ideológico se basó en los datos del CIS de noviembre de 2015, siendo 1 izquierda extrema y 10 derecha extrema (Gonzalo, 2016). Se alcanzó una valoración media de casi 850.000 votos (Reputation, 2015). Los resultados fueron los siguientes:

Tabla 3: Medición del posicionamiento ideológico de los lectores de prensa

Periódico	Preferencias
Expansión	6,0
ABC	5,6
El Confidencial	5,2
El Mundo	5,1
El Economista	4,7
El español	4,5
Voz pópuli	4,3

Fuente: Elaboración propia. Los datos han sido extraídos de manera aleatoria del análisis de encuestas del estudio sobre el perfil ideológico de los lectores de prensa publicado por el Centro de Medios, Reputación e Intangibles (Reputation, 2015).

La Tabla 3 muestra parte de la información obtenida en el estudio. Se aprecia que la mayoría de los periódicos están categorizados en torno a una ideología de centro, pero con inclinaciones a derecha o izquierda.

Sin embargo, varios estudios han reflejado que las noticias no son la única variable que promueven los movimientos en el mercado. Lawrence Summers, secretario del Tesoro de los Estados Unidos durante el mandato de Bill Clinton, publicó un estudio denominado “*What Moves Stock Prices*” que intentaba demostrar hasta qué punto las noticias podían explicar las fluctuaciones bursátiles. En ese estudio se extrajeron cincuenta movimientos de bolsa entre 1947 y 1987 y se buscó en la prensa titulares de noticias que explicaran alguna de esas alteraciones en el mercado (Luis, 2019). Los resultados reflejaron que más de la mitad de los movimientos no presentaban ninguna relación con las noticias. Es decir, que existían otras variables que influían en el mercado y que no atienden a criterios objetivos.

Por su parte, Citigroup ha intentado medir las sorpresas económicas a través de la creación de un índice, el Economic Surprise Index, que compara las noticias económicas con las expectativas de los analistas (Bespoke Investment Group, 2019). En 2015 y 2016

se demostró que este índice y el S&P500 presentaban una correlación positiva. Sin embargo, desde 2018 esta tendencia cambió, provocando una correlación negativa entre ambos índices.

Esto hace muy complicado medir el impacto de los acontecimientos como el *Brexit* en los mercados financieros, porque ese impacto también depende de las circunstancias y de las opiniones que se generan al respecto en las redes sociales.

Como hemos señalado con anterioridad, Twitter es una red social que combina tanto opiniones como noticias. Es por ello por lo que se ha elegido este medio para realizar el análisis del sentimiento en esta investigación.

Además, diversos estudios confirman la relación entre la bolsa y los sentimientos en las redes sociales. Por ejemplo, la Universidad de Múnich, tras un estudio de alrededor de 250.000 tweets sobre la bolsa llegó a la conclusión de que se puede predecir su valor con un 87% de fiabilidad (Sayans, 2019). Wall Street, entre otras bolsas de mercado, han expresado su deseo de utilizar la enorme cantidad de datos para hacer predicciones de inversión. De hecho, distintas empresas de análisis de opiniones han satisfecho su deseo. Es el caso de Senseitrade e Investmood, *start-ups* financieras de análisis de inversiones, las cuales han firmado un acuerdo sobre un sistema de trading algorítmico en bolsa que opera en función del sentimiento del inversor en Twitter. Todos los sentimientos que se expresan en Twitter hacen que los mercados no sean racionales.

Se han explicado diversos estudios que muestran la relación entre los movimientos bursátiles y los sentimientos en redes sociales. También se han mostrado los cambios existentes en la economía británica y en el Banco Santander tras el anuncio del *Brexit*. Sin embargo, ninguna investigación se ha enfocado en el verdadero impacto que crean los sentimientos de Twitter en el mercado bancario a partir de un único acontecimiento. De este modo, este trabajo de investigación pretende comprobar si realmente existe una correlación positiva entre el sentimiento de Twitter con respecto la influencia del *Brexit* en el Banco Santander y el precio de cotización de este último.

3.3. Análisis de sentimientos

Para aportar rigor a esta investigación es preciso intentar definir un concepto tan diverso como es el de análisis de sentimientos.

El análisis de sentimientos, también conocido como *opinion mining*, es el estudio de la información contenida en un texto para identificar reacciones, actitudes, contexto y emociones automáticamente (Fernández., 2018). El contexto sería el conjunto de palabras que rodean a otra y que implica el desarrollo de un significado a partir de un proceso de desambiguación usando tesauros (Rayón, 2020b)¹⁰.

Sin embargo, el 90% de la información que aparecen en los medios mundiales, especialmente en Twitter, se calcula que está en formatos no estructurados¹¹ (Vicenteño, 2017)¹², es decir, necesita pasar a un enfoque de descubrimiento de información y conocimiento para poder ser analizada y sacar conclusiones al respecto. De ahí la importancia de técnicas como la categorización de texto, el procesamiento del lenguaje natural, la extracción y recuperación de la información o el aprendizaje automático.

El procesamiento del lenguaje natural (PLN) se define como el campo de conocimiento de la Inteligencia Artificial cuyo objetivo es hacer eficaz la comunicación entre las personas y las máquinas mediante el procesamiento de las comunicaciones humanas e identificando los elementos más relevantes del mensaje (Decide, 2019). Se utiliza actualmente para diversas funciones, entre ellas destaca la detección de sentimientos.

Del mismo modo, el análisis de sentimientos presenta dos enfoques: el aprendizaje automático (*machine learning*) y el enfoque semántico.

El primero consiste en entrenar un modelo a través de un algoritmo de aprendizaje supervisado a partir de una secuencia de textos (corpus). Los corpus llevan asociados unas características semánticas. Los algoritmos pueden ser de regresión o de

¹⁰ Colección de palabras o términos con relaciones entre ellos, empleados para representar conceptos y usados en aplicaciones de PLN como Wordnet.

¹¹ Información que por su naturaleza no está almacenada en bases de datos o que está basada en texto (Vicenteño, 2017).

¹² Juan Vicenteño es parte del área comercial de Oracle México responsable de posicionar productos de *Business Intelligence*.

clasificación: de regresión, proporcionan un valor numérico, los de clasificación asignan una categoría. Se conoce como un procesamiento de lenguaje natural estadístico (Rayón, 2020b).

El segundo se caracteriza por el uso de un diccionario de *lexicons*, conjuntos de palabras puntuadas positiva, neutra y negativamente. Desde este enfoque, es necesario pre procesar el texto, dividirlo en palabras con la apropiada eliminación de *stopwords* y normalización lingüística por *stemming*, para posteriormente asignar un valor de polaridad a las palabras derivadas del texto. Es lo que se ha denominado procesamiento de lenguaje natural simbólico (Rayón, 2020b).

Este trabajo de investigación ha decidido utilizar para la comparación final el segundo enfoque al poder modificar el *lexicon* de manera manual en caso de que fuese necesario, pudiendo obtener una precisión tan alta como se quisiera del sentimiento en Twitter. Si se hubiera optado por utilizar el primer enfoque, no habría sido posible, dado que existe una mayor complejidad para modificar la secuencia de textos, crear una estructura de reglas de la gramática del lenguaje natural y garantizar la validez de los resultados. No obstante, el estudio también ha mostrado un ejemplo de este enfoque y ha comprobado que existen otras limitaciones para su uso en los resultados.

3.3.1. Limitaciones en Twitter

El análisis de los sentimientos en Twitter destaca por ser más complejo para las técnicas usuales de PLN por diversos motivos. Twitter nos proporciona una magnitud de datos enorme, datos que se publican de diversos medios diferentes lo que provoca una mayor frecuencia de errores ortográficos y mayor complejidad para captar los sentimientos en la red social. Además, los usuarios de Twitter publican mensajes cortos sobre una variedad de temas siendo la longitud media de los tweets de 14 o 78 caracteres. Esto hace que las técnicas de PLN deban adaptarse para poder entender la esencia del mensaje. (Go & Bhayani, 2009).

A la hora de analizar y clasificar los sentimientos de twitter hay que tener claro una serie de desafíos a los que se enfrentan las técnicas de PLN, muchos de ellos relacionados con la extracción del sentimiento oculto basado en el uso de la lengua:

- Determinar si el tweet expresa una opinión o es una noticia con intencionalidad alguna.
- Tomar en consideración si la información proporcionada es útil y trata sobre el tema que nos interesa.
- Determinar la presencia de *bots*, aplicaciones que automatizan funcionalidades o simulan la actividad de un usuario.
- Detectar palabras de la misma familia léxica que aparecen representadas por variantes morfológicas y de las cuales los diccionarios para analizar sentimientos solo guardan algunas.
- Determinar la polaridad de los textos pudiendo tener palabras positivas y negativas en la misma frase (García., 2014). Ej. “**Me alegre** de que se haya acabado, **qué horror** de actuación”. Aplicable a todos los textos.
- El dilema que genera el uso de distintos idiomas en la red social.
- Las frases sarcásticas son muy difíciles de interpretar: “*What a great car! It stopped working in two day*”.

A pesar de estas limitaciones, Twitter como hemos explicado con anterioridad, es una red de uso masivo, con grandes cantidades de datos e información lo que le convierte en la mejor red social para analizar los sentimientos respecto a un tema. En el capítulo de metodología se explicará que en este trabajo se han podido corregir algunas de estas limitaciones.

4. Metodología

La metodología escogida para este estudio es compleja y adaptada a las técnicas de *Business Analytics*. En este capítulo se ha redactado con detalle el distinto uso de estas herramientas y los procesos que se han llevado a cabo. Dado que para el análisis se ha requerido el uso de técnicas diversas, este capítulo lo hemos dividido en:

- **Descarga de datos.** En este apartado se expone cómo se ha realizado la descarga de los datos para el posterior análisis, así como métodos alternativos que se han valorado para este estudio.
- **Tratamiento de los datos.** Aquí se ha explicado cómo se han abordado las limitaciones de los datos de Twitter para impedir una incoherencia y un sesgo en el análisis.
- **Análisis exploratorio.** En esta sección, se expone una introducción de cómo se ha llevado a cabo un análisis exploratorio de los datos través de R, el cual se explicará con profundidad en el capítulo 5.
- **Análisis de sentimiento.** Este apartado explica los distintos pasos a realizar para llevar a cabo un análisis de sentimiento más preciso: análisis léxico, que consiste en la normalización del texto y la identificación de las palabras raíces; y el análisis semántico, donde se exponen los distintos *lexicon*. Por último, se presenta un enfoque alternativo basado en el aprendizaje automático.

4.1. Descarga de datos

En esta investigación se han realizado dos descargas de datos importantes. En primer lugar, hemos extraído los datos de la cotización del Banco Santander. En segundo lugar, hemos realizado una descarga de tweets históricos a partir de una librería de Python denominada *GetOldTweets3*.

4.1.1. Datos financieros del Banco Santander

En un primer momento, se han descargado los datos de *Yahoo Finance*. Éstos representan el valor de la cotización del Banco Santander en el NYSE desde enero de 2016, año cuando tuvo lugar el referéndum del Reino Unido, hasta marzo de 2020, después de la

salida oficial del Reino Unido de la Unión Europea. Todos los datos son semanales y están dispuestos en orden temporal. Su valor está representado en dólares estadounidenses y en formato de distintas columnas: *Date*, *Open*, *High*, *Low*, *Close*, *AdjClose* y *volumen*. El precio de cierre (*Close*) es el precio de cierre después de los ajustes por *splits*. El precio de cierre ajustado (*AdjClose*) es el precio de cierre ajustado por dividendos y *splits*.

Estos datos han sido tratados antes de hacer uso de ellos. El motivo de su limpieza ha sido que *Yahoo Finance* trabaja con los decimales en formato americano, es decir, el decimal es punto y no coma. Cuando estos datos se exportan a Excel en formato español, los decimales son comas y no puntos. El punto representa un millar lo que supone que el precio de las acciones de Santander estaba representado en millares. Finalmente, tras realizar ciertos ajustes, se obtuvo el *dataset* para su posterior análisis en R.

4.1.2. Datos de Twitter

En ocasiones se confunde el *text mining* con la recuperación de la información. Sin embargo, son diferentes. La recuperación de información consiste en la recuperación automática de documentos relevantes mediante indexaciones de textos, clasificación, categorización, etc. (Rayón, 2020a). Generalmente se utilizan palabras clave para encontrar una página relevante. En cambio, el *text mining* se refiere a examinar una colección de documentos y descubrir información no contenida en ningún documento individual (Rayón, 2020a). En este caso hemos utilizado Twitter para recuperar información y recoger aquellos mensajes que coinciden con un criterio dado: “Banco Santander” y el “*Brexit*”.

Desde un primer momento, la intención de este estudio fue utilizar Twitter por los motivos explicados con anterioridad. En distintas investigaciones relacionadas para obtener datos de Twitter, se ha demostrado que Twitter cuenta con diferentes métodos para recolectar datos a tiempo real, entre ellos, dos *Application Programming Interface* (API¹³): *Rest API*, que permite hacer consultas de información específica de usuarios y

¹³ *Application programming interface*: se trata de un conjunto de peticiones http estructuradas que devuelven datos en un formato ágil y eficiente.

tweets y buscar tweets recientes; y *Streaming API* la cual establece una conexión al “stream” de tweets mientras éstos son generados (Twitter developer).

Sin embargo, la extensión de ambas APIs es muy limitada. Solo permite obtener los últimos 3.200 tweets en tiempo real (Twitter developer). En el estudio, esto suponía que había muy pocas probabilidades de encontrar un tweet relacionado con estos dos conceptos, dado el volumen de mensajes que se generan cada día en Twitter. Además, también existía la imposibilidad de recoger un periodo histórico suficientemente amplio.

Tras distintas búsquedas y alternativas diferentes, se descubrió una librería de *Python* personalizada que permite extraer todos los tweets históricos relevantes a través de una búsqueda sin limitaciones. Esta librería se conoce como *GetOldTweets3*. Fue creada por Dmitry Mottl como una bifurcación de mejora del *GetOldTweets* de Jefferson Henrique (GitHub, 2019) y actualmente se utiliza por grandes comunidades de datos y profesionales del aprendizaje automático como *Kaggle* (Kaggle).

A partir de esta librería, la descarga de datos ha sido muy sencilla. Otra ventaja que proporciona *GetOldTweets-3* es que permite hacer una búsqueda a partir de dos o más palabras clave, cuestión que era muy complicada de hacer a través de la API de Twitter. Además, también permite establecer el periodo y el idioma de búsqueda, lo que facilita el análisis de sentimientos.

A continuación, se muestra un ejemplo de lo que se hizo en *Python* para este estudio.

Ilustración 3: Extracto del código usado en *Python* para la descarga de tweets

```
bash-3.2$ GetOldTweets3 --querysearch "Santander Brexit" --lang es --since 2016-01-01 --
until 2017-01-01
Downloading tweets...
Saved 1912
```

Fuente: Elaboración propia en Python

Como se aprecia en la Ilustración 3, la búsqueda se ejecutó a partir de dos palabras clave: “Santander” y “Brexit”. La decisión de utilizar únicamente la palabra “Santander” en lugar de “Banco Santander” se basa en que la filial británica de la entidad se conoce como “Santander UK” entonces, si hubiéramos añadido la palabra “Banco”, se habrían omitido tweets que hablan sobre el tema que tratamos.

El rango de fechas se estableció en función del periodo que se impuso con anterioridad para la descarga de datos del precio de cotización del Banco Santander. Por otro lado, con el propósito de facilitar el análisis de sentimientos, la búsqueda se ha producido a partir de un idioma. Para asegurarnos de que se establecía el idioma correcto, se hicieron dos descargas de tweets, unos con idioma español y otros con idioma inglés. Finalmente, se escogieron los tweets con idioma español al ser su cantidad superior que la de los tweets con idioma inglés (3.286 tweets vs 2.375).

Sin embargo, Twitter no es perfecto como cualquier otro sistema software. Esto hace que, en ocasiones, las consultas a partir de este código de *Python* se limiten a un determinado número de tweets. Esto ha hecho que los tweets se hayan tenido que descargar mes a mes manualmente. Aun así, el número de tweets que hablan sobre el Brexit y el Banco Santander es muy reducido (3.286) en comparación con los miles de millones de tweets que se pueden generar durante cuatro años. El estudio considera que esa limitación está determinada por aquellos tweets que se han borrado por los usuarios o por la propia plataforma, lo que provoca defectos en la descarga.

Los datos descargados conforman un *dataset* que incluye distintas columnas: la fecha y la hora, el nombre de usuario, las respuestas, los retweets, los favoritos, los mensajes de los tweets, la geolocalización, las menciones, los hashtags, el id, y el enlace. Cabe señalar que, por el momento, esta librería no tiene acceso a los datos geográficos de un tweet (Beck, 2020), lo que significa que en la columna de geolocalización del *dataset* descargado no hay datos disponibles y, por tanto, no es posible su análisis.

4.1.2.1. Curiosidades en la descarga

En un principio, ante la imposibilidad de utilizar las API de Twitter para el estudio, se barajó la posibilidad de utilizar otras alternativas para la descarga de los datos.

Se quiso hacer *webscrapping*, una técnica con la que mediante software se extrae información o contenido de un sitio web (Rayón, 2020a). Sin embargo, a parte de la complejidad que implicaba, algunos sitios webs no permitían esa extracción por cuestiones legales.

En segundo lugar, después de la consulta a distintos expertos, se dio la opción de usar una API de noticias como *Newsapi*¹⁴, que permite obtener todas las apariciones en medios de comunicación a partir de una *keyword*. Esta opción se consideró interesante pero también estaba limitada. Su extensión gratis abarcaba únicamente 100 búsquedas. A través de una cuenta *premium* o de pago se conseguía una ampliación de dicha extensión, sin embargo, el uso de este tipo de herramientas fue descartado desde un primer instante dado que el fin de este trabajo no era “comprar” una parte de la investigación sino obtener conclusiones únicas mediante los datos que se lograran conseguir a través de los propios medios. Además, el estudio garantizaba de esta manera que los datos descargados eran los adecuados para la investigación.

Por último, también se antojó el uso de herramientas de *social listening* como *Awario*, que, a partir de una palabra clave, permiten acceder a todo lo que han encontrado en Internet de ello, en este caso el Banco Santander y el *Brexit*.

Awario es una de las herramientas de *social listening* más completas y permite buscar dos palabras clave en una misma noticia en cualquier red social o medio de comunicación. Sin embargo, esta búsqueda no sigue un procedimiento normal, sino que es una búsqueda booleana. Se define por ser manual y permitir a los usuarios combinar palabras clave con operadores booleanos para producir resultados más relevantes. A través de una prueba gratis de siete días, se descubrió que este tipo de herramientas proporcionan todo tipo de datos, desde el número total de opiniones o noticias que tratan sobre el tema búsqueda, el alcance de las cuentas a las que van asociadas las noticias hasta el porcentaje de las menciones de las palabras clave por país para considerar cuál es el que genera más información sobre el tema al respecto.

Además, se pudo visualizar que el número total de datos que existen sobre el *Brexit* y el Banco Santander de manera conjunta es muy reducido en comparación con lo que se puede encontrar de ambos conceptos de manera separada, por lo que es lógico que el número de datos en Twitter sobre esta cuestión sea limitado.

¹⁴ Fuente de datos para obtener noticias y otro tipo de aplicaciones donde se desea mostrar a sus usuarios titulares en vivo. Se hace un seguimiento de titulares en 7 categorías en más de 50 países.

El problema que entraña este tipo de herramientas es que tienen un coste muy elevado y, además, no permiten exportar datos para analizarlos a través de un lenguaje de programación.

4.2. Tratamiento de los datos

El análisis de datos extraídos de redes sociales se enfrenta a varios obstáculos y, por ello, hay que tratarlos. Entre ellos, la suciedad de los datos, que se evidencia en los signos de puntuación, las etiquetas HTML, los caracteres sueltos, los patrones no informativos, las etiquetas div o los URLs de páginas web. También nos enfrentamos a la presencia de *bots*, software automático que publica contenidos simulando ser una persona, que pueden sesgar el contenido o hacer que los datos sean inconsistentes.

Es por ello por lo que antes de llevar a cabo el análisis de sentimientos, el proceso de limpieza de texto y tratamiento de datos se ha constituido como un apartado fundamental en este estudio.

4.2.1. Suciedad de los datos

Para el desafío que genera la suciedad de los datos, se ha utilizado una función propia de R que tiene la ventaja de poder adaptarse fácilmente dependiendo del tipo de texto analizado. Esta función se conoce como *Gsub*. Lo que hace es “buscar” por patrones de texto dentro de cada uno de los elementos de un vector, para luego reemplazarlos con otro texto que queramos (digitalcookhouse , 2014).

A continuación, vemos algunos ejemplos de la limpieza del texto propio de los tweets, el cual será el que se analizará para captar los sentimientos.

Ilustración 4: Extracto del código de Rstudio de limpieza de texto

```
#Eliminamos los "rt"
datostfg_2$text <- gsub("RT", "", datostfg_2$text)
#Eliminamos los usuarios
datostfg_2 $text <- gsub("@", "", datostfg_2 $text)
#Eliminamos los enlaces
datostfg_2 $text<-gsub('https://', "", datostfg_2 $text)
#Eliminamos los tabuladores
datostfg_2 $text <- gsub("[ \\t]{2,}", " ", datostfg_2 $text)
#Eliminamos los espacios en blanco del principio y del final.
datostfg_2 $text <- gsub("^ ", "", datostfg_2 $text)
datostfg_2 $text <- gsub(" $", "", datostfg_2 $text)
datostfg_2 $text <- gsub(" #", "", datostfg_2 $text)
datostfg_2 $text <- gsub("w", "", datostfg_2 $text)
datostfg_2 $text <- gsub("W", "", datostfg_2 $text)
datostfg_2 $text <- gsub("http://", "", datostfg_2 $text)
datostfg_2 $text <- gsub(">>", "", datostfg_2 $text)
```

Fuente: Elaboración propia en lenguaje de programación de Rstudio

Por otro lado, el *dataset* obtenido de Twitter también contenía valores ausentes en columnas tales como las menciones o los *hashtags*. Con el propósito de constituir un *dataset* con filas completas, a través de técnicas de análisis exploratorio en R se han eliminado esos valores ausentes (también conocidos como *NAs* o *missing values*) mediante su sustitución por un “0”.

4.2.2. Identificación de bots

A continuación, es preciso analizar la posible existencia de *bots* puesto que, en Twitter, al ser una de las redes sociales más usadas, podremos encontrar de muy diversas intencionalidades. Los *bots* no suelen expresar cargas de sentimientos y si lo hacen es porque son manipulados (no reales) para crear un sentimiento concreto. De este modo, si no se eliminasen, el estudio estaría sesgado.

Zi Chu, Steven Gianvecchio y Haining Wang propusieron en 2012 unos criterios para reconocer si una cuenta de Twitter podría ser un *bot* (Chu, Gianvecchio, & Wang, 2012).

Entre ellos:

- Los tiempos periódicos y reguladores son un indicador de la automatización.
- El Tweet contiene spam conocido.

Dado que actualmente no existe ninguna forma para detectar los *bots* automáticamente (las máquinas los tratan como si fueran humanos), este estudio ha investigado individualmente los posibles *bots* existentes en el conjunto de tweets a tratar (3286).

Tras un intensivo análisis de los datos siguiendo tanto los criterios expuestos con anterioridad como los del propio estudio, se han identificado una serie de patrones de comportamiento cuyo objetivo difiere al del resto de mensajes que hablan sobre el impacto del *Brexit* en el Banco Santander:

Tabla 4: Patrones de comportamiento identificados como Bots

Patrones	Tweets
<p>Mensajes que no hablan del Brexit ni del Banco Santander. Se refieren al concepto de “Santander” como ciudad o puerto (suelen mencionar paralelamente al presidente del Gobierno) o hacen referencia a la Liga Santander.</p>	<p><i>Ej: Tres jugadores de LaLiga Santander, con el futuro en el aire por el Brexit</i></p>
<p>Mensajes que se repiten automáticamente al acosar a ciertos hashtags. Los hashtags más destacados son: #EstafaBancoPopular, #UE, #ppcorrupto, #Deguidos, #GuindosPinocho.</p>	<p><i>Ej: "Consigna" salvar a vida o muerte de la quiebra al Banco Santander con la sangre de 300.000 familias. #EstafaBancoPopular #UE #ppcorrupto #Deguidos #ECB #bancosantander #UE #Commission #ParlamentEU #ECB #CNMV #BdE #BCE #brexit #GuindosPinocho</i></p>
<p>Mensajes que buscan ganar seguidores.</p>	<p><i>Ej: Bolso bandolera de cuero .. Precioso #moda #estilo #tendencias #bandolera #bolso #cuero #artesanía #MeGustaGHVIP13D #madrid #barcelona #logroño #oviedo #santander #nimes #marseille #vinted #lyon #paris #zara #europe #uk #brexit #zaragoza #salamanca</i></p>
<p>Usuarios que se identifican ellos mismos como bots.</p>	<p>Ej:</p> <ul style="list-style-type: none"> • Usuario: BotEconomía: • Tweet: El Brexit aleja aún más la fusión de gestoras de Santander AM y Pioneer (Unicredit): La estancada negociación

Mensajes del mismo autor que se repiten periódicamente a causa del spam.	Ej: <ul style="list-style-type: none"> • Usuario: Bancaynegocios • Tweet:Moody's cree que el "brexit" impactará en el Santander UK y el TSB
---	---

Fuente: Elaboración propia. Datos procedentes de Twitter

Durante el análisis, se ha percibido que se repiten un gran volumen de noticias publicadas por los periódicos como El Expansión o El Economista. Esto es debido a que diversos usuarios las publican con el fin de informar a sus seguidores. Estos usuarios no se pueden identificar como *bots* porque no repiten los tweets ni formulan hashtags para ganar seguidores, sino que, al contrario, comparten las noticias porque se identifican con ellas y poseen un sentimiento.

Frente a estas circunstancias, se han eliminado todos aquellos tweets que seguían los patrones de comportamiento descritos en la Tabla 4. Al disponer de un volumen muy reducido de datos, esta eliminación se ha podido realizar manualmente a medida que se realizaba el análisis. En total se han suprimido un 9,89% de los tweets por lo que el conjunto de datos final es de 2961 tweets.

Sin embargo, existen estudios que abarcan millones de datos donde no es posible explotar esta opción. Para ello la investigación considera que es posible crear funciones a través del lenguaje de programación que permiten filtrar el contenido.

A modo de conclusión de este apartado, el estudio considera que los *bots* desaparecerán en un futuro gracias a la tecnología *blockchain*, que permitirá tener una única identidad digital y un único perfil validado para navegar por todo internet de tal manera que se incrementará la seguridad y no se podrá cometer actos delictivos y similares en Twitter o cualquier red social.

4.3. Análisis exploratorio

Antes de comenzar el análisis de sentimientos, cabe realizar un análisis exploratorio para conocer características como la distribución temporal de los tweets o el número de tweets publicados. En este sentido, se han utilizado diversas herramientas de R que han permitido una visualización muy precisa. Este apartado se expondrá con detalle en el capítulo 5.

4.4. Análisis de sentimientos

Para llevar a cabo el análisis de sentimiento a través del enfoque semántico, se ha hecho uso de diversos recursos en R. Esto se debe a que para este tipo de análisis es necesario realizar dos procesos diferenciados (un análisis léxico y un análisis semántico) y R cuenta con excelentes herramientas para abordarlos con rigor. Por otro lado, también se ha realizado un *topic modelling* como ejemplo de método de aprendizaje automático.

4.4.1. Análisis léxico

En primer lugar, se ha de preparar el entorno para conseguir unos buenos resultados finales. En este subapartado existen dos fases de preprocesamiento: la normalización de texto y el *stemming*.

La normalización del texto consiste, en primer lugar, en separar los tweets en las palabras de las que se compone, eliminando las palabras de “parada” (también denominadas *stopwords*). Para ello R cuenta con un paquete denominado *stopwords*, el cual devuelve vectores del carácter de este tipo de palabras por diferentes idiomas. Este proceso se conoce como tokenización.

En nuestro estudio se ha creado un vector de idioma español, el cual se ha convertido en un *dataframe*. A éste le hemos añadido palabras como “banco” o “Brexit” y sus variantes morfológicas, puesto que son palabras que se repiten en la mayoría de los tweets y no aportan ningún tipo de significado al análisis. De este modo, hemos creado un *lexicon* de *stopwords* en español a medida, que utilizaremos con posterioridad.

A continuación, hemos realizado *stemming*. Este proceso consiste en reducir los términos de los tweets a sus raíces o forma base. El motivo de su realización es que permite aumentar la búsqueda sobre el número de palabras con sentimiento¹⁵. En R existe una función denominada *wordStem* dentro del paquete de *SnowballC* que realiza este proceso a través de algoritmos de *stemming* con múltiples idiomas. En nuestro caso hemos utilizado el idioma español.

4.4.2. Análisis semántico

Para realizar el análisis semántico la investigación ha utilizado varios diccionarios de Rstudio: *Afinn*, *bing* y *nrc*. Cada uno categoriza las palabras de los tweets de una manera diferente pero su propósito es el mismo. Como apunte, este sistema para presentar el análisis no es perfecto, dado que la mayoría de las palabras están en idioma inglés, sin embargo, son una buena aproximación para conseguir el fin de este estudio.

El *lexicon* del diccionario *Afinn* es quizás uno de los léxicos más simples y populares que puede ser utilizado ampliamente para el análisis de sentimientos (Rodrigo, 2017). Una de sus ventajas es que se puede descargar en español. Asigna a cada palabra una polaridad con un valor entre - 5 y 5, indicando las cifras negativas un sentimiento negativo y viceversa. Tiene alrededor de 2500 palabras. En su origen, las palabras han sido etiquetadas manualmente por Finn Årup Nielsen en 2009-2011 (O.Perry).

Los diccionarios *bing* y *nrc* se suelen utilizar de manera conjunta dado que ambos categorizan a las palabras en positivo, neutro y negativo respectivamente. La diferencia entre ambos se encuentra en el hecho de que el diccionario *bing* categoriza a las palabras en positivo, negativo y neutro, mientras que el diccionario *nrc* ahonda más en la emoción del mensaje al asignar a cada una de las palabras un tipo de sentimiento entre los siguientes: *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, *trust*. Además, cuenta con alrededor de 14.000 palabras. Por el contrario, el diccionario *bing* contiene 6.786 palabras.

¹⁵ Por ejemplo, una consulta sobre “continuidad” también encuentra mensajes en los que aparezca “continuo”, porque el *stem* de las dos palabras es “continuo”.

En el estudio se han juntado éstos últimos para mejorar el análisis y se han valorado todas las palabras en función de los once sentimientos expuestos anteriormente. Por otro lado, también se ha hecho uso del diccionario *Afinn* en español para facilitar la comparación con los movimientos de mercado del Banco Santander.

Dado que estos diccionarios no están especializados en el sector financiero, se han tomado en consideración distintas modificaciones para percibir de una manera más precisa el sentimiento de Twitter. Estas modificaciones han pasado por añadir a los distintos diccionarios (*lexicons*) de R palabras financieras con su correspondiente valor asignado de polaridad. La elección de los términos financieros se ha hecho de manera arbitraria de entre las palabras contenidas en el glosario financiero de la CNMV, del cual hemos extraído 50. A continuación hemos observado individualmente 200 tweets de nuestro *dataset* y tras la indagación hemos concluido añadir únicamente 22 de 50 palabras, al ser las más reiteradas entre esos tweets.

4.4.3. Método alternativo al enfoque semántico

El *topic modelling* es una herramienta de inteligencia artificial del subcampo de aprendizaje automático cuyos algoritmos son capaces de detectar y extraer relaciones semánticas latentes (o *topics*) de amplios corpus (Rayón, 2020b). Los *topics* son un conjunto de palabras que tienden a aparecer juntas en los mismos contextos. Los corpus, como se ha señalado con anterioridad, llevan asociados unas características semánticas.

Lo que se ha realizado con este método ha sido clasificar el texto de los tweets en función de uno de los ocho sentimientos que proporciona el diccionario *nrc*. En este caso no se ha añadido el diccionario *bing* ni las modificaciones correspondientes para que los resultados (se exponen en una nube) fueran más precisos y legibles.

Dado que lo que interesa es recoger las palabras semánticas y no las palabras gramaticales, el primer paso que se ha realizado en este método es eliminar las *stopwords*. En cuanto a la limpieza del texto de los mensajes, el *dataset* que se ha utilizado ya incluía el tratamiento anterior basado en la tokenización, por lo que no se ha realizado ninguna modificación. A partir de esto, se ha creado el corpus a través de R, compuesto del conjunto de palabras de los tweets o *topics* que aparecen en el diccionario *nrc*, las cuales

llevan asociadas un sentimiento. El corpus se ha convertido en un vector para poder interpretarlo como un documento. Este último proceso se ha realizado gracias a la función *VectorSource* de R.

A continuación, hemos convertido el corpus en una matriz de términos, para identificar asociaciones semánticas entre palabras a través de la función *TermDocumentMatrix* (tdm). Esta identificación se realiza a través de algoritmos y es lo que le hace diferente del enfoque semántico explicado con anterioridad.

Ilustración 5: Características de la matriz de términos

```
<<TermDocumentMatrix (terms: 5244, documents: 8)>>  
Non-/sparse entries: 18177/23775  
Sparsity           : 57%  
Maximal term length: 243  
Weighting          : term frequency (tf)  
> |
```

Fuente: Elaboración propia en Rstudio con la función *tdm*

Como podemos observar en el extracto de Rstudio (Ilustración 5), hemos obtenido 5.244 términos, una cifra bastante alta para los pocos datos de los que disponemos. Por último, hemos convertido esa matriz en un objeto de clase *matrix*, cuyas columnas corresponden con los ocho sentimientos del diccionario *nrc*, las filas están compuestas por las palabras del corpus, que como hemos visto son 5.244, y las celdas contiene el número de veces que la palabra aparece en los datos de twitter.

A continuación, hemos creado una nube de palabras a través de la librería *wordCloud* de R. La nube nos muestra la frecuencia de palabras por sentimiento, siendo aquellas con un tamaño más grande, las que más aparecen en los mensajes de Twitter.

El principal problema de la nube de palabras es que no nos devuelve cantidades, solo nos muestra una visualización de la frecuencia de *topics* en nuestro corpus. Los resultados no son fáciles ni evidentes de interpretar. No siempre es posible asignar un nombre único a un *topic* y existe una mala calidad de los *topics* por palabras intrusas y palabras genéricas.

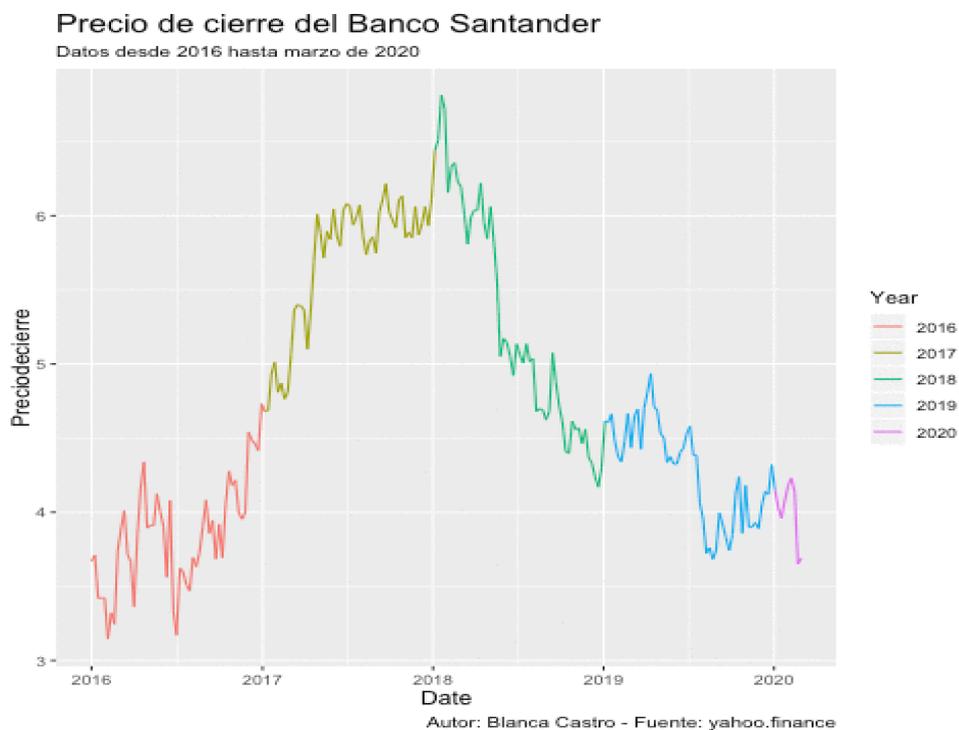
5. Análisis de Resultados

En este capítulo se ha expuesto, en primer lugar, un análisis sobre el valor de la cotización del Banco Santander a lo largo de los últimos cuatro años. En segundo lugar, un análisis exploratorio general para entender cuestiones como el volumen de tweets promedio a lo largo del periodo, o las cuentas de Twitter más mencionadas, de las cuales se sacan diversas conclusiones. En tercer lugar, se ha detallado el punto que aquí nos interesa, la correlación entre el precio de cotización del Banco Santander y el sentimiento de los tweets. Por último, también se ha estudiado, la correlación entre el volumen y el precio.

5.1. Análisis del precio de cotización del Banco Santander

El *dataset* construido tras la descarga de datos de *Yahoo Finance* ha permitido realizar la Ilustración 7, en la cual se observan los movimientos bursátiles del Banco Santander para el periodo elegido.

Ilustración 7: Precio de cierre ajustado de las acciones del Banco Santander entre 2016 y 2020



Fuente: Figura elaborada mediante ggplot+ ganimate en Rstudio.

A primera vista se aprecia que los valores más bajos de la cotización del Banco Santander en los últimos cuatro años corresponden con el mes del referéndum sobre el *Brexit* (junio de 2016). Este precio resulta alarmante y más aún cuando en los años anteriores el valor del Banco Santander era sumamente alto (rondaba los \$8) y era considerado el banco de mayor confianza en el RU. En este sentido, se puede decir que el *Brexit* ha sido un acontecimiento con gran trascendencia sobre la entidad financiera.

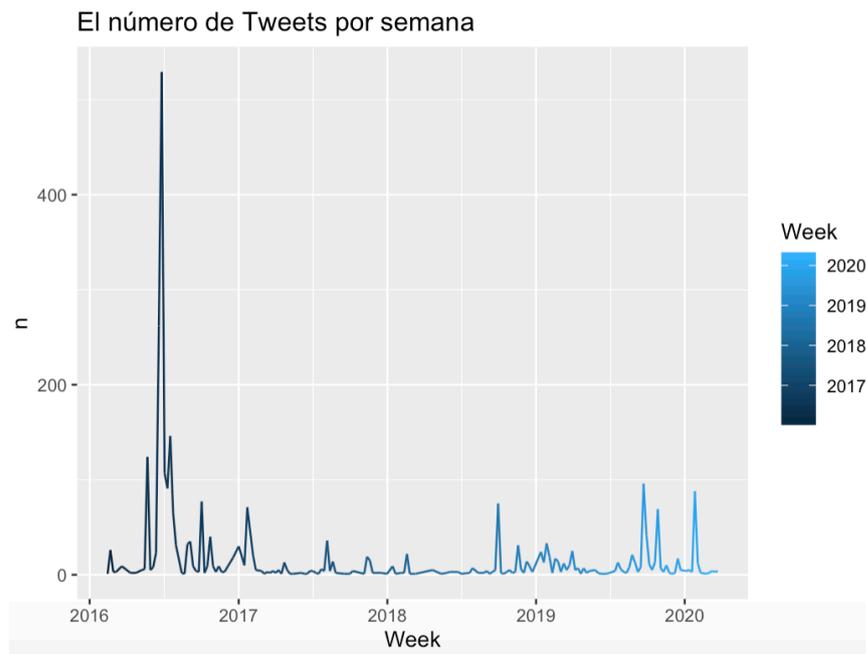
No obstante, a finales de 2016 comienza su periodo alcista en el mercado bursátil. Este periodo se mantiene durante el 2017 gracias entre otras cosas a la compra del Banco Popular por la entidad, considerado por Ana Botín como una gran oportunidad para salvar el *Brexit*. No obstante, mediados de 2018 se caracteriza por un periodo bajista, cuando la filial británica del Banco Santander se ve obligada a cumplir con las exigencias regulatorias. El periodo bajista se mantiene a principios de 2020 al cual le ha influido en gran medida la salida del RU de la UE y la pandemia del coronavirus. Si actualmente la eurozona está viviendo una recesión sin precedentes, el Reino Unido incrementará probablemente dicha desaceleración a causa del *Brexit*.

Este análisis está basado en suposiciones procedentes de la revisión de literatura. Es manifiesto que a simple vista el *Brexit* ha sido un acontecimiento relevante que ha condicionado en gran medida a la cotización en periodos concretos, pero no en su totalidad. Los cambios de *ring-fencing* y la pandemia del Covid-19 también han contribuido. Este trabajo trata de evaluar únicamente el *Brexit*. Por ello, a continuación, vamos a valorar si el impacto del *Brexit* es mayor gracias al análisis de los sentimientos en Twitter respecto al *Brexit*, posible elemento que también ha influido en las oscilaciones bursátiles.

5.2. Análisis exploratorio

En primer lugar, como se ha dicho, vamos a exponer el número de tweets por semana para conocer dónde se encuentra el grueso de los tweets que hablan sobre el Banco Santander y el *Brexit*.

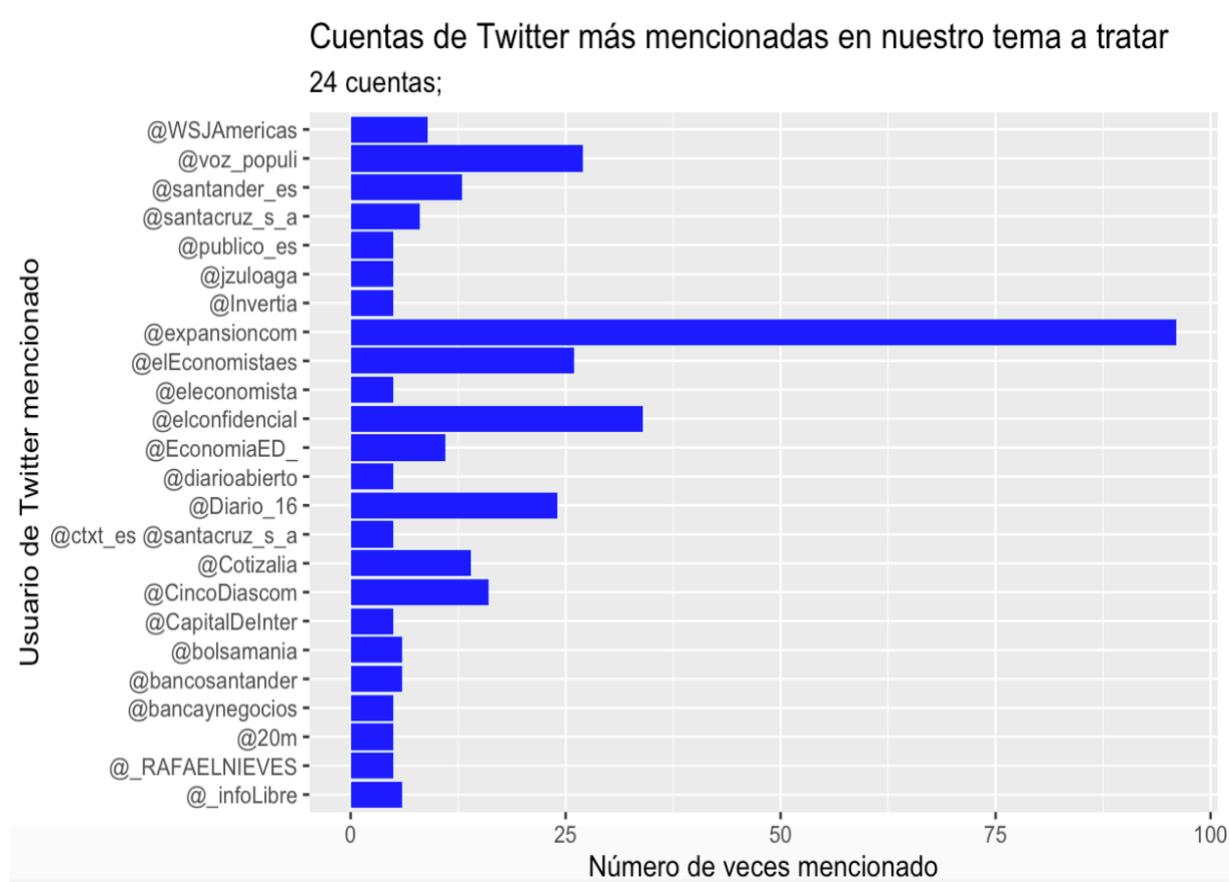
Ilustración 8: Volumen de Tweets por semana



Fuente: Elaboración propia en Rstudio

Como se puede apreciar en la Ilustración 8, el mayor volumen de los tweets se localiza entre 2016 y 2017. En concreto, en 2016, hemos hallado el 56% de los tweets que hablan sobre el Banco Santander y el *Brexit*, lo que nos hace pensar que el mayor número de sentimientos que se compartieron en Twitter respecto al posible impacto del *Brexit* en el Banco Santander fue en el año cuando se produjo el referéndum en el Reino Unido, seguramente por el desconcierto de sus resultados. No obstante, el resto de los tweets se han generado en los 3 años y medio próximos. Esto significa que se ha generado una cantidad de tweets por año muy inferior a la media, posiblemente por la demora del proceso. Sin embargo, estos tweets también son dignos de ser analizados. Más adelante, se indicará el sentimiento para comprobar cómo se han distribuido el tipo de emociones según la fecha.

Ilustración 9: Diagrama de Barras sobre los usuarios de Twitter más mencionados en el tema sobre el Banco Santander y el *Brexit*



Fuente: Elaboración propia con Ggplot+ geom_col en Rstudio

La cuenta de Twitter con mayor número de menciones (96) corresponde al periódico “Expansión”. A esta cuenta le sigue por número, “El Confidencial”, “Voz Pópuli” y “El Economista”, Todas estas cuentas, como se puede apreciar, corresponden a periódicos muy influyentes. Tras haber analizado individualmente cada mensaje de Twitter en el que dichas cuentas han sido mencionadas, todos ellos corresponden a titulares con hipervínculos a noticias de tales periódicos que se han compartido por el usuario.

Con anterioridad, hemos expuesto que las noticias también influyen en el mercado porque están motivadas por sentimientos. Por lo que el hecho de que usuarios de Twitter compartan dichas noticias es que están de acuerdo con ellas y, además, nos proporciona una información adicional: los periódicos que los usuarios suelen leer para informarse de noticias relacionadas con el Banco Santander y el Brexit se corresponden con aquellos económicos.

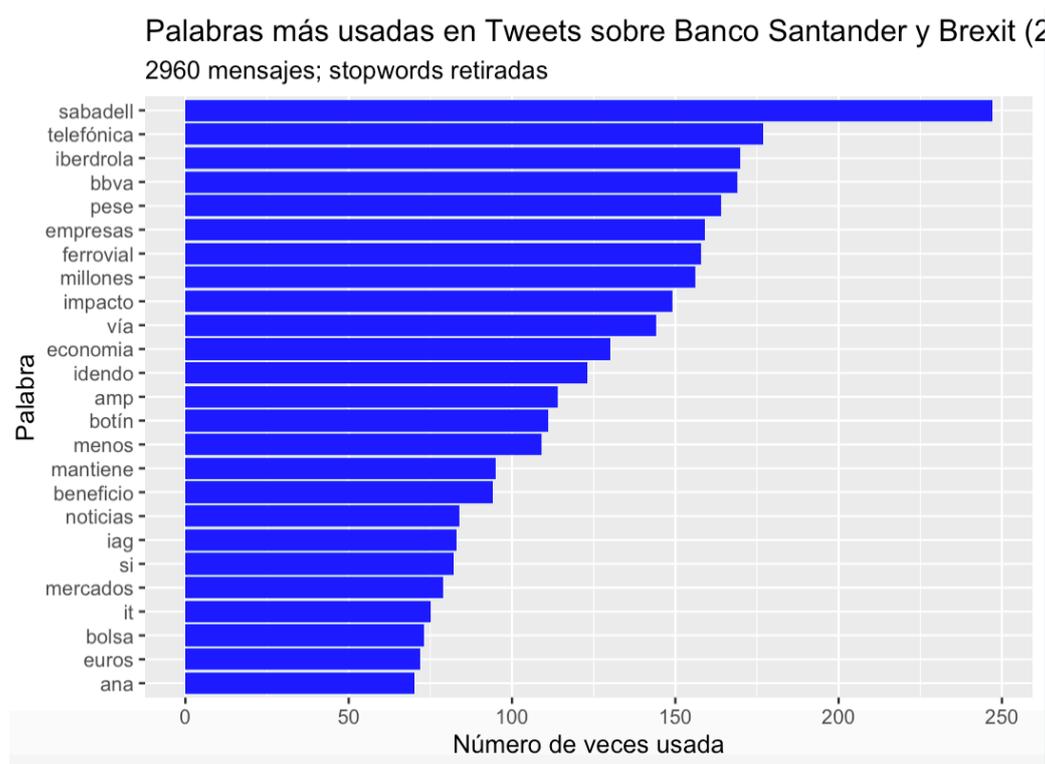
Atendiendo a los resultados del estudio publicado en 2015 sobre la medición de posicionamiento ideológico de los lectores de prensa expuesto anteriormente, podemos resaltar algunas conclusiones respecto a los periódicos que aquí nos concierne. Los resultados indican lo siguiente:

- “Expansión”. Este periódico tuvo como puntuación media un 6, lo que hace que se le asigne un perfil medio de derecha.
- “El Confidencial”. En este caso se le asignó un perfil de 5,2, puntuación que también corresponde con un perfil medio derechista, pero con tendencia a centro.
- “Voz Pópuli”. Con una puntuación media de 4,3, los lectores de este periódico presentan un perfil de centroizquierda.
- “El Economista”. Este medio obtuvo una valoración media de 4,7, lo que significa que se le ha asignado un perfil centroizquierdista, pero cercano a centro.

Después de haber analizado estos periódicos, podemos concluir que la mayor parte de usuarios que comparten las noticias sobre el tema a tratar son lectores con un perfil ideológico de centro, con una tendencia superior “derechista”, al haber “96” menciones de El Expansión y “34” de El Confidencial.

A continuación, pasamos a analizar un gráfico de barras donde se visualizan las palabras más usadas respecto al impacto del *Brexit* en el Banco Santander en Twitter. Esta ilustración (Ilustración 10) se ha realizado sin las *stopwords* porque, al contrario, aparecerían como aquellas palabras más usadas los artículos, determinantes o conjunciones que forman esa lista de palabras de “parada”.

Ilustración 10: Diagrama de barras de las palabras más usadas en el dataset de Twitter durante el periodo 2016-2020



Fuente: Elaboración propia con Ggplot + geom_col en Rstudio.

Como se aprecia en la Ilustración 10, las palabras más usadas en los mensajes de Twitter corresponden a cuatro entidades muy conocidas “Sabadell”, “Telefónica”, “Iberdrola”, y “BBVA”. El estudio considera que el motivo detrás de esto corresponde a que estas empresas se hallan en una situación similar (no igual) al Banco Santander respecto al Brexit, por lo que se entiende que en la mayoría de los tweets que nos concierne se hable también de estas empresas. Las siguientes palabras que más se repiten son aquellas relacionadas con Ana Botín, beneficios posibles, previsiones; sin embargo, no se puede deducir ninguna conclusión exacta.

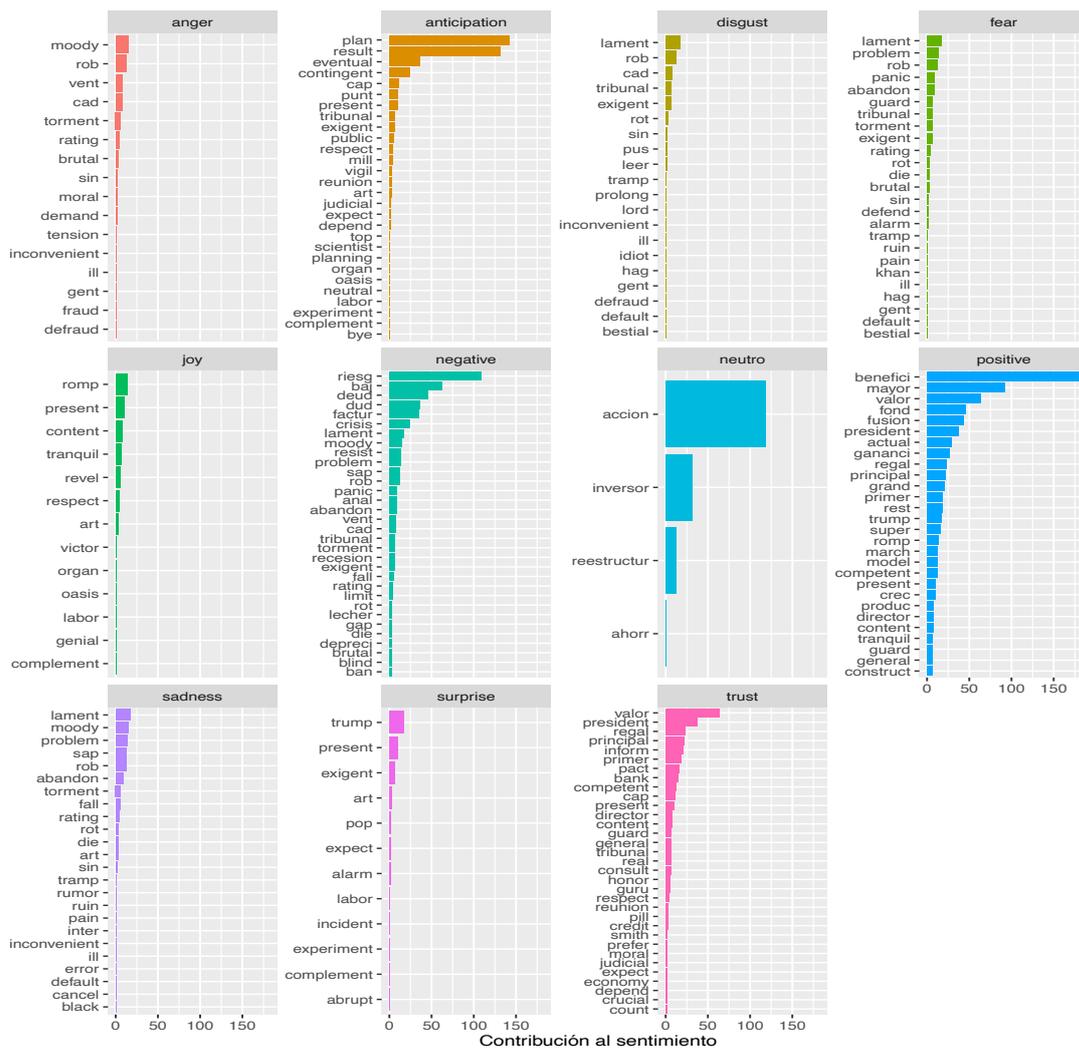
Cabe destacar que esta ilustración también se ha realizado con las palabras más empleadas haciendo *stemming*. La única diferencia entre ambas son las variantes morfológicas, por lo que no hay ninguna palabra que por su raíz destaque por ser más repetida en los mensajes.

5.3. Sentimientos

En este apartado se expondrá el análisis de los sentimientos de los tweets del estudio. En un primer lugar, se han analizado las palabras más repetidas en función de once sentimientos, los cuales se han asignado gracias a los diccionarios de *bing* y *nrc*.

Si se hace una comparación con la nube de palabras que se creó mediante *topic modelling*, hay palabras que coinciden, pero hay otras que varían. La principal razón de esto último es el método. En la nube se producen asociaciones semánticas en un contexto a través de un algoritmo para luego clasificar esos topics de los corpus en un sentimiento establecido en el diccionario. En cambio, en el análisis semántico se analizan todas las palabras de los tweets (con las correspondientes modificaciones) a partir de los términos de los distintos *lexicon*, los cuales llevan asociados un sentimiento.

Ilustración 11: Contribución de los tweets al sentimiento



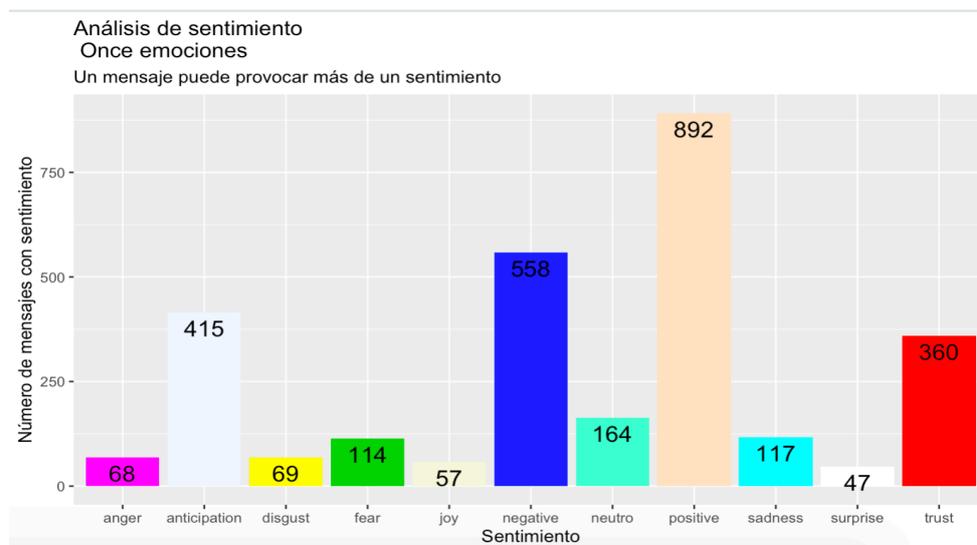
Fuente: Elaboración propia con Rstudio

Un ejemplo concreto de las diferencias entre ambos métodos se aprecia en el sentimiento de miedo (*fear*). En la nube de palabras (Ilustración 7) no aparece ningún término, en cambio, en la Ilustración 11 destacan palabras o *stem* como “lament”, y “problem”. Por el contrario, en el caso del sentimiento de anticipación (*anticipation*) sí que coinciden palabras como “plan”, “result”.

La nube de palabras no incluye los sentimientos positivos y negativos por posible opacidad en la visualización, lo que ha producido que muchas palabras recogidas en el sentimiento de anticipación en la nube se encuentren en la Ilustración 11 en sentimiento negativo. Por tanto, se concluye que es más preciso el análisis de sentimientos escogido dado que aporta un mayor volumen de información y es más preciso. Además, permite contabilizar la cantidad de palabras más frecuentes por sentimiento.

Con el propósito de conocer el tipo de sentimiento que predomina en nuestro análisis, se han unificado los diccionarios *bing* y *nrc* para aumentar el espectro de palabras y los tipos de sentimientos. A continuación, se han añadido las modificaciones relacionadas con los términos financieros a estos diccionarios, los cuales se han clasificado únicamente en negativo, positivo y neutro. Por último, se ha creado un gráfico de barras, donde en el eje X se ha representado el número de palabras acumuladas y en el eje Y cada tipo de sentimiento.

Ilustración 12: Frecuencia por tipo de sentimiento respecto al Banco Santander y al Brexit



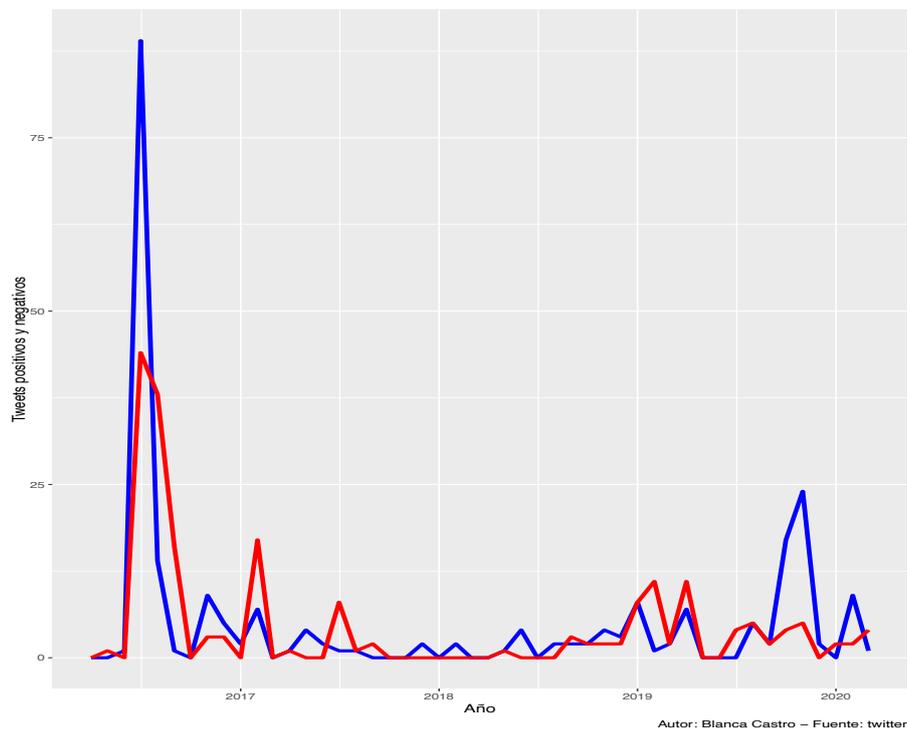
Fuente: Elaboración propia con *Ggplot+geom_bar* de *Rstudio*

Entre todos los mensajes que abundan en Twitter, los resultados de la Ilustración 12 muestran que el sentimiento predominante respecto al impacto del Brexit sobre el Banco Santander es positivo. Para ahondar en detalle, se ha calculado que el sentimiento positivo representa un 32% del total. A este sentimiento le sigue el negativo, el cual simboliza un 20 % del total. El estudio considera que la razón por la que estos sentimientos prevalecen por encima de los demás, se debe a que las palabras financieras que se han añadido solo se han categorizado en positivo, neutro y negativo. El tercer sentimiento que sobresale es anticipación, lo que quiere decir que abundan las palabras como “prever” o “planificación”, cosa que tiene sentido puesto que todavía no se han materializado las negociaciones sobre el *Brexit*. Por último, la investigación considera relevante analizar el cuarto sentimiento más frecuente, “confianza”, porque esto deja entrever que existe esperanza en que el *Brexit* no afecte en gran medida a la entidad presidida por Ana Botín.

A partir de este momento, se analizarán los sentimientos promedios de Twitter desde una serie temporal, que corresponde con el periodo del valor de cotización que hemos escogido para el Banco Santander.

En primer lugar, se ha realizado un análisis para calcular cuál es la frecuencia de los tweets con sentimientos positivos y con sentimientos negativos a lo largo del periodo elegido. Como se puede observar en la Ilustración 13, durante 2016 los tweets que más han destacado han sido los que inspiran positividad sobre el impacto del *Brexit* en la Banco Santander, concretamente representan el doble respecto a los negativos. En otras palabras, en el año cuando tuvo lugar el referéndum la mayoría de las personas en Twitter no consideraban al *Brexit* como una amenaza para el Banco Santander, a pesar del asombro de sus resultados y la caída del valor de cotización. También es posible que la población estuviera desinformada de la totalidad de las consecuencias que significaba una salida de un país como Reino Unido de la Unión Europea en términos económicos. Por otro lado, en esta ilustración también se aprecia que el mayor volumen de los tweets publicados se concentra en 2016.

Ilustración 13: Frecuencia de Tweets por sentimiento: Positivo vs Negativo



Fuente: Elaboración propia con Ggplot+geom_bar de Rstudio. Datos extraídos de Twitter.

A partir de 2017, el sentimiento es más fluctuante, siendo a principios y mediados de 2017 el periodo que más negatividad se proyecta respecto a dicho impacto. A finales de 2019, la positividad vuelve a destacar en Twitter. El motivo detrás de la subida en el volumen de tweets positivos se debe posiblemente a la cercanía de la fecha de culminación del *Brexit*, la cual ha despertado el sentimiento que se mantenía al principio. En términos de volumen, desde 2017 hasta 2019 es muy escaso. Solo en las fechas cercanas a la salida del Reino Unido de la Unión Europea ha habido un repunte en el volumen de tweets publicados.

A continuación, se va a tratar el tercer objetivo de esta investigación: estudiar si existe correlación entre el sentimiento de los tweets y el precio de cotización del Banco Santander.

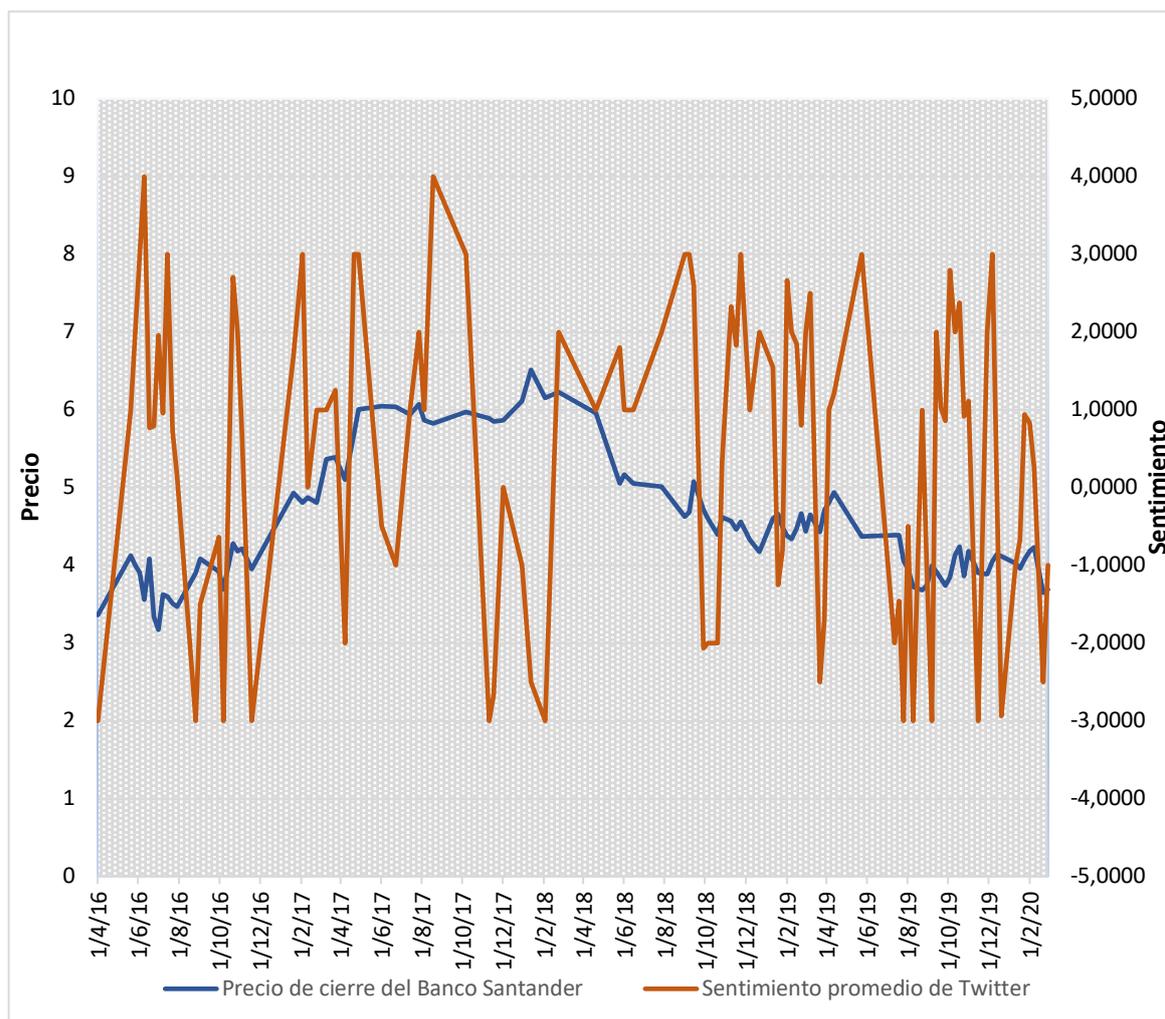
5.4. Correlación entre dos variables: sentimiento vs precio

Para estudiar si finalmente existe una correlación positiva o negativa entre los sentimientos que se proyectan en Twitter y el precio de cierre del Banco Santander, el sentimiento promedio se ha constituido como un factor fundamental para alcanzar dichos resultados.

Para ello se ha utilizado el *lexicon Afinn* en español al cual se ha incorporado el diccionario financiero que se ha construido con anterioridad en el estudio. Como se ha indicado en el capítulo 3, este tipo de diccionario califica las palabras de los tweets en función de un intervalo: -4 (sentimiento más negativo) y 4 (sentimiento más positivo). Esta categorización ha permitido facilitar el cálculo del valor del sentimiento promedio, el cual se ha llevado a cabo a través de R. Para hallar la media, se han agrupado los datos por semanas lo que ha permitido finalmente obtener el valor del sentimiento promedio por semana para el análisis.

La decisión de agrupar los datos por semanas en lugar de por meses se encuentra en el hecho de que la extracción de los datos referentes al precio de cotización del Banco Santander se ha realizado también semanalmente, lo que permite obtener un mismo eje del que partir para visualizar si coexiste una correlación entre ambas variables. Además, ante el bajo volumen de tweets que se han publicado durante estos últimos años sobre el tema a tratar, un análisis mensual podría resultar en una pérdida de información relevante.

Ilustración 14: Correlación entre el sentimiento promedio de Twitter vs Precio de cierre del Banco Santander



Fuente: Elaboración propia. Datos extraídos de R Studio.

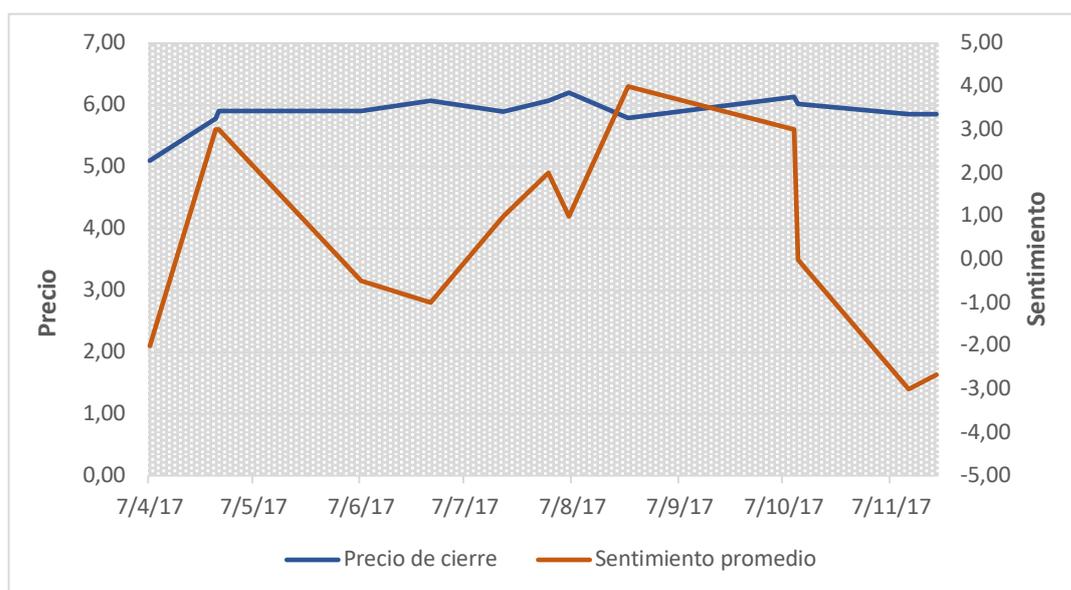
Por otro lado, para simplificar la visualización se ha escogido el periodo de tiempo habitual a lo largo del estudio, pero con ciertas modificaciones. Aquellas semanas que no se ha publicado ningún tweet (hecho muy frecuente dado los pocos datos extraídos de Twitter) se han omitido para que el gráfico fuera consistente.

En términos de resultados, se puede apreciar en la Ilustración 14 que no existe ningún tipo de correlación entre el sentimiento promedio de Twitter y los movimientos bursátiles del Banco Santander: hay periodos en los cuales existe correlación positiva y hay otros donde sobresale la correlación negativa. Por ejemplo, a principios de 2016, donde aparece el valor mínimo del Banco Santander en la bolsa respecto a los últimos cuatro años, el sentimiento que se comparte en Twitter es especialmente positivo. Paradójicamente, en

Twitter como hemos resaltado también se incluyen noticias, la mayoría compartidas por usuarios con un perfil ideológico de centro derecha según los periódicos más mencionados en Twitter (análisis procedente del estudio del marco teórico sobre el perfil ideológico de los lectores), por lo que se asume que ese tipo de noticias también proyectan positividad. En este sentido, la caída del precio de cierre del Banco Santander no estaría especialmente motivada por ese tipo de noticias que mencionan al *Brexit*. Sin embargo, hay otros periodos en los que existe correlación positiva. Es el caso de principios de 2018, donde se publican mensajes en Twitter con un sentimiento promedio positivo al mismo tiempo que el precio de cierre alcanza el pico más alto de los últimos cuatro años. Sin embargo, al no existir una correlación precisa, todas las asunciones establecidas quedarían anuladas. En otras palabras, el sentimiento proyectado en Twitter respecto al impacto del *Brexit* sobre la filial británica no influye en los mercados financieros.

Para verificar nuestros resultados, se puede observar en la Ilustración 15 la relación entre el sentimiento promedio y el valor del Banco Santander cuando tuvo lugar la compra del Banco Popular. El intervalo de fechas recorre 7 meses del año 2017, un periodo pequeño en comparación con el de la ilustración anterior, y se ha elegido con el propósito de presentar con una mayor claridad la inexistencia de dicha correlación.

Ilustración 15: Sentimiento promedio de Twitter vs Precio de cierre Banco Santander (año compra del Banco Popular por el Banco Santander)

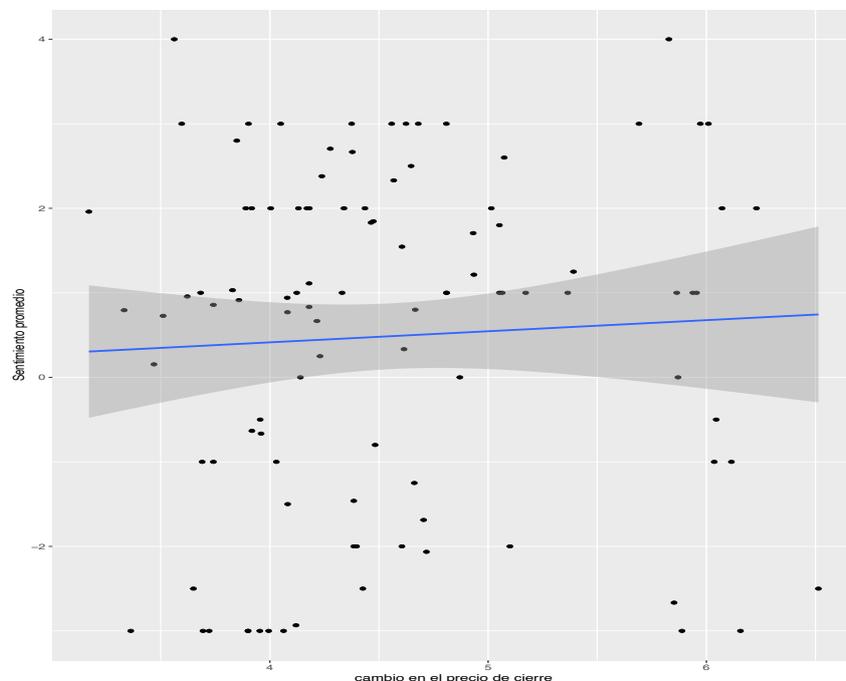


Fuente: Elaboración propia. Datos extraídos de Rstudio

La compra del Banco Popular fue un gran activo para el Banco Santander pues le permitió compensar las posibles consecuencias negativas del *Brexit*, de ahí que su valor en el mercado bursátil fuera alto y relativamente estable. Sin embargo, el sentimiento relacionado con dicha compra se caracteriza por ser muy inestable, desde parcialmente negativo en el mes de la compra (junio 2017) hasta sumamente positivo en el siguiente mes. Sin embargo, el sentimiento positivo no se mantiene, sino que tiende a bajar a partir de septiembre de 2017. Tomando en consideración el análisis individual que se efectuó para la identificación de los *bots*, el sentimiento negativo se identifica con aquellas conversaciones en línea que consideran la adquisición como una gran estafa para el Banco Popular por parte del Banco Santander, el cual adquirió el primero por un euro al verlo como una oportunidad para salvarse del *Brexit*.

La Ilustración 16 representa un diagrama de dispersión en el cual se aprecia que no hay ninguna correlación entre las dos variables puesto que no hay un patrón definido entre los puntos. De este modo, podemos afirmar que la relación entre los movimientos bursátiles y el sentimiento de Twitter respecto a un acontecimiento es aleatoria.

Ilustración 16: Diagrama de dispersión del sentimiento respecto al precio



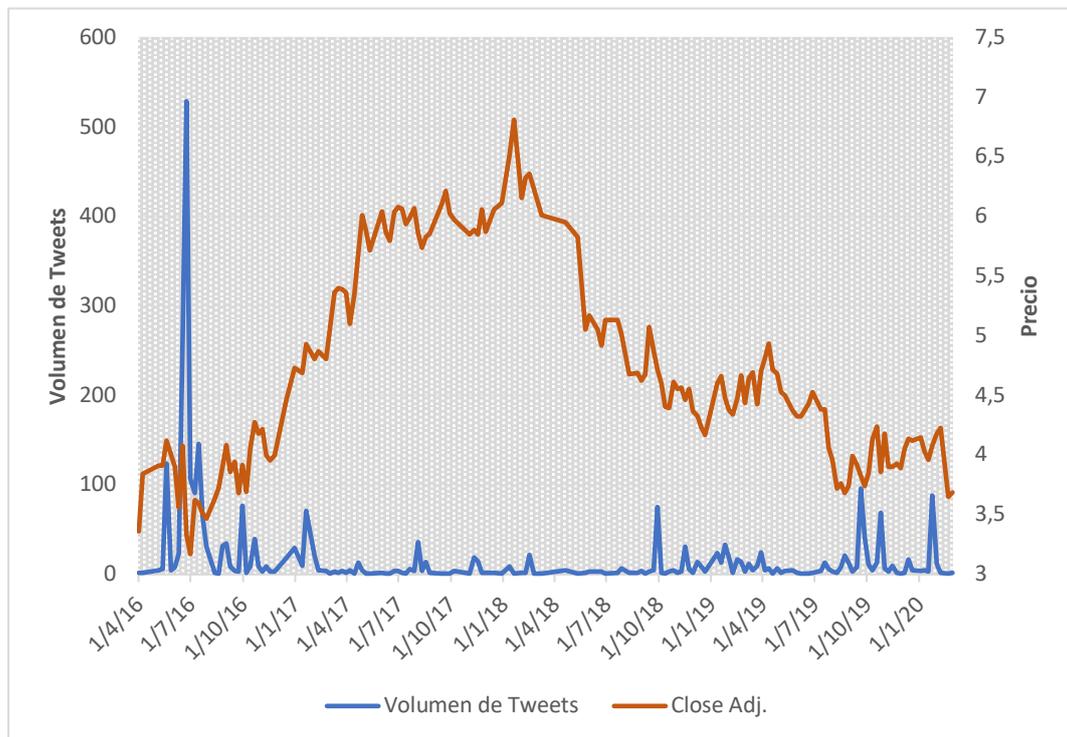
Fuente: Elaboración propia con *Ggplot* + *Geom_point* en *RStudio*

5.5. Correlación entre dos variables: volumen vs precio

A pesar de la escasez de los datos extraídos, el estudio ha considerado relevante realizar una comparación entre el volumen de los tweets y el precio de cierre del Banco Santander. Sin embargo, se puede apreciar en la Ilustración 17 que no existe ningún tipo de correlación entre ambas variables, por lo que los movimientos de bolsa de la entidad han sido influenciados por otros factores, especialmente por los acontecimientos presentados con anterioridad.

Parece que en un primer momento el referéndum sobre la salida del Reino Unido de la UE supuso una revolución en Twitter dando lugar a varias conversaciones en línea sobre las posibles consecuencias sobre el Banco Santander pero que a medida que ha pasado el tiempo ha dejado de ser relevante.

Ilustración 17: Volumen de Tweets vs Precio de cotización (de cierre ajustado) del Banco Santander



Fuente: Elaboración propia. Datos extraídos de Rstudio

Si hubiéramos analizado los tweets del Banco Santander sin considerar el *Brexit* posiblemente la cantidad de mensajes hubiera sido copiosamente mayor. Por ello, no se puede descartar la existencia de una correlación entre los sentimientos del Banco

Santander en Twitter y su valor de cotización o incluso del *Brexit* sobre la bolsa. Solo se puede concluir que los tweets respecto al impacto del *Brexit* en el Banco Santander no son una variable explicativa de los movimientos bursátiles y que los acontecimientos inesperados son factores más influyentes.

6. Conclusiones

Este trabajo de investigación ha abordado un análisis acerca del impacto de la salida del Reino Unido de la UE sobre el precio de mercado del Banco Santander. Desde su incorporación a la institución más grande de Europa, el dinamismo de la economía del RU se ha convertido en una realidad, al igual que el crecimiento del sector bancario europeo. Sin embargo, el *Brexit* ha introducido un nuevo paradigma en la región que ha provocado la introducción de nuevas medidas por parte de las entidades bancarias como una señal de preparación. En el caso del Banco Santander, a simple vista se ha descubierto que el *Brexit* ha sido junto a otros agentes (como los cambios regulatorios) un incidente que ha afectado al valor de cotización en los últimos cuatro años. No obstante, diversos estudios han reflejado que el valor de cotización está condicionado por una gran variedad de factores como es el caso de los sentimientos de Twitter.

A lo largo del tiempo, Twitter ha demostrado ser una red social que combina la formulación de opiniones y la publicación de noticias por parte de los usuarios lo que no es de extrañar que se utilice para medir los sentimientos de sus usuarios a través de técnicas de *data analytics*. Este fue el motivo por el que se eligió como elemento de medición del impacto real del *Brexit* sobre el precio bursátil del Banco Santander.

Esta investigación ha probado que estas herramientas de análisis de información son de una gran utilidad para distintas funciones: extraer datos históricos de redes sociales como Twitter; analizar sentimientos de tweets para determinar si una noticia de un periódico o un comentario en una red social tiene una inclinación positiva neutra o negativa; y evaluar una posible correlación entre los sentimientos de Twitter y el precio de mercado de cualquier entidad financiera.

Tras este estudio cualitativo y analítico, se puede concluir que el *Brexit* ha influido de alguna manera en el valor de la cotización del Banco Santander por la incertidumbre que genera y las medidas que ha llevado a cabo Ana Botín. Sin embargo, los sentimientos de Twitter respecto al tema que nos concierne no presentan ningún tipo de correlación con el precio de mercado del Banco Santander. Sabemos que, en general, abundan los sentimientos positivos en Twitter en comparación con los negativos y que las conversaciones en línea sobre este dilema son muy escasas, pero también que este

intercambio de información no ha implicado ningún cambio en los movimientos bursátiles de la entidad financiera en el NYSE.

Por otro lado, durante la investigación, nos hemos encontrado con distintas limitaciones que son importantes de señalar con el propósito de que estudios posteriores las tengan en consideración o simplemente traten de resolverlas.

6.1. Limitaciones del estudio

6.1.1. Problemas éticos en la gestión de los datos

Twitter es una red social donde se elaboran perfiles de personas a partir de los datos que se facilitan a través de tweets, mensajes privados e información compartida. Esto es lo que se conoce como la identidad digital, es decir, la discriminación de una persona basada en sus datos (reales o falsos) y es uno de los principales problemas éticos de la era de la información. Los usuarios que crean una cuenta en Twitter aceptan que su información pase a ser pública sin considerar las consecuencias éticas que implica dicha aceptación. A veces no se sabe qué información se cede, qué utilidad puede prestar o hasta que punto dicha exposición de datos puede atentar contra la seguridad y privacidad de las personas.

En este punto se alza la cuestión ética de si se puede comerciar con derechos tales como la dignidad o la intimidad de una persona. Al haber aceptado los términos y condiciones de seguridad, existe una confirmación por parte del usuario para poder hacer uso de la información para cualquier propósito. No obstante, este estudio no ha publicado la base de datos extraída de Twitter en ningún lugar público con el propósito de proteger al usuario.

6.1.2. Falta de datos

La investigación ha presentado problemas en la descarga de datos a través de la librería expuesta con anterioridad. La librería de datos utilizada no ha permitido obtener todos los tweets históricos, puesto que la plataforma suele borrar aquellos que considera que son *bots* o incluso los propios usuarios borran antiguos tweets. De este modo, el estudio también se enfrenta a una falta de datos que podrían influir en los resultados del análisis

de sentimiento. Unido a esta limitación, se encuentra el escaso volumen existente de datos sobre este tema en Twitter en comparación con la media habitual por año que recoge miles de millones de tweets.

Por otro lado, es posible que los tweets no capten el sentimiento y las intenciones de voto de los analfabetos demográficos incapaces de utilizar la tecnología como algunos ancianos, o de un nivel socioeconómico más bajo.

6.1.3. Falta de exactitud en el análisis de sentimientos

A pesar de que se han eliminado algunos de los desafíos a los que se enfrentan las técnicas de PLN, determinar la polaridad de los textos en un mensaje de 140 caracteres es muy complicado y más si existen palabras positivas y negativas en la misma frase. Al igual que los recursos retóricos u la ironía. Además, los mensajes que se han descargado de la red social están escritos en español dado que la interpretación de los sentimientos a través de un *lexicon* solo es posible si todos los mensajes se encuentran en el mismo idioma. Esto, además de omitir tweets en otros idiomas que pueden aportar información relevante, también crea una limitación en el estudio puesto que los recursos más adecuados para el análisis son en inglés. No obstante, el estudio ha logrado descargarse el diccionario *Afinn* en español y ha elaborado un diccionario muy preciso de términos financieros, también en idioma español que ha contribuido a mejorar la precisión del análisis.

6.1.4. Falta de una ponderación justa de las opiniones y las noticias

Como se ha expuesto a lo largo del trabajo, en Twitter se comparten tanto noticias como opiniones de los usuarios. El problema que se genera al tomar en consideración los datos en su conjunto (sentimiento promedio) es que omitimos la posibilidad de que las noticias compartidas por los usuarios puedan incidir en mayor medida que las propias opiniones en el precio de la cotización. De ahí que sea relevante señalar aquellos “nodos” con un mayor alcance en Twitter puesto que nos permite conocer los posibles usuarios con una mayor influencia o incidencia en el valor bursátil. Esto se realiza a través de la Teoría de Grafos, concepto matemático subyacente a la idea de red que nos permite describir como las cosas están conectadas. Si hubiéramos descubierto el alcance de las cuentas de Twitter de nuestro *dataset*, podríamos haber efectuado esta teoría.

6.1.5. Variación del precio de cotización en función del índice bursátil a analizar

Este trabajo de investigación ha utilizado los precios del Banco Santander originarios del NYSE para el análisis. Sin embargo, este banco cotiza en distintos índices bursátiles como se ha indicado en el estudio. Este hecho testifica que el precio de cotización puede variar en función del mercado en que opere. En este sentido, los resultados obtenidos en este trabajo son únicamente válidos para los precios del NYSE. Si se hubiesen elegido por ejemplo los datos del IBEX-35, las conclusiones hubiesen sido otras, con la posibilidad de haber obtenido una correlación entre los sentimientos de Twitter y la cotización bursátil.

6.1.6. Otros factores involucrados

Al comienzo del estudio se optó por elegir un periodo de tiempo para el análisis enfocado en el *Brexit*. Por ello, toda la investigación se estudia desde la fecha del referéndum respecto a la salida del Reino Unido de la Unión Europea hasta principios de marzo de 2020, poco después de la salida oficial. Sin embargo, no se ha tenido en cuenta otro factor que está influyendo tanto en la salida del Reino Unido de la UE como en los estados financieros del Banco Santander: la pandemia del coronavirus. Con este apunte, el estudio remarca que pueden existir factores imprevisibles que cambien por completo las expectativas sobre el posible impacto del *Brexit* en una entidad bancaria. En otras palabras, no se puede medir el impacto de un acontecimiento internacional sin tener en cuenta otras cuestiones.

7. Bibliografía

- Aguado, J. (2018). *Santander shifts assets to Spanish parent to ring-fence UK unit*. United Kingdom : Reuters .
- Armour, J. (2017). Brexit and financial services. *Oxford review of Economic Policy*, Volume 33, Number S1, pp.S54-S69.
- Ashurst. (2016 de Marzo de 2016). *Brexit: potential impact on the UK's banking industry*. Obtenido el 20 de febrero de 2020 de Ashurst:
<https://www.ashurst.com/en/news-and-insights/insights/brexit-potential-impact-on-the-uk-banking-industry/>
- Asociación de Medios de Información. (2017). *Libro Blanco de la Información 2017*. AMI.
- Banco Santander SA. (2019). *Informe Anual 2019*. Madrid.
- Banco Santander. (s.f.). *Santander UK plc*. Obtenido de Banco Santander:
<https://www.santander.co.uk/about-santander/our-business>
- Bank of England. (2015). *EU membership and the Bank of England*. Report. Londres.
- Bankinter. (27 de Septiembre de 2016). *Blog de referencia financiera*. Obtenido el 3 de marzo de 2020 de Llega MiFID II: ¿qué es y cómo me afecta?:
<https://blog.bankinter.com/economia/-/noticia/2014/06/04/nueva-normativa-financiera.aspx>
- Beck, M. (12 de January de 2020). *How to Scrape Tweets from Twitter*. Obtenido el 4 de abril de 2020 de Towards data science: <https://towardsdatascience.com/how-to-scrape-tweets-from-twitter-59287e20f0f1>
- Bespoke Investment Group. (7 de September de 2019). *Economic Surprise Index Turns Positive*. Obtenido 1 de abril de 2020 de Seeking Alpha:
<https://seekingalpha.com/article/4290302-economic-surprise-index-turns-positive>
- Blanco, M. J. (3 de Septiembre de 2013). Internet y Redes Sociales: ¿Un reemplazo o complemento para los medios de comunicación? *Periodismo Digital*.
- Brandwatch. (10 de febrero de 2015). *Entendiendo el análisis de sentimiento: qué es y para qué se usa*. Obtenido el 4 de febrero de 2020 de Brandwatch:
<https://www.brandwatch.com/es/blog/analisis-de-sentimiento/>

- Chu, Z., Gianvecchio, S., & Wang, H. (2012). Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? *IEEE Transactions on Dependable and Secure Computing*, Vol.9, Number 6, pp.811-824.
- Condado, N. L. (17 de Febrero de 2017). Brexit y MiFID II: Welcome to Spain. *Expansión* .
- De, D. (2020). *United Kingdom: Optimism up but risks remain*. United Kingdom: Deloitte Insights.
- Decide. (12 de septiembre de 2019). *Procesamiento del Lenguaje Natural: qué es y para qué se utiliza*. Obtenido el 10 de abril de 2020 de decidesoluciones: <https://decidesoluciones.es/procesamiento-del-lenguaje-natural-pln-o-nlp-que-es-y-para-que-se-utiliza/>
- Deloitte. (2019). *Brexit Industry Insights*. London: Core Creative Services.
- digitalcookhouse . (23 de agosto de 2014). *Limpiando data in R- Funcion SUB o GSUB*. Obtenido el 11 de abril de 2020 de digitalcook house word press: <https://digitalcookhouse.wordpress.com/2014/08/23/limpiando-data-in-r-funcion-subgsub/>
- Fernández., C. J. (19 de noviembre de 2018). *Analisis de sentimiento en R*. Obtenido el 15 de abril de 2020 de LinkedIn.: <https://www.linkedin.com/pulse/análisis-de-sentimiento-en-r-carlos-jáuregui-fernández/>
- Fernandez, R. (20 de marzo de 2020). *Twitter en España-Datos estadísticos*. Obtenido el 6 de marzo de 2020 de Statista: <https://es.statista.com/temas/3595/twitter-en-espana/>
- Garcia., L. M. (2014). *Analisis de sentimientos y predicción de eventos en Twitter*. . Memoria para optar al título de ingeniero civil eléctrico Universidad de Santiago de Chile: Universidad de Chile. Obtenido el 2 de abril de 2020 de: http://repositorio.uchile.cl/bitstream/handle/2250/130479/cf-montesinos_lg.pdf
- GitHub. (27 de november de 2019). *GetOldTweets3*. Obtenido el 3 de abril de 2020 de <https://github.com/Mottl/GetOldTweets3>
- Go, A., & Bhayani, R. (2009). *Twitter Sentiment Classification using distant supervision*. CS224N Project Report, Stanford 1 (12).
- Gonzalo, C. (2016). *El perfil ideologico de los medios de prensa españoles*. Obtenido el 2 de abril de 2020 de: <https://www.carlosgonzalo.es/el-perfil-ideologico-de-los-medios-de-prensa-espanoles/>

- Hawksworth, J., & Teow, J. (2019). *UK Economic Outlook: Explore the causes and drivers of the UK's productivity challenge and the gains from closing the productivity gap*. London: PwC.
- Instituto Nacional de Estadística. (16 de octubre de 2019). *Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares*. Obtenido el 7 de marzo de 2020 de Instituto Nacional de Estadística: https://www.ine.es/prensa/tich_2019.pdf
- Inversor Ediciones, SL. (2020). *finanzas.com*. Obtenido el 14 de marzo de 2020 de economía digital: https://www.finanzas.com/mercados/santander-uk-cierra-140-oficinas-en-reino-unido-con-efecto-en-1-270-empleos_13981967_102.html
- Kaggle. (s.f.). *Trump Tweet Analysis*. Obtenido de kaggle: <https://www.kaggle.com/rahulanand0070/trump-tweet-analysis>
- Luis, A. d. (23 de diciembre de 2019). *¿Cómo afecta al mercado las noticias?* Obtenido el 18 de marzo de 2020 de Sala de Inversión: <https://www.saladeinversion.com/articulos/como-afecta-al-mercado-las-noticias/>
- Macro, D. (2019). *Reino Unido-Balanza Comercial* . Obtenido de 3 de marzo de 2020 datosmacro.com: <https://datosmacro.expansion.com/comercio/balanza/uk>
- Moreno, M. (2015). *Cómo triunfar en las redes sociales*. Barcelona: Grupo Planeta .
- O.Perry, P. (s.f.). *AFINN Sentiment Lexicon*. Obtenido el 3 de abril de 2020 de Corpus Text: http://corpustext.com/reference/sentiment_afinn.html
- OCDE. (2019). *Economic Outlook*. París: OECD Publishing.
- OECD. (2015). *OECD Economic Outlook*. París : OECD Publishing .
- OECD. (2015). *OECD Economic Surveys: United Kingdom*. París : OECD Publishing .
- Omicore. (10 de February de 2020). *Twitter by the Numbers: Stats, Demographics & Fun Facts*. Obtenido el 14 de marzo de 2020 de Omicore: <https://www.omnicoreagency.com/twitter-statistics/>
- PwC. (2017). Planning for Brexit: Operational impacts on wholesale banking and capital markets in Europe. *Association for Financial Markets in Europe*.
- Rayón, A. (2020a). *API y Webscraping: técnicas de enriquecimiento de bases de datos*. Materiales docentes Diploma en Fundamentos en Business Analytics, Universidad Pontificia Comillas, Madrid.
- Rayón, A. (2020b). *Text Analytics*. Materiales docentes Diploma en Fundamentos en Business Analytics, Universidad Pontificia Comillas, Madrid.

- Reputation, N. (9 de diciembre de 2015). *El perfil ideológico de los lectores de prensa. Análisis encuestas #7 Del debate decisivo*. Obtenido el 20 de marzo de 2020 de SM Reputation Metrics:
<https://smreputationmetrics.wordpress.com/2015/12/09/el-perfil-ideologico-de-los-lectores-de-prensa-analisis-encuestas-7deldebatedecisivo/>
- Rodrigo, J. A. (diciembre de 2017). *Text mining con R: ejemplo práctico Twitter*. Obtenido el 5 de abril de 2020 de Rpubs: https://rpubs.com/Joaquin_AR/334526
- Sánchez, M. G. (13 de marzo de 2017). *La importancia de Twitter y los mercados*. Obtenido el 20 de marzo de 2020 de BBVA: <https://www.bbva.com/es/la-importancia-de-twitter-y-los-mercados/>
- Santander Board. (2019). Santander Board approves first 2019 dividend and goodwill adjustment of its UK subsidiary. Ciudad Grupo Santander, Madrid, España: Corporate Communications.
- Santander UK Group Holdings plc. (2019). *2019 Half Yearly Financial Report*.
- Santander UK Group Holdings plc Q2. (2015). Investor Update Q2 2015 Results. London, United Kingdom.
- Santander UK plc. (2015). *Annual Report 2015*.
- Sayans, D. C. (15 de julio de 2019). *¿Y si la operativa sobre el sentimiento bursátil en Twitter genera un retorno del 62%?* Obtenido de 3 de marzo de 2020 Estrategias de inversión.: <https://www.estrategiasdeinversion.com/analisis/bolsa-y-mercados/informes/y-si-la-operativa-sobre-el-sentimiento-bursatil-n-427977>
- Sowels, N. (2018). The Uncertain Consequences of Brexit for UK-Based Financial Services. *Observatoire de la société britannique*, Vol. 21, number 1, pp. 195-211.
- Statista. (5 de febrero de 2019). *Uso de medios de comunicación*. Obtenido el 20 de febrero de 2020 de Statista: <https://es.statista.com/estadisticas/491058/consumo-diario-de-medios-de-comunicacion-en-espana-por-tipo/>
- Statista. (2 de marzo de 2020). *Evolución anual del valor de la capitalización bursátil del Banco Santander a nivel global desde 2010 hasta 2019*. Obtenido 20 de mayo de 2020 de Statista:
<https://es.statista.com/estadisticas/549408/capitalizacion-bursatil-del-banco-santander-global/>

- Twitter. (23 de mayo de 2019). *Guía de estilo para la selección*. Obtenido el 23 de marzo de 2020 de Twitter: <https://help.twitter.com/es/rules-and-policies/curationstyleguide>
- Twitter developer. (s.f.). *Things developer should know*. Obtenido de Twitter: <https://developer.twitter.com/en/docs/basics/things-every-developer-should-know>
- Veiga, E. R. (16 de diciembre de 2016). *De periodísticas, hechos y opiniones*. Obtenido el 20 de marzo de 2020 de <https://www.investigacionyciencia.es/blogs/ciencia-y-sociedad/98/posts/de-periodistas-hechos-y-opiniones-14804>
- Vicenteño, J. (2017). Oracle Big Data Discovery: Análisis Exploratorio y descubrimiento de información. . *Oracle Technology Network*. Obtenido el 3 de abril de 2020 de <https://www.oracle.com/technetwork/es/articles/database-performance/big-data-discovery-analysis-3437179-esa.html>