



Facultad de Ciencias Económicas y Empresariales
(ICADE)

*¿Puede afectar la opinión en Twitter sobre la
fusión de CaixaBank y Bankia en el precio de
cotización de CaixaBank?*

Autor: Santiago Sánchez Martínez-Corbalán

Director: María Coronado Vaca

MADRID | JUNIO de 2021

Resumen

Tanto Bankia como CaixaBank se encuentran entre las entidades financieras con mayor cuota de mercado en España. Durante los años 2020 y 2021 han protagonizado uno de los mayores movimientos en el sector bancario. La idea de este movimiento provoca una gran cantidad de opiniones entre los distintos usuarios haciendo este hecho atractivo para el estudio e investigación.

Por ello, en el presente trabajo se ha procedido a explicar la historia del sector bancario y los movimientos que preceden a este movimiento con el objetivo de que el lector comprenda el por qué de esta fusión. Tras una breve contextualización, se ha tratado de llevar a cabo un análisis empírico empleando las distintas técnicas que nos proporciona el Big Data, con el objetivo de realizar un análisis de sentimientos con la red social de Twitter y como ha evolucionado el precio de cotización de la entidad financiera CaixaBank. Para llevar a cabo el análisis empírico ha sido necesario la utilización de los lenguajes de programación R y Python. Tras realizar el análisis, las opiniones analizadas de Twitter respaldan en cierta manera el movimiento propuesto por las entidades, así como se ha reflejado de una manera positiva en la cotización de CaixaBank.

Palabras clave

CaixaBank, Bankia, fusión, análisis de sentimientos, Twitter, Python, R, precio de cotización, *Business Analytics*, *Machine learning*.

Abstract

Both Bankia and CaixaBank are among the financial institutions with the largest market share in Spain. During the years 2020 and 2021 they have been the protagonists of one of the biggest movements in the banking sector. The idea of this movement provokes many opinions among different users, making it an attractive object for study and research.

Therefore, in this paper we have proceeded to explain the history of the banking sector and the movements that preceded this movement for the reader to understand the reasons for this merger. After a brief contextualization, we have tried to carry out an empirical analysis using the different techniques provided by Big Data, with the aim of carrying out a sentiment analysis with the Twitter social network, and how the share price of the financial institution CaixaBank has evolved. To carry out the empirical analysis it was necessary to use the programming languages R and Python. After carrying out the analysis, the opinions analyzed on Twitter support to a certain extent the movement proposed by the entities, as well as being reflected positively in CaixaBank's share price.

Key words

CaixaBank, Bankia, merger, sentiment analysis, Twitter, Python, R, share price, *Business Analytics, Machine learning.*

ÍNDICE DE ILUSTRACIONES

<i>Ilustración 1: Tasa de crecimiento real del PIB europeo entre 1997 y 2005.....</i>	<i>11</i>
<i>Ilustración 2: Evolución de los tipos de interés reales entre 1996 y 2011 (porcentaje).....</i>	<i>12</i>
<i>Ilustración 3: Reformas en el sistema financiero español entre 2008 y 2012.....</i>	<i>18</i>
<i>Ilustración 4: Extracto de código Python para la extracción de Tweets.....</i>	<i>30</i>
<i>Ilustración 5: Extracto de código limpieza “dataset”.....</i>	<i>31</i>
<i>Ilustración 6: Nube de palabras respecto a CaixaBank y Bankia.....</i>	<i>34</i>
<i>Ilustración 7: Evolución precio de cotización CaixaBank comprendido entre septiembre de 2020 y abril de 2021.....</i>	<i>36</i>
<i>Ilustración 8: Evolución del volumen de cotización CaixaBank entre septiembre de 2020 y abril de 2021.....</i>	<i>37</i>
<i>Ilustración 9: Topics más importantes de análisis de sentimientos respecto a CaixaBank y Bankia.....</i>	<i>38</i>
<i>Ilustración 10: Frecuencia de los tipos de sentimientos respecto a CaixaBank y Bankia.....</i>	<i>39</i>
<i>Ilustración 11: Comparación entre sentimientos positivos o negativos respecto a CaixaBank y Bankia.....</i>	<i>40</i>
<i>Ilustración 12: Evolución de los sentimientos a lo largo del periodo determinado sobre CaixaBank y Bankia.....</i>	<i>41</i>
<i>Ilustración 13: Nube de palabras con sentimientos respecto a CaixaBank y Bankia.....</i>	<i>42</i>

ÍNDICE DE ACRÓNIMOS

API	<i>Application programming interface</i>
BCE	<i>Banco Central Europeo</i>
BE	<i>Banco de España</i>
CCOO	<i>Confederación Sindical de Comisiones Obreras</i>
CNMC	<i>Comisión Nacional de los Mercados y la Competencia</i>
ERE	<i>Expediente de Regulación de Empleo</i>
FMI	<i>Fondo Monetario Internacional</i>
FROB	<i>Fondo de reestructuración ordenada bancaria</i>
MEDE	<i>Mecanismo Europeo de Estabilidad</i>
PIB	<i>Producto interior bruto</i>
UE	<i>Unión Europea</i>
UGT	<i>Unión general de Trabajadores</i>

Índice de contenidos

1. Introducción.....	7
1.1 Objetivo.....	7
1.2 Justificación del tema objeto de estudio	7
1.3 Metodología y estructura	8
2. Análisis del sector bancario CaixaBank y Bankia.....	10
2.1. El sector bancario en España a lo largo del tiempo	10
2.1.1 Sector bancario tras la crisis financiera de 2008	13
2.1.2 Rescates bancarios españoles	15
2.1.3 Crisis del COVID-19 y Nueva oleada de fusiones	19
2.2 Fusión CaixaBank y Bankia.....	21
3. Análisis de sentimientos y Twitter como red social.....	23
3.1 Twitter como medio de comunicación e influencia	23
3.2 Análisis de sentimientos	24
3.2.1 Problemática con Twitter	26
4. Metodología	28
4.1 Descarga de los datos necesarios.....	28
4.1.1 Datos financieros CaixaBank.....	28
4.1.2 Datos de Twitter.....	29
4.2 Tratamiento de los datos	31
4.2.1 Limpieza de datos	31
4.3 Análisis de los resultados obtenidos	32
4.4 Análisis de sentimientos	32
4.4.1 Análisis sintáctico y léxico.....	32
4.4.2 Procesamiento semántico	33
5. Resultados del análisis y visión del precio de cotización	36
5.1 Análisis de la cotización de CaixaBank.....	36
5.2 Análisis de sentimientos acerca de la fusión entre CaixaBank y Bankia	38
6. Conclusiones	43
6.1 Limitaciones surgidas durante el desarrollo del estudio.	44
6.1.1 Limitación de datos.....	44
6.1.2 Dificultad de análisis de sentimientos.....	44
6.1.3 Momento de realización del estudio	44
7. Bibliografía	46
8. Anexos	50
8.1 Anexo 1: Código de análisis de sentimientos	50
8.2 Anexo 2: Código de descarga de Tweets mediante python.....	60

1. Introducción

1.1 Objetivo

El objetivo que se persigue en el presente trabajo de investigación tiene dos enfoques claros, pero conjuntos. Por un lado, el estudio del precio de la acción de CaixaBank y como este ha podido evolucionar según las opiniones surgidas en Twitter. Por otro lado, un análisis de sentimientos mediante metodologías de Data Analytics, text mining, mediante la realización de un análisis de sentimientos de Twitter para ver como los clientes de ambas entidades piensan sobre la fusión, como pueden fidelizarlos, como pueden prepararse para el futuro y si confían en el movimiento.

1.2 Justificación del tema objeto de estudio

Durante 2020 y 2021 la COVID-19¹ ha provocado una de las situaciones más críticas en el mundo. Esto ha provocado una serie de fenómenos en el ámbito de las finanzas, acelerando procesos que antes se miraban a largo plazo. Tras esta crisis, los grandes bancos españoles temen que la tasa de morosidad crezca de una manera descontrolada. Por ello, han unido fuerzas y se prevé una nueva ola de fusiones. Como resultado de ello son las fusiones de CaixaBank-Bankia, objeto de estudio, BBVA-Sabadell, recientemente paralizada y rota, Unicaja-LiberBank y otros grandes proyectos de fusiones.

Asimismo, la popularidad de las redes sociales como medios de comunicación e influyentes para la sociedad ha crecido en magnitudes inimaginables. Estos, están creciendo a una media de 6,7% cada año. En 2019, su uso se incrementó en un 1,4% con respecto al año anterior, en un 6% desde 2018, y hasta un 15% en 2016 (Cooper, 2020). Cabe destacar que, las redes sociales que cuentan con más usuarios son Twitter y Facebook (Galeano, 2021). La manera en la que los distintos usuarios expresan sus opiniones está sujeta a unos sentimientos, algunos más arraigados y otros menos. Es inevitable que un titular de un periódico o la opinión de una persona no este sujeta a una serie de sentimientos.

En la época en la que vivimos, la globalización es un fenómeno cada vez mayor y la cantidad de información que circula por todos ellos es infinitamente superior que hace

¹ Es la enfermedad causada por el nuevo coronavirus conocido como SARS-CoV-2.

unos años. Esta situación provoca que la información llegue a un mayor número de personas y que pueda llegar a manipular sus sentimientos acerca de algunos temas.

Los primeros impactos provocados por las opiniones son reflejados en el precio de cotización de las compañías. Sin embargo, la existencia de las redes sociales permite a gran parte de usuarios mostrar sus sentimientos hacia los distintos movimientos que realiza la entidad en su día a día. Esto provoca que la utilización de técnicas de Big Data permita la evaluación de riesgos y prevención de fraudes. Además, mediante un análisis de sentimientos ofrecen a las entidades la posibilidad de anticiparse a lo que sus usuarios demandan y como perciben ciertos movimientos.

Por todo ello, se ha decidido como objeto de estudio la fusión entre las entidades financieras españolas CaixaBank y Bankia. A lo largo de este estudio se pondrán en práctica las técnicas de *Machine learning* y *Business analytics*, las cuáles son consideradas como herramientas fundamentales para los estudios futuros y presentes. El objetivo de dicho estudio será el de determinar si la operación se percibe de una manera positiva o negativa y como podría afectar al precio de cotización de la entidad futura y presente, CaixaBank.

1.3 Metodología y estructura

La metodología empleada para este estudio se divide en dos partes muy claras. En primer lugar, se ha empleado una metodología general, de donde se ha podido obtener una serie de recursos, como son el BCE y BE. Además de ello, se ha utilizado todo tipo de documentos y artículos relevantes sobre el tema relacionado, así como *Google Scholar*, que se trata de un buscador de contenido puramente académico y de investigación. Por otro lado, la metodología *Business analytics* y *Machine learning*, la cual está enfocada a resolver la parte empírica del estudio. Se encontrará desarrollada de una forma más detallada en el apartado 4.

La estructura seguida en el siguiente estudio se divide en cinco partes, incluyendo esta breve introducción. En primer lugar, se presenta una contextualización sobre el sector bancario español a lo largo de la historia, indagando sobre su historia previa a la crisis financiera de 2008, su etapa de la reestructuración y sus rescates, además de como ha afectado la crisis de la COVID-19 y como se ve su futuro acompañado de un breve resumen acerca de la fusión entre ambas entidades. En segundo lugar, se justifica la elección de Twitter como red social y las limitaciones que ello implica. Posteriormente,

en el capítulo 4, se ha explicado la metodología empleada para el estudio empírico, en la cuál destaca como se procedió a: la descarga tanto de los datos financieros como la de la base de datos de twitter; como se ha llevado a cabo el tratamiento de los datos, es decir, la limpieza del *dataset*; la introducción del punto cinco y la explicación del análisis de sentimientos. Finalmente, en el último capítulo empírico, 5, se exponen los resultados obtenidos del estudio y las reflexiones que obtenemos de ellos. Para terminar, se exponen las conclusiones del estudio y una serie de limitaciones surgidas durante su desarrollo.

2. Análisis del sector bancario CaixaBank y Bankia

En este apartado vamos a tratar de contextualizar las entidades bancarias en el ámbito español. Para comenzar, se realiza un pequeño análisis de la importancia del sector bancario en nuestro territorio, tratando de explicar la banca anterior a la crisis de 2008. Posteriormente, se desarrolla un breve análisis de los rescates que surgieron derivados de esta crisis (teniendo en cuenta que Bankia fue rescatado por el gobierno) para finalmente acabar con la crisis de la COVID-19 y como se pronostica la posterior recuperación en este sector. Asimismo, vamos a analizar lo que ha sido toda la trayectoria de la fusión entre CaixaBank y Bankia, desde su anuncio hasta su integración.

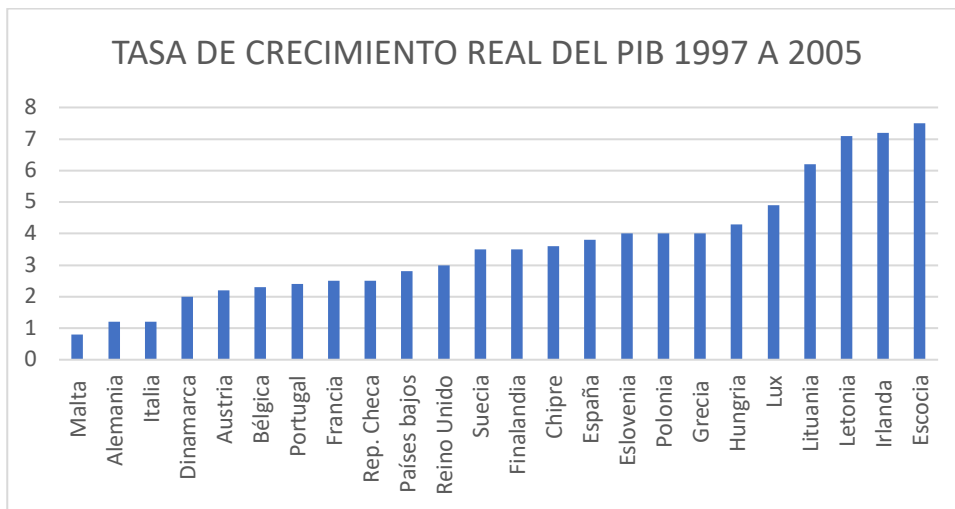
2.1. El sector bancario en España a lo largo del tiempo

El sector bancario, que actualmente conocemos en España, ha sufrido grandes transformaciones a lo largo de los años. La primera transformación se inició en los años sesenta, la cuál se vio acelerada tras la apertura de los mercados al exterior, es decir, el libre comercio entre países, provocando la entrada de otros grandes bancos extranjeros aumentando así el número de activos circulantes en la economía. Todo ello, llevó a la especialización de las distintas entidades en cada una de las áreas financieras como son el área de fusiones y adquisiciones, gestión de carteras, gestión de patrimonios...

A pesar de que el sector financiero poseía un papel clave en las economías europeas, en España el peso era relativamente bajo, representando, en 2005, el 4,64% del valor añadido bruto y el 1,96% del empleo de la economía. Con todo ello, el crecimiento del sector bancario fue mucho superior al crecimiento del sector privado al igual que en varios países del resto de la unión europea. Además, a mediados de los años noventa se produjo un aumento de los activos financieros, multiplicándose por 1,6. Sin embargo, este crecimiento había que tenerlo en cuenta a términos absolutos con el crecimiento de la economía español, ya que crecía a un ritmo muy elevado provocando que los activos se multiplicasen por 2,4, puesto que el PIB español crecía por encima del resto de Europa (Maudos, 2008).

A continuación, se muestra una ilustración sobre el crecimiento del PIB (porcentaje) europeo ofreciendo una visión aguda de lo expresado en el párrafo anterior.

Ilustración 1: Tasa de crecimiento real del PIB europeo entre 1997 y 2005



Fuente: Fundación BBVA, *el sector bancario español en el contexto internacional* (2008).

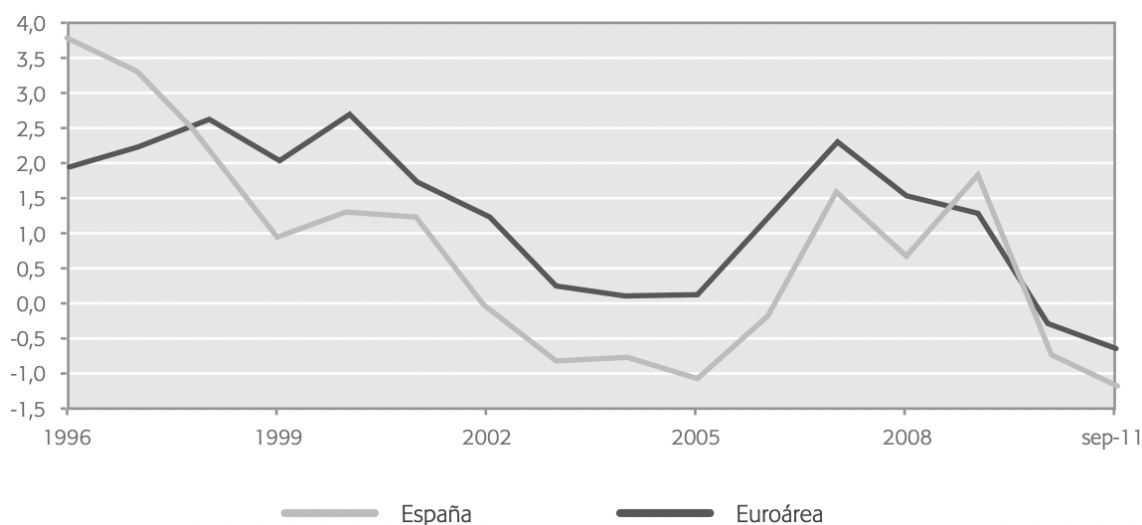
Observando la ilustración, los bancos ocupaban una posición central en la estructura financiera de la economía española, donde los préstamos bancarios suponían un 141% del PIB, mientras que el mercado de renta fija y de renta variable suponían el 82% y el 85% respectivamente. Estos activos financieros en manos de la banca pasaron de suponer un 167% del PIB durante 1997 a representar el 258% del PIB en 2006. Con todo ello, el peso del sector bancario en España es relativamente reducido de acuerdo con los estándares europeos, ya que en la UE-25 la ratio de los activos totales/PIB era del 323% en 2006 (Guevara, 2008).

El crecimiento de la banca española superaba al crecimiento de los demás países europeos hasta 2007, donde la bancarización de la economía española se intensificaba. Aunque, la crisis iniciada en EE. UU. durante 2008 afectó a este sector, no solo en España, sino en toda Europa. Esta gran crisis afectó a España de una manera muy agresiva debido a una serie de factores como son:

- Abundancia de liquidez debido a los reducidos tipos de interés:

La entrada del euro en España produjo que los tipos de interés y las tasas de inflación cayesen, lo que provocó que los tipos de interés reales se situasen en valores negativos en el año 2006. Esto provocó que el nivel de endeudamiento de toda la población ascendiese puesto que resultaba muy atractivo. Para mostrarlo de una manera más clara. A continuación, se va a ilustrar el movimiento de los tipos de interés reales en la zona euro desde 1998 y 2011.

Ilustración 2: Evolución de los tipos de interés reales entre 1996 y 2011 (porcentaje)



Fuente: Maudos, J. El sector bancario español en el contexto internacional: el impacto de la crisis (2011)

- Creciente endeudamiento:

Con el escenario ideal de endeudamiento anteriormente descrito, la economía española era incapaz de soportar esas inversiones sino era aumentando la deuda privada, lo que la posición financiera neta negativa paso de representar el 32% del PIB a el 89%, muy por encima del 11% de las demás economías europeas. Para hacer frente a este aumento de deuda, la economía española se apoyo en los grandes bancos, los cuales tuvieron que acudir a los mercados mayoristas y aumentar su deuda externa.

- Crecimiento del crédito bancario, por encima del PIB:

El crédito al sector privado en España creció a niveles muy elevados y por encima de la media de los demás países de la unión europea. Mientras en los países de Europa el crecimiento se encontraba en torno al 7,4% en España crecía a un 17,8% anual, solo superado por Irlanda y Grecia, y muy diferente a Alemania que crecía a un 1,1%.

- Riesgo en el sector de la construcción y promoción:

La gran apuesta fue el sector de la construcción y promoción, así como la vivienda donde el 60,8% de los créditos de las entidades de crédito iban destinadas a esta actividad.

- Gran número de oficinas conlleva elevado empleo.

Con el crecimiento de los créditos, las cajas de ahorros decidieron aumentar las sucursales, lo que se tradujo en una aparición de nuevas sucursales (alrededor del 30%). Este hecho provocó una gran cantidad de puestos de empleo que, posteriormente, se vería reducido por el cierre causado por la crisis. Sin embargo, redujeron las sucursales en un 1 %.

2.1.1 Sector bancario tras la crisis financiera de 2008

La banca española se encontraba en una difícil situación a causa de la crisis por lo que, el Estado, adoptó una serie de medidas a corto-medio plazo para combatir las graves consecuencias. Esta serie de medidas seguían las directrices impuestas por la UE y la cumbre del G-20 (18 de noviembre de 2008).

El Estado español promovió un gran gasto público para apoyar a las entidades financieras. Este movimiento se realizó con la finalidad de apoyar la oferta de créditos a la actividad empresarial y a los ahorradores mediante el Plan E², constituido en noviembre de 2008. Además, este plan aportó un conjunto de soluciones de las cuales, algunas no llegaron a ser ejecutadas.

Esta serie de medidas situaron a España en una posición favorable en términos de rentabilidad, eficiencia, volúmenes de provisiones y niveles de capital en comparación con el entorno europeo. Sin embargo, la actividad empresarial se vio reducida por este fenómeno poniendo de manifiesto la incapacidad y la viabilidad futura de las entidades de crédito, debido a las restricciones de liquidez, así como las dificultades de entrar a financiación del mercado por el deterioro de ciertos activos. Todos estos problemas, llevaron a la mayoría de las entidades de crédito al cierre de oficinas y sucursales bancarias, e impulsó las reestructuraciones bancarias dejando la ventana de las fusiones abiertas para todas ellas. Esta medida fue impulsada por el gobierno mediante el establecimiento del *Real Decreto-Ley 9/2009*, de 26 de junio, persiguiendo un claro objetivo de mantener la confianza en estos organismos.

Más tarde, se decretaría el *Real Decreto-Ley 11/2010*, a 9 de julio, el cual afectaba a las cajas de ahorro. Estas, desempeñaban un papel fundamental en nuestro sistema

² El Plan Español del Estímulo de la Economía y el Empleo es una medida creada por el Gobierno para luchar contra el desempleo y recuperar puestos de trabajo e intentar recuperar así la economía a nivel mundial tras la crisis vivida recientemente.

financiero puesto que fomentaba el ahorro, movilizaba recursos y permitía acceso a crédito a todas las familias españolas. Con este decreto se perseguía que estas cajas se profesionalizasen más y se le animaba a la captación de capital externo debido al momento que atravesaba el país ya que su funcionamiento no atraía inversores por su escasa liquidez y pocos derechos de votos.

Así que estos dos factores claves (complejidad jurídica y dificultad de obtención de recursos) dieron lugar a una visión muy desfavorable para este tipo de organismos, y que, en un primer momento, causó movimientos societarios con el objetivo de tener una mayor liquidez y mejorar su eficiencia, para posteriormente mejorar su posición competitiva respecto a las entidades de crédito y mejorar su imagen en el mercado.

Más tarde, en el transcurso del año 2011 el gobierno español aprobó un nuevo decreto, *Real Decreto-Ley 2/2011*, de 18 de febrero, debido a la creciente incertidumbre causada por la nueva crisis de la deuda soberana de la Eurozona. Además, este buscaba asegurar la estabilidad de las entidades, facilitar el acceso a la financiación, fomentar las reestructuraciones de todas las entidades y canalizar el crédito hacia la economía real. Esto se conseguiría ya que el decreto proponía el refuerzo del nivel de solvencia de los distintos bancos españoles y, además se aceleraba el proceso de reestructuraciones.

Sin embargo, todas estas medidas fueron inútiles para mejorar nuestro sistema bancario y en cierto modo recuperarlo lo antes posible. Por ello, se optó por la elaboración de otro nuevo *Real Decreto-Ley 2/2012*, el 3 de febrero, para llevar a cabo un saneamiento mayor de este sector financiero español, donde se empezó reconociendo que las medidas impuestas anteriormente no causaron ningún efecto en su recuperación y que, tras cuatro años de haberse iniciado la crisis, el sector financiero seguía viéndose afectado. También, las restricciones crediticias y los activos tóxicos de las entidades de crédito seguían muy presentes sobre el sector inmobiliario debido a la concesión de hipotecas en los años anteriores a la crisis. Todo ello, dificultó la mejora de su eficiencia y su competitividad.

Así pues, la normativa contenía una estrategia integral de reforma financiera que pretendió profundizar en la recuperación de la credibilidad, confianza y la fortaleza del sistema financiero español (Cuervo, 2012). Esta reforma pretendía sanear los balances de las distintas entidades, la creación de incentivos para propiciar el ajuste adecuado y eficiente de los excesos de la capacidad instalada; fortalecer la gobernanza de las entidades que resultaran tras los procesos de reestructuración, y garantizar que los costes

que conllevaran este saneamiento no recayesen en el sector público y recayesen en el sector bancario.

Todas estas medidas anteriormente mencionadas y, a las que se le unieron modificaciones en el FROB fueron utilizadas para lograr el objetivo de una recuperación bancaria rápida y eficaz, que tras la crisis de 2008 no evitaron que las autoridades españolas solicitasen a Europa ayudas para llevar a cabo un rescate bancario, el cual se desarrollará a continuación.

2.1.2 Rescates bancarios españoles

Al inicio del año 2012, España se encontraba en un momento de grave tensión en sus mercados, ya que la rentabilidad exigida a la deuda pública española cada vez era mayor, esto se tradujo en un incremento de la prima de riesgo del país. Por todo ello, el Consejo de ministros de nuestro país decidió que el ministro de economía debía realizar un profundo análisis sobre la resistencia del sector bancario provocado por el continuo deterioro de la economía. Para el desarrollo del estudio, el ministerio de economía decidió apoyarse en *Roland Berger* y *Oliver Wyman (2012a)*, ambos consultores independientes, para la realización de una estimación de las necesidades líquidas del sector bancario con el fin de hacer frente a cualquier situación de incertidumbre que pudiera surgir.

Los resultados originados del profundo estudio se publicaron a finales del segundo trimestre del año donde afectaban a más de 14 entidades bancarias, cabe resaltar Caixabank&BancaCívica (objeto de estudio), y que afectó también a grandes bancos españoles como son el Banco Santander, BBVA y Sabadell entre otros. Las medidas respondían a dos premisas macroeconómicas claras obtenidas de las conclusiones del Programa de Evaluación del Sector Financiero (Calvo y Martín, 2014), donde se pone de manifiesto que los graves problemas se encontraban en aquellas entidades que estaban controladas por el FROB o que ya habían recibido alguna ayuda pública previa. En un escenario de crecimiento esperado del 9% las necesidades se situaron en torno a los 16.000 y 26.000 millones de euros, pero por el contrario en un contexto desfavorable de crecimiento del 6% las necesidades se encontrarían entre los 51.000 y 62.000 millones de euros.

Tras el profundo análisis el gobierno español decidió solicitar asistencia financiera externa para llevar a cabo el proceso de reestructuraciones planteado y la recapitalización del tejido bancario español. Esta ayuda fue otorgada tras la reunión de los países del

Eurogrupo, en el cual se estableció que sería el Mecanismo Europeo de Estabilidad (MEDE) el encargado de prestar esa asistencia financiera solicitada.

Más adelante, se llevaron a cabo una serie de negociaciones que acabaron con la aprobación de un memorándum de entendimiento, el cual establecía un claro camino a seguir con una serie de fechas establecidas. Los grandes objetivos marcados en este memorándum consistían en incrementar la flexibilidad de las entidades bancarias, mediante una mayor transparencia de sus balances comerciales, con el objetivo de reducir la inversión en el sector inmobiliario y diversificar más su cartera; restablecer la financiación basándose en el mercado; aumentar la independencia de los bancos españoles con respecto al BE para aumentar la liquidez; y, finalmente, perfeccionar cada uno de los mecanismos para mitigar futuras crisis financieras y tener un mayor control de riesgos.

Por otro lado, uno de los aspectos clave pactados fue la continua revisión de las debilidades y vulnerabilidades del tejido bancario español donde se fijaron tres claros objetivos:

- Identificar aquellos bancos con necesidades de capital mediante un análisis genérico de sus activos en propiedad y un análisis más estricto para comprobar la resistencia de cada uno de los bancos (análisis bottom-up) en un entorno macroeconómico adverso.
- Reestructuraciones de aquellas entidades más vulnerables, entendidas como aquellas que tienen déficits de capital y han sido detectadas mediante las distintas pruebas de resistencia establecidas.
- Segregar los activos dañados de aquellos bancos que necesiten apoyo de la administración pública, para posteriormente, desviarlos a una sociedad externa de gestión de dichos activos.

Tras establecer las premisas del memorándum de entendimiento, y después de haber realizado un primer análisis top-down, se llevó a cabo el análisis bottom-up, el cual era realizado por Oliver Wyman(2012b) al que se le sumaban otro número de expertos independientes, el Banco de España y el ministerio español de Economía. Todo ello era supervisado a su vez por la Comisión Europea, el Banco Central Europeo, la autoridad Bancaria Europea y el Fondo Monetario Internacional (FMI).

A su vez, el estudio de las entidades del sector bancario español se realizó de una manera más detallada e intensa. Se incorporaron métodos de valoración de las carteras de crédito del sector privado y de los activos inmobiliarios, a lo que se sumaba un estudio sobre la capacidad que tendrían para soportar pérdidas en los próximos tres años del entorno tan adverso sobre las 14 entidades que se encontraban como objeto de estudio en aquel momento.

Los resultados de estos estudios fueron positivos para la banca española, donde el 62% de las entidades de crédito españolas podrían soportar una situación macroeconómica desfavorable futura y, que se estimaron unas necesidades de 53.745 millones de euros en ayudas, en los que se incluían los gastos de los procesos de reestructuraciones de las distintas entidades.

Asimismo, se realizó una clasificación en cuatro grupos de las entidades bancarias. En el grupo 0 se encontraban las entidades que no precisaban ninguna ayuda económica, aquí se situaba la entidad CaixaBank. El grupo 1 lo comprendían los bancos rescatados por el FROB, en el cual se localizaba la entidad madrileña Bankia. El grupo 2 presentaba compañías con graves necesidades de capital. Finalmente, el grupo 3 no lo conformo ningún banco.

Para cumplir con el procedimiento del memorándum de entendimiento, el gobierno español estableció un nuevo Real-Decreto Ley 9/2012, a 14 de noviembre, con el fin de garantizar su principal objetivo de reestructuración de las entidades de crédito y convertirlas en herramientas más solventes y viables. De modo que, si dichas entidades no cumplían con su objetivo, el gobierno actuaría de manera inmediata. A su vez, se estableció el nuevo régimen jurídico para el FROB.

Al mismo tiempo, este Real-Decreto estableció la creación de las sociedades de gestión de activos, las cuales tenían el objetivo de acumular aquellos activos tóxicos, (resultados previamente a la crisis) y que dañasen el balance de las entidades dando lugar así a una gestión más sencilla y una transacción rápida y efectiva. Para garantizar con una mayor fuerza este hecho, el gobierno decretó otro nuevo *Real-Decreto Ley 1559/2012*, el 15 de noviembre, para que aquellas entidades que habían recibido ayudas públicas, entre las que se encuentra Bankia, con el fin de que este tipo de entidades optasen a sacar sus máximos beneficios y pudiesen devolver el préstamo efectuado por el estado.

A todo ello se le suma una serie de medidas para fortalecer el marco regulador y la supervisión del sector bancario español por parte de Estado, las cuales habían sido impuestas en el memorándum de entendimiento.

A continuación, se va a mostrar un cuadro con las medidas empleadas entre los años 2008 y 2012:

Ilustración 3: Reformas en el sistema financiero español entre 2008 y 2012

Plan E (noviembre 2008)	<ul style="list-style-type: none"> - Creación del Fondo de Adquisición de Activos Financieros. - Aumento de los importes máximos garantizados por los FGD: 100.000 € por depositante. - Avales del Estado para las nuevas emisiones de deuda de las entidades de crédito. - Adquisición estatal de títulos para reforzar los recursos propios de las entidades (no se llegó a emplear).
Real Decreto-Ley 9/2009, de 26 de junio	<ul style="list-style-type: none"> - Reestructuración ordenada del sistema bancario. Objetivo: mantener la confianza en el sistema financiero nacional e incrementar su fortaleza y solvencia. - El modelo propuesto se articuló en torno a: <ul style="list-style-type: none"> ➢ Los tres FGD. ➢ Una nueva institución: el FROB. Funciones: la gestión de los procesos de reestructuración y el reforzamiento de los recursos propios en determinados procesos de integración.
Real Decreto-Ley 11/2010, de 9 de julio	<ul style="list-style-type: none"> - Reforma del régimen jurídico de las cajas de ahorros: mayor profesionalización de sus órganos de gobierno. - Refuerzo de su capacidad para obtener recursos propios de máxima calidad, con el objetivo de favorecer su posición competitiva y su credibilidad en los mercados.
Real Decreto-Ley 2/2011, de 18 de febrero	<ul style="list-style-type: none"> - Refuerzo del nivel de solvencia de todas las entidades de crédito, con la finalidad de salvaguardar la estabilidad y eficiencia del sistema financiero. - Fase final de los procesos de reestructuración de entidades.
Real Decreto-Ley 2/2012, de 3 de febrero	<ul style="list-style-type: none"> - Saneamiento de los balances de las entidades de crédito de activos relacionados con el sector inmobiliario, para que puedan canalizar el ahorro hacia proyectos de inversión eficientes. - Eliminación de los excesos de capacidad instalada. - Fortalecimiento de la gobernanza de las entidades resultantes de los procesos de integración, especialmente de las cajas de ahorros. - Garantizar que los costes del saneamiento del sistema financiero no recaigan sobre el sector público.
Real Decreto-Ley 18/2012, de 11 de mayo	<ul style="list-style-type: none"> - Nuevo refuerzo de las exigencias de provisiones para los activos vinculados al sector inmobiliario. - Nueva modificación de la normativa del FROB. - Previsión de la creación de Sociedades para la Gestión de Activos ("bancos malos").

Fuente: Calvo Bernardino, Antonio y De Vidales Carrasco, Irene Martín y (2014), "Crisis y cambios estructurales en el sector bancario español: Una comparación con otros sistemas financieros." Estudios de Economía Aplicada".

Tras el fuerte compromiso del gobierno español por restaurar el sector con las medidas impuestas, a finales del 2012 recibió la cantidad 39.468 millones de euros del

MEDE para el FROB. Más adelante, en el segundo mes de 2013 se recibieron bonos del MEDE al FROB para el rescate de los grupos bancarios del grupo 2.

Todas estas medidas provocaron el cierre de 38 entidades de crédito, donde destacan las transformaciones de las cajas de ahorro y los continuos procesos de consolidación y redimensionamiento por ellas vivido (Calvo, De Vidales y Martín, 2014).

2.1.3 Crisis del COVID-19 y Nueva oleada de fusiones

El 31 de diciembre de 2019 surgió un acontecimiento que nos cambiaría la vida, se detectó el primer caso de coronavirus del mundo en Wuhan, China. Nadie se imaginaba que este fenómeno modificaría nuestra forma de vivir, y con ello de trabajar. Dicho fenómeno ha provocado la mayor recesión económica vista tras la crisis de 2008 debido a que las medidas que se tuvieron que llevar a cabo para intentar controlar el virus paralizaron el mundo laboral en la gran mayoría de los sectores.

Este hecho ha provocado una incertidumbre del sector bancario como ocurrió en 2008. Surge la necesidad de saber si las entidades financieras serían solventes a la hora de afrontar esta crisis. De cualquier modo, la banca española lleva preparándose para una situación como esta mucho tiempo, gracias a la experiencia de la anterior crisis económica. Asimismo, los poderes públicos utilizan los bancos como una herramienta clave para canalizar sus ayudas (Esteban, 2020).

Por otro lado, no todo va a ser positivo para este sector, la primera pregunta que todo el mundo se plantea es la rentabilidad que la banca va a tener tras este impacto y con los continuos bajos tipos de interés, lastre que existía antes de la crisis de la COVID-19.

Por ello, la banca tendrá que reponerse del efecto de la rentabilidad y de la solvencia que la crisis esta suponiendo. El mayor indicador son los índices de morosidad, ya que continúan y seguirán en crecimiento en un futuro cercano, y por mucho que las ayudas públicas puedan contribuir para que este efecto no sea tan relevante, los expertos dicen que si será importante.

Se marcan así una serie de claves para llevar a cabo la recuperación de la banca en este tiempo, pero con un menos apoyo y obligación por parte del estado. Vamos a marcar tres claves:

- **Digitalización:** La crisis de la COVID-19 ha acelerado el proceso de digitalización, el cual llevaba realizándose poco a poco. Este fenómeno se ha

producido no solo de los bancos, si no también de todas las personas físicas, ya que se han tenido que adaptar a trabajar desde sus casas. Este campo supone una gran ocasión para consolidar o perfeccionar aquellos puntos de mejora, llevando a cabo nuevas inversiones para sacar, mejorar y ofrecer nuevos servicios que se adapten a todo tipo de situaciones extraordinarias. La digitalización ha impulsado a los bancos a buscar sinergias mediante sus fusiones, con la posibilidad de aumentar sus tamaños y convertirse de una manera más digital llegando a un mayor número de personas. Además, ayuda a aquellos bancos con problemas de rentabilidad y de solvencia ante la morosidad sobreviviendo así a este periodo. Finalmente, este factor puede situar a España como una de las bancas digitales más importantes de Europa.

- **Sostenibilidad:** La crisis sanitaria ha puesto en evidencia que los bancos deben optar por una inversión en activos más sostenibles, es decir, en el desarrollo de fármacos de la salud e investigación sanitaria. Asimismo, pone evidencia que deben abrir su abanico de inversión a activos que protejan los riesgos medioambientales y, en definitiva, una apuesta por modelos de negocios sostenibles.
- **Inversión:** Tras esta crisis se abren nuevos frentes de inversión. Ciertamente es que siempre quedarán sectores que continuarán siendo competitivos y que seguirán creciendo, pero con este nuevo estilo de vida, hay que adaptarse a los cambios y ver nuevas oportunidades en el ámbito de la inversión. Muchas inversiones se están dirigiendo a la llamada “*Nueva Economía después del COVID-19*”, entre las cuales se integran empresas tecnológicas con economía digital, así como tratamientos de prevención y/o desarrollo de fármacos.

Todo ello, ha llevado al sector bancario español a una nueva ola de fusiones, un plan que llevaba persiguiendo Luis de Guindos desde hace mucho tiempo tratando de crear gigantes bancos españoles. Esta oleada está presidida por CaixaBank y Bankia, los cuales han realizado el primer movimiento impulsando así a sus competidores. Tras el fallido intento de fusión entre BBVA y Sabadell, entró en juego la fusión de Unicaja y LiberBank. Además, Santander podría tener un as bajo la manga, donde todo hace indicar que podría llevar a cabo una fusión con Bankinter. Todo ello empuja a BBVA a circular

en este mundo para no quedar relegado a sus competidores. Otros grandes bancos que suenan son Kutxabank, Abanca e incluso Ibercaja.

Asimismo, los supervisores bancarios tratan de incentivar a los bancos europeos a que lleven fusiones dentro de la Unión Europea, no solo a nivel nacional, aunque resulte una tarea difícil de conseguir.

2.2 Fusión CaixaBank y Bankia

El pasado 17 de septiembre de 2020, se acordó la fusión entre Bankia y CaixaBank donde la entidad absorbente sería CaixaBank y la absorbida Bankia, pero la principal incógnita que se trataba en las negociaciones no era esta, sino el precio de compra y la ecuación de canje.

Como resultado de esta negociación, la ecuación de canje acabó cerrándose en 0.6845 acciones nuevas de CaixaBank por cada acción de la compañía Bankia, suponiendo que el 74,2 % de la nueva entidad pertenece a la antigua CaixaBank y el 25,8% a Bankia. Por otro lado, la fundación Caixa será el principal accionista manteniendo su participación del 30% mientras que la participación del FROB llegará a alcanzar el 16,1%.

Este movimiento va a suponer un ahorro de costes de más de 770 millones de euros, que a su vez vera aumentado el beneficio por acción en un 28% con respecto a las estimaciones para 2022. Asimismo, se observa un paso hacia delante del gobierno hacia la privatización de las entidades bancarias, ya que poseía el 61,8% de Bankia mediante el FROB (Hernández, 2020).

Desde hace mucho tiempo, el presidente de la firma CaixaBank, Isidro Fainé, persiguió este movimiento tan relevante para la compañía. Se trata de un movimiento estratégico enorme, ya que supone que la nueva entidad se convierta en la entidad con mayor cuota de mercado en España contando con un 25% del total. Otro de los aspectos más relevantes es la gran cantidad de activos que esta nueva compañía manejará, convirtiéndose así en la que más volumen de activos posee, valorados en más de 664.000 millones de euros.

La nueva compañía tendrá su sede social en la capital de la Comunidad Valenciana y el consejo de administración constará de 15 personas, la cual una de ellas será un representante del FROB por su participación en ella.

El 26 de marzo de 2021, culminó la ansiada fusión entre CaixaBank y Bankia. Se figuraba que el acontecimiento supondría el crecimiento de toda la sociedad y de la compañía, pero CaixaBank tenía planeado realizar un ERE y llevar a cabo una reestructuración de sus sedes bancarias. Para realizar estos movimientos precisaban del consentimiento de la CNMC.

El 20 de abril del mismo año, tras haber transcurrido apenas un mes de la fusión, CaixaBank anunció el mayor ERE de la historia de la banca española, acumulando la salida de 8.291 empleados, el 18,67% de su plantilla total, además del cierre de 1.534 oficinas, lo que supone un 27% de su red. Este movimiento ha provocado la reacción de los sindicatos. El secretario general de CC.OO. de CaixaBank, Ricard Ruiz, explica que no desea que los mayores afectados sean aquellos con una edad mayor a 50 años, ya que pretende mantener profesionales con la experiencia que los más jóvenes aún precisan desarrollar.

Es por ello por lo que el responsable de la UGT, Victoriano Miraverte, explica que esta reestructuración es “inasumible” puesto que la mitad de los ERES son para personas con una edad inferior a 50 años, condición aplicada por el banco anteriormente mencionada (Ángeles, 2021).

CaixaBank justifica este movimiento como necesario, justificándose con cuestiones productivas y organizativas, puesto que se producen duplicidades tras el movimiento de la fusión entre dos bancos que competían en el mismo sector. Además, la previsibilidad de que los tipos de interés continúen negativos hasta 2025, reduce el margen financiero para la firma, así como la digitalización del mundo y de los clientes provoca que el número de oficinas físicas tengan que verse reducidas. Sin embargo, el banco continuará teniendo beneficios y no lo argumenta por las pérdidas que se podrían generar.

3. Análisis de sentimientos y Twitter como red social

Tras un breve análisis del sistema bancario español y haber explicado las principales causas y motivos que llevaron a CaixaBank y Bankia a realizar esta fusión, a continuación, vamos a explicar por qué Twitter es el medio seleccionado para la base de datos con el fin de examinar la fusión y como se va a desarrollar el trabajo.

3.1 Twitter como medio de comunicación e influencia

Desde hace más de 10 años han aparecido nuevos medios de comunicación que están revolucionando la manera de informar a la población, todo empezó con Facebook y continuó con el nacimiento de Twitter. Esto genera una importante controversia alrededor de los medios de comunicación tradicionales de si llegarán a adaptarse a estas nuevas tendencias o se quedarán obsoletos (Universidad Andrés Bello, 2020).

Hoy en día, las fuentes a través de las que la población recaba más información sobre artículos periodísticos son Facebook y Whatsapp (Digital News Report, 2019), en los cuales la confianza recibida por sus consumidores asciende hasta el 34%. En cambio, existen otras plataformas como Instagram, que son utilizadas para captar la atención de los consumidores de una manera rápida. Twitter sigue todavía con una cuota baja de usuarios comparada con estas grandes plataformas, pero en estos últimos años se está empleando para influenciar y llegar de una manera más directa. Un ejemplo de ello ha sido el caso de las elecciones de Estados Unidos.

El crecimiento de este tipo de medios de comunicación ha provocado que aquellos periodistas profesionales se conviertan en puros espectadores anónimos, ya que anteriormente, solo era válida la información oficial. Esto, ha provocado que los medios tradicionales se estén adaptando a estas nuevas circunstancias.

Esta situación se ha visto agravada por la crisis del coronavirus, provocando una rápida innovación en los medios comerciales y en las técnicas de mantenimiento de espectadores. Plataformas como Twitch y Youtube han provocado un mayor impacto en las personas, pasando de ser plataformas únicamente destinadas al entretenimiento a convertirse en medios de información surgiendo canales de tendencias política, histórica y financiera entre otras.

El auge de Twitter como uno de los nuevos medios de comunicación sociales más relevantes en el mundo, se debe a que las noticias se narran en tiempo real, además de

que permite a sus usuarios participar en este tipo de noticias dando sus opiniones con diferentes puntos de vistas. Esto da la opción de que los consumidores tengan un amplio abanico de opiniones sobre cada noticia, sin la existencia de una verdad absoluta. Esta red nos demuestra que todo depende del cristal con que se mire (agencias de comunicación, 2020).

Por otro lado, la simplicidad de esta red social lo hace más atractivo, ya que solo necesitas un *smartphone* con conexión a internet. Asimismo, te ordena las noticias según la repercusión que tengan en el mundo o en tu país en el momento que te conectas, evitando a los usuarios una serie de información irrelevante que no es de su interés. Es un medio que además influye en gran medida a la imagen que percibimos tanto de las marcas como de ciertos medios de comunicación que se han adaptado e involucrado en esta red social. Los bulos y rumores son numerosos y las marcas tienen que poseer un gran equipo para detectarlos consiguiendo así que su imagen no se vea perjudicada por un comentario incierto. Asimismo, cualquier comentario inoportuno por medio de ellos puede ser causa de provocación.

Finalmente, el propósito de Twitter fue nacer como una red de microblogging³ (Castro, 2020) encargada de intercambiar información de una manera más informal. Sin embargo, el paso del tiempo y el avance de las tecnologías provocaron que se fuese adaptando a la demanda de los cibernautas. No solo usuarios anónimos emplean esta red social, si no que otras entidades públicas con peso en la sociedad también hacen uso de esta. Un claro ejemplo son los propios partidos políticos que utilizan este medio para realizar sus campañas electorales y compararse con sus rivales más directos, tratando de llamar la atención de una parte superior de la población, pues entienden que es una importante fuente para llegar a grandes masas con una mayor eficiencia.

3.2 Análisis de sentimientos

Para entender mejor el trabajo que llevamos a cabo, hay que clarificar este término que puede parecer algo ambiguo y complicado de comprender.

³ es un servicio que permite a sus usuarios enviar mensajes de sólo texto, con una longitud máxima de 140 caracteres

El análisis de sentimientos (“Sentimental analysis”) aplicado en el mundo Web se puede entender como el tratamiento computacional de las opiniones, sentimientos y fenómenos subjetivos en los textos (Vallverdú, 2011). En otras palabras, un programa computacional es capaz de entender, interpretar, discernir y traducir el significado de los distintos textos digitales que se quieren analizar, haciéndolos operativos y pudiendo clasificarlos en objetivos y subjetivos y, dentro de subjetivos entre positivos o negativos.

Esta técnica se emplea cada vez más en los diferentes procesos de las compañías por las distintas aplicaciones que tiene, así como la utilidad de sus resultados. Es por ello, que es una herramienta muy utilizada en compañías de marketing digital, puesto que permite un seguimiento del uso de las redes sociales, y en general, de todos los medios. Además, es una manera muy rápida y eficaz en comparación con las técnicas más rudimentarias empleadas hasta ahora. Todo ello, ha llevado a esta herramienta a implementarse en diversas áreas del trabajo.

Por ejemplo, las empresas pueden llevar un seguimiento de lo que los usuarios opinan sobre sus servicios o productos de una manera más subjetiva, otorgándoles libertad, sin que sientan presión y sin saber que realmente sus opiniones están siendo tratadas directamente.

Otra posible aplicación es la que pueden llevar a cabo todo tipo de compañías e incluso instituciones públicas. Entre las que se encuentra la posibilidad de intuir la intención de voto de la población, lo que tienen que mejorar y lo que más critican sobre sus políticas.

Las técnicas utilizadas en este tipo de actuaciones se basan en algoritmos de técnicas de “*Machine learning*”, los cuales se caracterizan por administrar documentos mediante la librería “*tm*”, por el que recibe el nombre de corpus, que es el texto de los documentos en el que se basará nuestro análisis (Rodríguez, 2019). En el caso del análisis de sentimientos dichos corpus se clasifican de forma binaria entre positivos y negativos.

Aparte de los algoritmos que emplean una selección binaria, existe la posibilidad de emplear otro tipo de algoritmos que emplean una clasificación en forma de rango, es decir, entre 1 y 4. En este caso, el algoritmo se entrena por los diferentes corpus asignados, previamente, a cada rango y aplica una serie de tipología a los textos que analiza.

Además de las técnicas de machine learning, existe la posibilidad de realizar un análisis semántico de las opiniones que cuenta con tres fases distintas. La primera fase se caracteriza por la realización de un análisis léxico y sintáctico, donde se normalizan las palabras y se “*tokenizan*”, es decir, se divide la frase en las distintas palabras que la componen, al que posteriormente se le eliminan una serie de palabras con la función de R “*stopwords*”, las cuales son palabras de relleno dentro de cada una de las opiniones, se continua con una “*lemmatization*”, reducir las palabras a su forma base, y acaba con un “*stemming*”, que consiste en reducir cada uno de los términos a sus *stems*. Posteriormente, se procede a un análisis sintáctico que mediante el “*parsing*” se determina la estructura de las frases y mediante el “*parts of speech tagging*” se etiqueta a cada palabra con su categoría gramatical. La segunda fase consiste en un procesamiento semántico, en el que se extrae el significado de las palabras. Finalmente, la última fase identifica los distintos sentimientos (Rayón, 2021).

El presente trabajo de investigación trata de emplear el mayor número de técnicas con el objetivo de dar una mejor explicación sobre lo que engloba esta fusión, tratando de examinar tanto los sentimientos como los temas más importantes. Por ello, se realiza el análisis y se compara con el precio de la acción de CaixaBank y como evolucionan con este tipo de comentarios, en función de si cuando hay buenos comentarios asciende su precio o por el contrario disminuye.

3.2.1 Problemática con Twitter

Sin embargo, realizar un análisis sobre las opiniones que surgen en twitter tiene grandes problemas por diversas razones. Los tweets de las personas están condicionados por diversos factores que pueden influir de una manera negativa para realizar un análisis objetivo:

- **Facilidad de influir en la opinión de los demás:** Tus amigos más cercanos pueden apoyar tu opinión con facilidad, tratando de cambiar la opinión de otros usuarios, e incluso evitar que estos juzguen de una manera objetiva.
- **Situaciones de spam:** La gran facilidad que aporta la red social para twittear, hace imposible controlar este tipo de situaciones, apareciendo en muchas ocasiones opiniones sin relevancia. Este hecho provoca que el nivel de los tweets, que los usuarios consumen, estén por debajo del nivel que puedan llegar a demandar.

- **Limitación de caracteres:** Este factor provoca que los grandes usuarios no puedan expresar todo lo que piensan en un solo tweet, provocando que se tengan que crear hilos para poder seguir comentando. Esto suscita que la información este distorsionada por opiniones de otras personas entre tweet y tweet, provocando que no se encuentre la información que realmente se esta buscando. Por ello, los usuarios demandan que se aumente la cantidad de caracteres. Sin embargo, ya se produjo un aumento de 140 caracteres a 280, que ha resultado simbólica.
- **Desaparición de tweets:** Al día se mandan más de 500 millones de tweets, provocando la desaparición de tweets anteriores, los cuales nunca volverán a aparecer. La única solución para que no desaparezcan es la utilización de un hashtag siempre que vayas a escribir un tweet. Este fenómeno ocasiona un consumo de tiempo que la gente no está dispuesto a asumir, limitando las opiniones de las personas. Este se convierte en el principal problema de twitter y de realizar un análisis de sentimientos mediante tweets (Atribus, 2020).
- **Existencia de perfiles falsos:** Hoy en día, es muy común la existencia de cuentas “fakes”, que tratan de informar a la población de una manera o de otra provocando que la credibilidad de los análisis y de las opiniones caigan en picado. Por ejemplo, en el mundo del fútbol hay una gran cantidad de cuentas que se dedican a la critica de una serie de jugadores, pero que realmente no se sustentan en ninguna razón firme.

Sin embargo, al ser una red social usada por millones de usuarios en todo el mundo, las ideas generan grandes efectos en temas importantes. Por ejemplo, personas tan influyentes como Elon Musk, fundador de Tesla, tienen un usuario y sus opiniones provocan grandes cambios en los precios de las criptomonedas como en el precio de la acción de su compañía y de otras. Por ello, Twitter ha sido la red escogida para la obtención de la base de datos. Esta base de datos será seleccionada como objeto de estudio para llevar a cabo el análisis de este trabajo.

4. Metodología

La metodología que implica realizar este trabajo es muy compleja por lo que se adaptan a las técnicas requeridas del *Business Analytics*. En este apartado, se trata de explicar los distintos pasos y herramientas que se han llevado a cabo. Para comprenderlo mejor vamos a dividirlo en cuatro subapartados:

- **Descarga de los datos necesarios:** Explicaremos como se ha llevado a cabo la descarga de la base de datos de Twitter y los datos financieros para el posterior análisis y las conclusiones sobre el trabajo.
- **Tratamiento de los datos:** En este apartado, se explica como se ha procedido a adaptar los datos previamente obtenidos para que el estudio se lleve de una manera más eficiente.
- **Análisis de resultados obtenidos:** Se realizará una breve exposición de lo que vendrá en el apartado 5 del trabajo. En este apartado, se ve como son los datos de una manera más analítica y clara.
- **Análisis de sentimientos:** Se explicarán todos los pasos dados para finalmente llegar al mayor propósito del trabajo. Normalizaremos el texto y eliminaremos las palabras que nos puedan estorbar a la hora de analizar los distintos textos. Finalmente, se realizará un análisis semántico, en el que se atribuirán las palabras a distintitas categorías.

4.1 Descarga de los datos necesarios

Para la investigación llevada a cabo en este proyecto ha sido necesaria realizar dos tipos de descargas de datos. En primer lugar, se han descargado los precios de cotización de CaixaBank. En segundo lugar, la descarga de los tweets, que es la más complicada se ha realizado a través de la librería de Python “*Snsrape*” tras haber probado otras técnicas.

4.1.1 Datos financieros CaixaBank

Los datos sobre las cotizaciones de CaixaBank se han extraído de la plataforma financiera *yahoo finance*. Por otro lado, los datos sobre la cotización de Bankia no han procedido a descargarse puesto que el movimiento perjudicaría en mayor medida a CaixaBank, la cual será la entidad nueva resultante. Estos datos representan sus valores de cotización desde septiembre de 2020, momento en el que tuvo lugar el anuncio de la fusión entre estas dos grandes entidades, hasta el 1 de abril de 2021, una semana después de cerrarse por completo la transacción.

Todos los datos vienen representados en distintas columnas: fecha, apertura, máximo, mínimo, cierre, cierre ajustado y volumen. El precio de cierre (cierre) es la cifra a la que asciende después de los ajustes por splits. El precio de cierre ajustado (cierre ajustado) es el precio de cierre concertado por dividendos y splits.

Los datos han sido tratados previamente antes de ser utilizados para el estudio. El principal motivo por el que se han limpiado los archivos es por qué *Yahoo Finance* trabaja con puntos como si fuesen decimales en vez de utilizar comas. Finalmente, con la realización de una serie de arreglos en R, se obtuvo el *dataset* que va a ser objeto de estudio.

4.1.2 Datos de Twitter

Esta sección de descargas es uno de los pasos más importantes del estudio y, por lo tanto, uno de los más complejos y costosos de tiempo del trabajo. El procedimiento de la descarga se intento llevar a cabo mediante el acceso a la API de twitter, en la cual se debe crear una cuenta de desarrollador. Dicha cuenta se admite con gran complejidad, y con esta cuestión solo era posible extraer los tweets de la última semana, no los de una serie histórica. Por ello, este proceso fue descartado.

A la imposibilidad de encontrar una solución en el lenguaje R, se procedió a intentar, por diversas librerías, su descarga mediante el lenguaje de programación de Python⁴. En primer lugar, se elaboró la descarga a través de la librería *GetOldTweets3*⁵, librería creada por Jefferson Henrique (GitHub,2019), ya que nos permitía recopilar una mayor cantidad de tweets en cualquier franja de tiempo, pero esta librería fue bloqueada. Ello nos imposibilitó la liberación de dichos tweets haciendo más complicado el trabajo y dando lugar a una demora a la hora de la realización del estudio.

Posteriormente, se procedió al uso de la librería *tweepy*. Esta librería era más compleja, ya que volvía a requerir el acceso a la API de Twitter anteriormente expuesta, acción que se complico en el trabajo debido a la inhabilitación por parte del equipo de ingenieros hacia mi cuenta. Asimismo, esta librería tenia el mismo problema que la

⁴Es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código

⁵ Una biblioteca de Python 3 y su correspondiente utilidad de línea de comandos para acceder a tweets antiguos. (pypi, 2020)

programación en R, ya que solo se podían obtener los tweets de la última semana haciendo imposible un buen análisis de sentimientos sobre el tema en el que se ha trabajado.

Finalmente, se procedió a la descarga de los tweets relacionados entre CaixaBank y Bankia, mediante la librería “*snsrape*”, considerada la sustitutiva de *GetOldTweets3* por su incapacitación (Beck, 2020). Esta librería nos permite la descarga de la cantidad de tweets que queremos sin restricciones en cualquier idioma y entre cualquier franja temporal. La extracción se realizó en el idioma español, ya que se trata de dos entidades españolas con escasa presencia internacional, pues los comentarios serán realizados en mayor parte por la comunidad hispano hablante.

A continuación, se va a mostrar un extracto del código utilizado:

Ilustración 4: Extracto de código Python para la extracción de Tweets

```
keyword = 'Caixabank Bankia'  
  
(sntwitter.TwitterSearchScrapper(keyword + ' lang:es since:2020-07-10 until:2021-03-30 -filter:links -filter:replies').get_items()):
```

Fuente: Elaboración propia mediante lenguaje de programación python

Como se puede observar en la ilustración 4 la búsqueda se ha realizado con las palabras claves conjuntas, “CaixaBank y Bankia”, puesto que tras probar por separado se obtenían una gran cantidad de tweets que no tenían relación de la fusión. Por ello, se procedió a examinarlas juntas pues se obtendrían los tweets claves desde una semana antes a la fecha en la que se anunció el acuerdo de fusión, 17 de septiembre de 2020, y una semana después de formalizarse la fusión, 23 de marzo de 2021, pues así hay un mayor feedback de como los usuarios de Bankia se han adaptado al cambio y como opinan sobre la fusión con solo una semana. El fin de escoger hasta una semana después es para ver como de rápida ha sido la adaptación de la nueva entidad.

Concluimos en que no ha sido tarea fácil la descarga de los tweets, más teniendo en cuenta que se ha tenido que ir posponiendo continuamente el “dataset” hasta el 1 de abril para que el estudio fuese más preciso lo que provocaba retrasar aún más el trabajo de investigación, del que se pretende sea lo más exacto posible.

4.2 Tratamiento de los datos

Los datos extraídos de twitter poseen una serie de complejidades a la hora de realizar un análisis de este calibre. Por ello, es necesario adaptar el “dataset” mediante la limpieza de este para que el estudio se puede llevar a cabo de una manera concreta. Entre estas impurezas destacan los signos de puntuación, las etiquetas HTML, los caracteres sueltos, los patrones no informativos, las etiquetas div o los URLs de pagina web (Castro, 2020). Asimismo, se pueden incluir aquellos símbolos que no representan nada a la hora de realizar un análisis como pueden ser: “@, &, \$, #”.

Este tratamiento se considera fundamental para llevar a cabo el análisis, pues sin el se obtendrían conclusiones poco relevantes y que no llevarían a ningún lugar.

4.2.1 Limpieza de datos

Para afrontar la suciedad generada en el “dataset”, se ha empleado una función del propio lenguaje R, la cual se caracteriza por poder adaptarse fácilmente a los diferentes tipos de texto analizados. La función *Gsub* es una de las funciones que vienen por defecto en R para trabajar con texto. 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, voy a mostrar la serie de limpieza que he llevado a cabo mediante un código de R, empleando la librería *tydr* para emplear *gsub*:

Ilustración 5: Extracto de código limpieza “dataset”

```
datos$tweet<-gsub("RT", "", datos$tweet)
datos$tweet<-gsub("@", "", datos$tweet)
datos$tweet<-gsub("https://", "", datos$tweet)
datos$tweet<-gsub("[\t]{2,}", "", datos$tweet)
datos$tweet<-gsub("^ ", "", datos$tweet)
datos$tweet<-gsub(" $", "", datos$tweet)
datos$tweet<-gsub(" #", "", datos$tweet)
datos$tweet<-gsub("W", "", datos$tweet)
datos$tweet<-gsub("w", "", datos$tweet)
datos$tweet<-gsub("http://", "", datos$tweet)
datos$tweet<-gsub(">>", "", datos$tweet)
```

Fuente: Elaboración propia en lenguaje de programación de R Studio

Como podemos observar en la ilustración 5, se trata de eliminar cualquier tipo de texto irrelevante para el análisis de la fusión entre ambas entidades. Por ello, omitimos los RT⁶ para que no se repita en exceso el mismo mensaje, el @ para no mostrar la entidad de donde provienen estas opiniones; su procedencia web, así como las tabulaciones que surgen de los datos extraídos de fuentes de internet. Finalmente, omitimos una serie de textos que nos resultan irrelevantes para este estudio concreto.

4.3 Análisis de los resultados obtenidos

En este apartado se van a mostrar los resultados obtenidos en el análisis de sentimientos, así como la visualización de las cotizaciones de CaixaBank durante el periodo de la fusión y qué ha podido influir en ellas. Este apartado se realiza para averiguar si es útil realizar un análisis de sentimientos en temas económicos, ya que, en el ámbito del marketing sabemos que el uso de estas nuevas técnicas ha superado con creces a los métodos antiguos. Este es uno de los motivos por los que decidí involucrarme en este tema, tan fascinante de investigación.

4.4 Análisis de sentimientos

En análisis de sentimientos se han empleado varias herramientas. Para empezar, se ha realizado un análisis léxico. Posteriormente, se ha continuado con un examen semántico y un estudio de *topic modelling* de donde se pueden extraer los principales temas tratados en los tweets con sus caracteres más comunes y relacionados. Finalmente, se extrajo la conclusión de si, en el conjunto de tweets, predomina el sentimiento positivo o negativo.

4.4.1 Análisis sintáctico y léxico

En primer lugar, vamos a preparar el entorno de trabajo para conseguir unos resultados coherentes con en el estudio. Por ello, se procede a la realización de la *tokenización* del texto y a la creación de una estructura de análisis.

⁶ publicar el blog de otro usuario en el sitio web de Twitter para sus propios seguidores.

La normalización del texto, también conocida como *tokenización*, consiste en separar los tweets en las palabras de las que se compone, eliminando las palabras de “parada” (también denominadas *stopwords*) (Castro, 2020). El paquete con el que contamos en R se denomina *stopwords*, donde los vectores del carácter de las palabras son devueltos en diferentes idiomas.

Para nuestro estudio hemos creado un vector en idioma castellano, que posteriormente se convirtió en un *dataframe* para poder elaborar el estudio. En este *dataframe* se encuentran palabras que pueden resultar irrelevantes para el estudio, como pueden ser CaixaBank, Bankia, entidades... y sus seudónimos, ya que son las palabras objeto de estudio y los tweets ya van relacionadas con ellas. El resultado de todo ello es la creación de un *lexicón* de *stopwords* en español, que será empleado en procesos futuros.

A continuación, se procede a eliminar esas *stopwords* para que nos proporcione un análisis más claro y sencillo de entender. Para ello, se ha creado una estructura de los datos y se ha implementado un código que nos permite anularlas, llegando a eliminar el 49,2% del total de las palabras que poseía el *dataframe*.

4.4.2 Procesamiento semántico

En este subapartado se van a realizar varios posibles procedimientos semánticos para el análisis. En primer lugar, se comienza con la realización de un análisis *topic modelling*, el cual 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, 2021).

Este análisis conllevará a la creación de una *Keywordcloud*, la cual consiste en la aparición de una nube de palabras, donde a medida que se nombran más en un mismo tweet se van concentrando entre ellas, y a medida que aparecen en más tweets aparecen en mayor tamaño, mediante la librería de R *wordcloud*. Para ello, es necesario la creación de un corpus de las palabras a lo que hay que añadirle la eliminación de las mencionadas anteriormente *stopwords*. Con este tipo de ilustraciones obtendremos información útil, pero que necesitaremos complementar con ilustraciones posteriores.

Un estudio adicional es el llamado análisis de biagramas, donde las palabras aparecen conectadas mediante puntos con sus palabras más repetidas mostrándose cuales tienen mayor relación. En este caso, ha resultado imposible la realización de este, debido

a la gran cantidad de tweets y caracteres que poseíamos y, por tanto, no se obtenían evidencias claras y se ha optado por abandonar este tipo de análisis.

Finalmente, llevamos a cabo un análisis clúster, como los demás análisis vamos a tener que realizar un corpus, para posteriormente eliminar las *stopwords*. Asimismo, para este procedimiento debemos convertir el corpus construido previamente en un topic para poder realizar el *topic modelling*.

El segundo paso, consiste en determinar el número óptimo de topics que queremos obtener para saber cuales son los más relevantes. Esta función es posible realizarla mediante la librería proporcionada por R, llamada *cvtools*. Al tratarse de un análisis de opiniones muy general y amplio este procesamiento nunca obtendría un número óptimo, ni si quiera empleando el método del codo, “*Elbow*”, donde nos podría mostrar, mediante la forma de un codo en su inicio, las desviaciones desde muy arriba hasta que hace el movimiento más brusco y se mantiene constante, pues ese punto sería el número óptimo que deberíamos escoger, pero resulta imposible como anteriormente he explicado.

Por ello, se ha optado por escoger diez topics como el número óptimo, puesto que se repetían y no aportaban más información. De este modo, he podido profundizar en cada topic que tipo de palabra es la que más se obtiene. También, podemos ver que topic es el que mayor repercusión tiene, es decir, el que tiene más opiniones y tweets relacionados con el mismo.

A continuación, se va a mostrar la nube de palabras obtenida, ya que se trata de un gráfico de iniciación:

Ilustración 6: Nube de palabras respecto a CaixaBank y Bankia



Fuente: Elaboración propia mediante R Studio.

De la ilustración 6, se pueden obtener una serie de conclusiones muy relevantes también, pero que deben sostenerse en otros análisis. Podemos observar que la palabra más comentada es fusión como era de esperar, a la que le siguen palabras con millones y dinero, lo que lleva a pensar que la opinión de los usuarios es que, esta operación, puede mover una importante cantidad de efectivo. Por otro lado, se sigue hablando de rescate, por lo que entendemos que la opinión pública es que dicha unión supone un rescate de CaixaBank hacia Bankia. Incluso, aparecen otras entidades bancarias, como Santander y BBVA, donde encontramos comentarios de que seguirán siendo superiores o, por el contrario, se verán superados por esta fusión si no realizan ningún movimiento a corto plazo. Por otro lado, parecen estar un poco preocupados por las comisiones resultantes de la nueva entidad, así como del rescate que supuso Bankia en un pasado y de como se integrará el gobierno en la entidad.

5. Resultados del análisis y visión del precio de cotización

En este apartado se va a llevar a cabo una serie de análisis sobre los resultados obtenidos, así como ciertas conclusiones que podemos obtener de este análisis comparado con la evolución de la cotización de la acción de CaixaBank desde el anuncio de la fusión hasta una semana después de efectuarse. Se ha cogido este periodo por la situación de la pandemia que hemos vivido durante el año 2020, la cual afectó a la totalidad de las entidades bancarias, e incluso a todo negocio.

5.1 Análisis de la cotización de CaixaBank

Los datos extraídos previamente de *yahoo finance*, después de haber sido manipulados para adaptarlos a las necesidades del estudio, nos ha permitido obtener la siguiente ilustración, donde se nos muestran los movimientos bursátiles de CaixaBank, desde el 17 de septiembre de 2020 y 30 de marzo de 2021.

Ilustración 7: Evolución precio de cotización CaixaBank comprendido entre septiembre de 2020 y abril de 2021



Fuente: Elaboración propia mediante RStudio a partir de datos extraídos de yahoo finance

A simple vista, en la ilustración 7 podemos apreciar como el precio de cotización de la entidad se encontraba en plena caída, pero este hecho fue causado en gran parte por la incertidumbre creada por la COVID-19, pandemia que provoco la parada de la economía mundial. Por ello, en los primeros momentos es complicado determinar si el precio de cotización sufriría un gran impacto por el anuncio de acuerdo de fusión entre las dos grandes entidades.

Sin embargo, tras su anuncio se puede observar como hubo ciertos puntos donde la caída se detuvo e incluso su precio aumentó, pero esto resultaría un simple espejismo

para la entidad que continuaría bajando, alcanzando su precio más bajo durante un periodo de corto plazo. A mediados de noviembre cuando la fusión iba cogiendo forma y contaba con el apoyo de analistas, y muchas personas influyentes empezaron a comentar de forma positiva, el precio de la acción subió de una manera notable durante unos días para luego estabilizarse y continuar cotizando al mismo precio durante los siguientes meses previos a la formalización de la fusión.

Finalmente, se puede seguir observando que el precio de cotización de la entidad está creciendo de una manera constante y estable, incluso después de haberse realizado la fusión entre ambas entidades; las dificultades que supone esa aceptación; y la reestructuración que deben llevar a cabo. El análisis se ha realizado desde un punto de vista subjetivo, basándonos en los conocimientos que tenemos sobre la situación que ha rodeado a toda esta fusión, así como los motivos que han llevado a la compañía a realizarla. Por ello, nos vamos a apoyar en el análisis de sentimientos para profundizar en que medida influye esta fusión en las personas, y por consiguiente, estas, en la cotización de la entidad. Por otro lado, podremos obtener como se percibió el anuncio y sus meses posteriores, además de su influencia en los clientes.

A continuación, se va a mostrar una ilustración sobre el volumen de las cotizaciones entre el periodo comprendido de septiembre de 2020 hasta abril de 2021.

Ilustración 8: Evolución del volumen de cotización CaixaBank entre septiembre de 2020 y abril de 2021



Fuente: Elaboración propia mediante RStudio a partir de datos extraídos de yahoo finance

Por otro lado, mediante la ilustración 8 podemos observar como la magnitud de volumen de cotización se intensificó al principio, cuando su precio de cierre estaba en

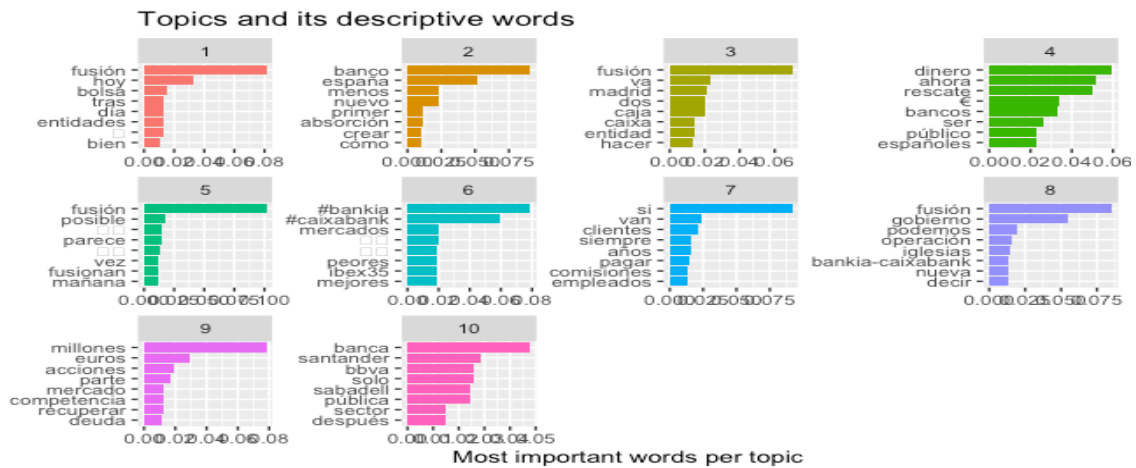
continúa caída, para posteriormente mantenerse estable hasta que, cuando empezó a subir su precio, su volumen también aumentó. Finalmente, el momento de mayor volumen en la cotización del precio de CaixaBank se observa que se produce tras la finalización del proceso de fusión de las entidades. Esto se debe a la reorganización de las acciones, y las negociaciones de estas en las transacciones de los accionistas de Bankia por la absorción.

5.2 Análisis de sentimientos acerca de la fusión entre CaixaBank y Bankia

A continuación, vamos a exponer el análisis de sentimientos de los tweets obtenidos para el estudio. En primer lugar, vamos a llevar a cabo un análisis de los diez topics más mencionados con sus palabras más relevantes. Continuaremos, con los sentimientos mostrados por la gran parte de los usuarios, así como que grado de influencia tienen cada uno de los sentimientos.

Comparando la nube de palabras creadas a partir del *topic modelling*, podemos observar que existen palabras que coinciden, pero otras que no varían. El método empleado en la nube de palabras provoca asociaciones semánticas a través de un algoritmo, donde clasifica esos topics según un sentimiento del diccionario. Por otro lado, en el análisis semántico se estudian todas las palabras de los tweets a partir de los distintos *lexicón*, los cuales llevan asociados un sentimiento (Castro, 2020).

Ilustración 9: Topics más importantes de análisis de sentimientos respecto a CaixaBank y Bankia



Fuente: Elaboración propia mediante RStudio.

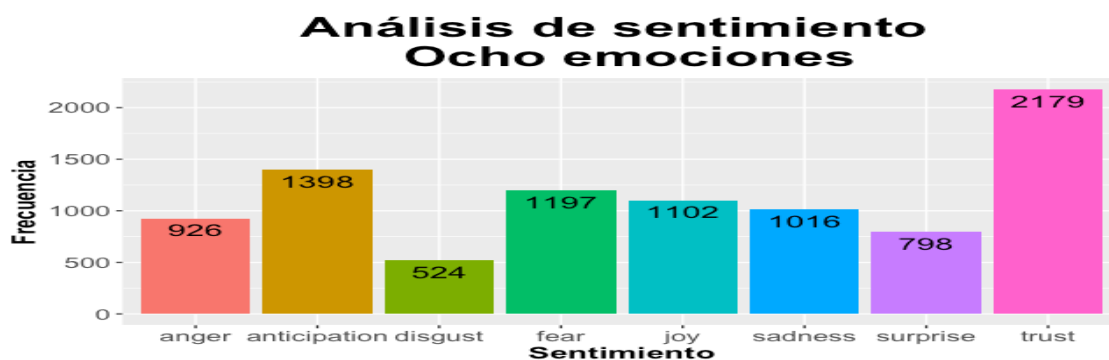
Como se puede observar en la ilustración 9, de los diez topics más importantes resultantes del análisis mediante el método *lexicón*, se llega a la conclusión que la palabra

predominante “fusión” en la nube de palabras, aparece como la palabra dominante en cuatro de los topics más importantes con una influencia de su uso muy grande. Esto quiere decir que, durante el periodo de tiempo estudiado, lo más comentado acerca de las entidades era sobre este tema, el de la fusión, por lo que hace de este estudio una investigación muy atractiva de analizar. Asimismo, se le asocia una operación del gobierno, con su posibilidad de hacerse efectiva, e incluso mencionando a cargos políticos como Pablo Iglesias.

Por otro lado, otro de los topics más repetidos es el relacionado con el dinero y la cantidad de efectivo que esta operación puede mover. La gente asocia estos términos también como un posible rescate hacia Bankia como ya se produjo en la anterior crisis, pero a través de una entidad privada como es CaixaBank, así como es una manera de recuperar deuda y aliarse contra la creciente morosidad que espera el futuro.

Como con ambos análisis no podemos medir de una manera efectiva que sentimientos reales muestra y con que frecuencia se manifiestan, se han unificado los diccionarios *bing* y *nrc* para aumentar la capacidad de palabras y los tipos de sentimientos que pueden aparecer en los tweets. A continuación, se va a mostrar un gráfico de barras donde aparecerán en comparación ocho sentimientos distintos con el grado que estos representan en el total de los tweets, haciendo los estudios más reales y proporcionándoles un enfoque real.

Ilustración 10: Frecuencia de los tipos de sentimientos respecto a CaixaBank y Bankia



Fuente: Elaboración propia mediante RStudio.

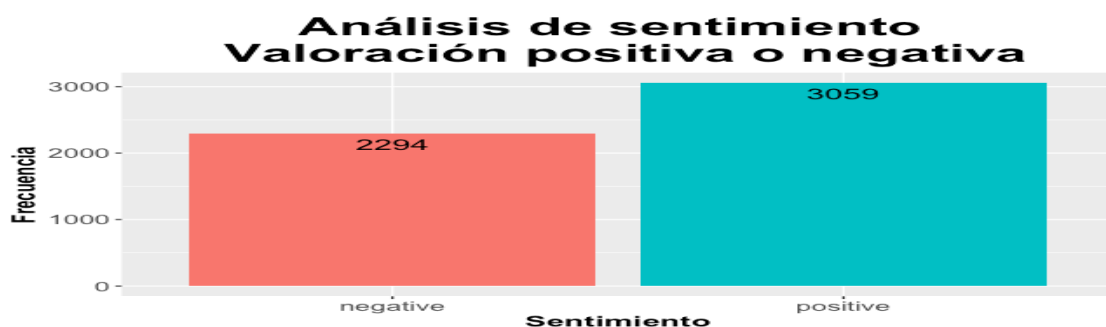
Lo primero de todo que se puede comentar de la ilustración 10 es que la entidad financiera resultante, CaixaBank, puede gozar de tranquilidad con respecto a sus clientes, puesto que tienen depositada demasiada confianza con este movimiento, lo que provoca

que no se producirá una salida muy elevada de estos, incluso puede atraer nuevos por el optimismo que se ha generado. En oposición, encontramos temor, lo que nos hace plantearos todo lo contrario. Con todo, nos decantamos por optimismo, ya que el temor es superado por este.

Como se puede observar en la ilustración, el sentimiento predominante, de entre todos los tweets, con mucha diferencia con respecto al segundo es el de confianza. La razón por la que pueda predominar este sentimiento se debe a la existencia de palabras como recuperación, crecimiento y respaldo dentro de los tweets analizados, mostrando que creen que la situación de la nueva entidad surgida de la fusión se convertirá en un seguro para sus clientes y una de las mejores y fiables entidades del país. Este sentimiento supone alrededor del 24% de los tweets. Asimismo, el segundo sentimiento más apreciado en el estudio es el sentimiento de anticipación, el cual surge por palabras donde se mencione la previsión, rapidez y anticipación, lo cual quiere decir que están previendo una situación difícil en el futuro provocada por la actual crisis y se están preparando contra cualquier tipo de peligro. Por otro lado, el tercer sentimiento más abundante es el de temor, lo que parece sorprendente viendo que los dos principales sentimientos son positivos hacia esta fusión. Esto puede deberse a la pésima situación de la economía provocada por la COVID-19, e incluso por la posible dificultad que se les preveía a ambas entidades. Finalmente, cabe destacar que el sentimiento de sorpresa es muy bajo, por lo que era un movimiento esperado por gran parte de la población.

Posteriormente, para continuar apoyando el estudio se va a realizar una comparación entre que sentimientos predominan más, basándonos en una variable binaria, es decir, distinguiendo los sentimientos entre positivos o negativos simplemente.

Ilustración 11: Comparación entre sentimientos positivos o negativos respecto a CaixaBank y Bankia

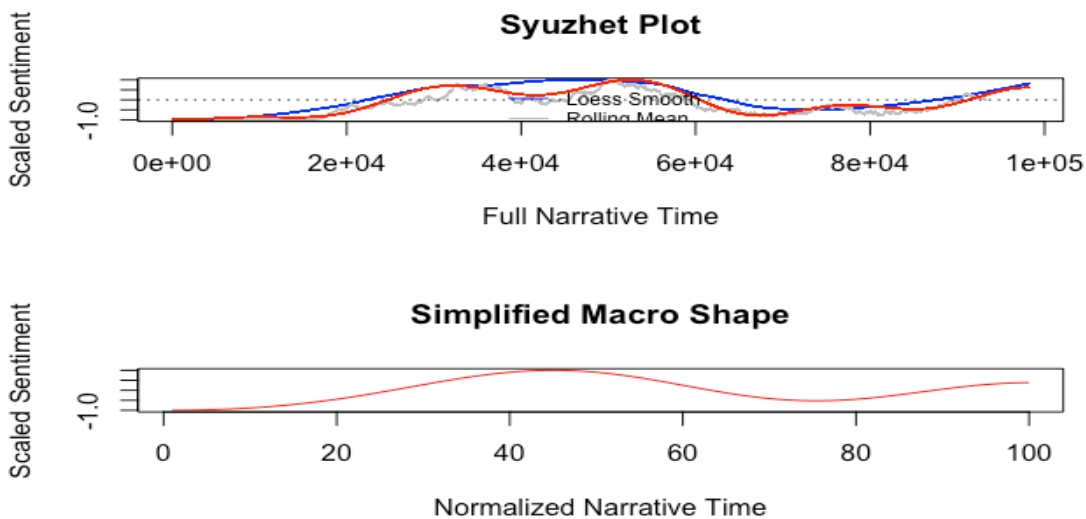


Fuente: Elaboración propia mediante RStudio.

En la ilustración 11, se muestra que la gran mayoría de los tweets tienen un sentimiento positivo, aunque no destaca de una manera sobresaliente sobre los negativos. Esto nos hace apoyar el por qué del crecimiento de la cotización de CaixaBank, pero que se trata de un crecimiento constante y con muy poca pendiente. Asimismo, era de esperar que esta fusión se viese de una manera positiva debido a la gran cantidad de entidades que existen en la actualidad, y con las reestructuraciones masivas que están por venir.

Por otro lado, vamos a observar la evolución de los sentimientos a lo largo del tiempo y con la cantidad de tweets sobre el tema en la siguiente ilustración.

Ilustración 12: Evolución de los sentimientos a lo largo del periodo determinado sobre CaixaBank y Bankia



Fuente: Elaboración propia mediante RStudio

En primer lugar, de la ilustración 12, se observa de una manera clara que los sentimientos respecto al tema de estudio son muy fluctuantes, es decir, varían mucho a lo largo del periodo determinado. Los sentimientos empiezan en una tendencia negativa que pronto se corregirá decantándose por el positivismo sobre la fusión de ambas entidades, posteriormente, aparece cierto temor respecto al tema para finalmente, tener un sentimiento neutro y alcista, cercano al positivismo.

Para acabar, vamos a volver a realizar una nube de palabras agrupándolas en dos sentimientos distintos, donde se podrá observar de una manera más clara los sentimientos positivos y negativos y que palabras los componen.

6. Conclusiones

En el presente trabajo de investigación se ha llevado a cabo un análisis sobre el impacto que puede generar las opiniones de los distintos usuarios de CaixaBank y Bankia a la nueva entidad correspondiente de la fusión entre ambas (CaixaBank). Desde la llegada de la crisis causada por la COVID-19, la incertidumbre se ha apoderado del mundo, provocando que surgieran movimientos de tal calibre a una gran velocidad. Por ello, esta fusión es la primera en realizarse, en el ámbito nacional, para combatir esta crisis y cubrirse contra futuras situaciones de tensión. Esto ha provocado que sus competidores se adaptasen a este fenómeno tratando de realizar operaciones similares. Todo ello, más los sentimientos mostrados por los usuarios en las distintas redes sociales o puntos de información, pueden influir en su precio de cotización y en otras maneras del ámbito de trabajo de la nueva entidad, CaixaBank.

Con el transcurso del tiempo, Twitter no solo ha dejado de ser una red social donde se compartían opiniones entre los jóvenes, si no que también podemos encontrar medios de comunicación contrastados, donde publican sus noticias y la gente, cada vez más, trata de dar su visión sobre la situación, con el firme objetivo de influir en ciertas facetas de la economía. Es por ello, que twitter ha sido la base de datos elegida, además de que es una de la red sociales con mayor volumen de usuarios, lo que proporciona una gran diversidad de opiniones.

Durante el desarrollo del estudio se han podido emplear diversas herramientas que tienen una gran utilidad para varias áreas de trabajo. Hemos podido extraer datos históricos de twitter mediante Python. Además, se ha podido realizar un análisis de sentimientos sobre estas opiniones y demostrar si tienen un impacto positivo y negativo. Por otro lado, que es lo que más les preocupa a los clientes de estas entidades en el término de la fusión. Por último, se ha podido observar como ha evolucionado el precio de cotización de CaixaBank, pudiendo obtener ciertas conclusiones que sin el previo análisis no se hubiesen soportado en ninguna evidencia empírica.

Tras haber realizado un profundo análisis sobre el objeto de estudio, fusión entre CaixaBank y Bankia y que opinan sus usuarios sobre este fenómeno, se puede concluir que la información que se transmite por los medios de comunicación puede influir en el precio de cotización de la entidad, pero también es una buena herramienta para predecir lo que les preocupa y tratar de solucionar esas inquietudes antes de que sean mayores. Por

otro lado, se observa que en este tipo de comentarios e informaciones predomina un sentimiento positivo lo que no debe preocupar a la entidad en gran medida, pero si tratar algunos temas preocupantes. Este tipo de sentimientos no ha influido mucho en su precio de cotización, más bien la incertidumbre de la crisis es lo que más cambio ha generado en este factor.

Por último, durante el desarrollo del estudio han surgido una serie de limitaciones que hacen falta señalar con el firme propósito de que estudios futuros sepan a qué se enfrentan realmente, y cuales son las mayores dificultades.

6.1 Limitaciones surgidas durante el desarrollo del estudio.

6.1.1 Limitación de datos

La extracción llevada a cabo obtiene todos los datos en la franja de estudio delimitada, pero muchos usuarios pueden haber eliminado tweets que con anterioridad eran relevantes y pudieron afectar a la entidad, incluso no pueden obtenerse de varios idiomas a la vez por lo que limita el estudio a un solo idioma, el castellano, cuando opiniones en otros idiomas también pueden influir de una manera positiva o negativa a los distintos factores de la entidad.

Finalmente, en estos datos puede colarse información poco relevante dejando en un segundo plano ciertas opiniones que puede resultar muy relevantes para el estudio.

6.1.2 Dificultad de análisis de sentimientos

La primera limitación provoca la imposibilidad de un análisis de sentimientos tan preciso como si los datos fueran todos los necesarios y fueran todos relevantes. Esto provoca que el patrón de programación seguido escoja ciertos valores positivos que los considera neutros o negativos y viceversa. Por ello, a pesar de que este análisis se trate de un estudio muy completo podría conseguir una exactitud aún mayor si la base de datos se fuese actualizando día a día, pues podríamos resaltar que tweets son los borrados y cuales son los considerados escritos por *bots*.

6.1.3 Momento de realización del estudio

Otra gran limitación con la que se encuentra este proyecto ha sido el espacio histórico de su realización debido a que se trata de un momento muy duro a causa de la COVID-19, el cual ha afectado a gran parte de la población y no solo a la industria bancaria. Por ello, se ha intentado contextualizar lo mejor posible la situación para

entender mejor el movimiento que han empleado estas entidades. Asimismo, el estudio se ha llevado a cabo desde el momento de su anuncio hasta que se ha realizado la absorción por parte de CaixaBank. El marco más favorable para el estudio será tras el paso de un año tras la absorción de la compañía (23 de marzo de 2021), esto permitiría aumentar las opiniones de los expertos y las consecuencias que el movimiento tendrá.

7. Bibliografía

Agencias de comunicación. (2020). *La influencia de twitter como medio de comunicación*. Recuperado de: <https://www.agenciasdecomunicacion.org/actualidad/la-influencia-de-twitter-como-medio-de-comunicacion.html>

Alcocer, J. (2020). *El futuro del sector financiero*. Recuperado de: <https://www.pwc.es/es/publicaciones/financiero-seguros/el-futuro-del-sector-financiero.html>

Atribus. (2020). *Principales redes sociales para empresas, ventajas y desventajas*. Recuperado de: <https://www.atribus.com/principales-redes-sociales/>

Beck, M. (2020). *How to scrape tweets with snsrape*. Recuperado de: <https://betterprogramming.pub/how-to-scrape-tweets-with-snsrape-90124ed006af>

Calvo, A y De Vidales, I. (2014). *Estudios de la Economía aplicada, vol.32, núm 2, ulio-diciembre, 2014, pp. 536-566*. Recuperado de: <https://www.redalyc.org/pdf/301/30130732004.pdf>

Casas, B. (2020). *Hacia una nueva oleada de fusiones en el contexto de la COVID-19*. Recuperado de: https://cincodias.elpais.com/cincodias/2020/09/08/opinion/1599567431_095762.html

Castro, B. (2020). *¿Se puede medir el impacto del brexit sobre la cotización del banco Santander a través de los sentimientos de Twitter?*.

Clavero, V. (2021). *La nueva CaixaBank tras la fusión con Bankia echa a andar con un alto coste laboral y dudas para los clientes*. Recuperado de: <https://www.publico.es/economia/fusiones-bancarias-nueva-caixabank-fusion-bankia-echa-andar-alto-coste-laboral-dudas-clientes.html>

Climent, S. (2013). *La restructuración del sistema bancario español tras la crisis y la solvencia de las entidades financieras. Consecuencias para las cajas de ahorro*. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S1138489113000046>

Cooper, P. (2020). *140 estadísticas de redes sociales que son importantes para los mercadólogos en 2020*. Recuperado de: <https://blog.hootsuite.com/es/125-estadisticas-de-redes-sociales/>

Digitalcookhouse. (2014). *Limpiando data in R- función SUB o GSUB*. Recuperado de: <https://digitalcookhouse.wordpress.com/2014/08/23/limpiando-data-in-r-funcion-subgsub/#:~:text=La%20funci%C3%B3n%20GSUB%20es%20una,con%20otro%20texto%20que%20queramos>

Esteban, J. (2020). *La banca después de la COVID-19*. Recuperado de: <https://www.tendencias.kpmg.es/2020/06/la-banca-despues-de-la-covid-19/>

Fernández, C. J. (2018). *Análisis de sentimiento en R*. Recuperado de: <https://www.linkedin.com/pulse/an%C3%A1lisis-de-sentimiento-en-r-carlos-j%C3%A1uregui-fern%C3%A1ndez/>

Fernández de Guevara, J. (2008). *El sector bancario español en el contexto internacional: Evolución reciente y retos futuros*, pp 259-267.

Galeano, S. (2021). *Cuáles son las redes sociales con más usuarios del mundo*. Recuperado de: <https://marketing4ecommerce.net/cuales-redes-sociales-con-mas-usuarios-mundo-ranking/>

GitHub. (2019). *GetOldTweets3*. Recuperado de: <https://github.com/Mottl/GetOldTweets3>

Go, A., & Bhayani, R. (2009). *Twitter Sentiment Classification using distant supervision*. CS224N project report, Stanford 1 (12).

González, C. (2020). *La nueva ola de fusiones en la banca y el paradigma digital*. Recuperado de: <https://www.tendencias.kpmg.es/2020/12/la-nueva-ola-de-fusiones-en-la-banca-y-el-paradigma-digital/>

Hernández, M. (2020). *CaixaBank y Bankia aprueban su fusión y arrancan la creación del mayor banco de España*. Recuperado de: <https://www.elmundo.es/economia/empresas/2020/09/17/5f63759f21efa0322d8b45f5.html>

Hidalgo, C. (2021). *CaixaBank y Bankia culminarán este viernes su fusión*. Recuperado de: <https://www.economiadigital.es/empresas/caixabank-bankia-consuman-fusion.html>

Lafraya, C. (2020). *CaixaBank y Bankia convocan en diciembre las juntas para la fusión*. Recuperado de:

<https://www.lavanguardia.com/economia/20201024/484250545769/caixabank-bankia-fusion-fechas-union.html>

Larrouy, D. (2020). *Las nuevas fusiones van a convertir a España en el país europeo con más despidos en banca desde la crisis de 2008*. Recuperado de: https://www.eldiario.es/economia/nuevas-fusiones-convertir-espana-pais-europeo-despidos-banca-crisis-2008_1_6280670.html

Lozano, E. (2021). *CaixaBank recortará con un ERE su plantilla tras la fusión con Bankia*. Recuperado de: <https://www.elindependiente.com/economia/finanzas/2021/02/15/la-nueva-caixabank-recortara-un-10-su-plantilla-y-un-20-su-red-tras-fusionarse-con-bankia/>

Maudos, J. (2013). *El sector bancario español en el contexto internacional: El impacto de la crisis*.

Maudos, J. (2011). *El sector bancario español en el contexto internacional: el impacto de la crisis*, Fundación de las Cajas de Ahorros, Madrid.

Maudos, J y Fernández de Guevara, J. (2008). *El desarrollo y la estructura financieros de la economía española: La importancia del sector bancario*. Recuperado de: https://www.fbbva.es/wp-content/uploads/2017/05/dat/c1_desarrollo%20y%20estructura%20financieros.pdf

Moran, P. (2020). *Bankia y la evolución del sistema bancario europeo*. Recuperado de: <https://www.ieb.es/bankia-y-la-evolucion-del-sistema-bancario-europeo/>

Patiño, M. (2021). *CaixaBank busca desatascar su fusión con Bankia*. Recuperado de: <https://www.expansion.com/empresas/banca/2021/03/18/605283e6e5fdea9a7f8b45c0.html>

Rayón, A. (2021). *Text Analytics*. Materiales docentes Diploma en Fundamentos de Business Analytics, Universidad Pontifica de Comillas.

Rayón, A. (2021). *API y Webscraping: técnicas de enriquecimiento de bases de datos*. Materiales docentes Diploma en Fundamentos de Business Analytics, Universidad Pontfica de comillas.

Rodríguez, C. (2019). *Clasificación de documentos de texto con R*. Recuperado de: <http://sigma.iimas.unam.mx/carloserwin/clasificacion.html#:~:text=La%20estructura%2>

0principal%20para%20administrar,que%20se%20basar%C3%A1%20nuestro%20an%C3%A1lisis.&text=Una%20nube%20de%20palabras%20es,texto%20C%20tienen%20un%20mayor%20tama%C3%B1o

Sánchez, H. (2017). *¿Qué es el sentimental analysis?* Recuperado de: <https://www.postedin.com/blog/que-es-el-sentiment-analysis/>

Sobrino, R. (2020). *De 55 a 11 entidades desde 2009: la consolidación bancaria entre dos crisis.* Recuperado de: https://cincodias.elpais.com/cincodias/2020/09/11/companias/1599849568_554343.html

Universo Machine Learning. (2017). *5 librerías de R para la limpieza y exploración de datos.* Recpuerado de: <https://conocemachinelearning.wordpress.com/2017/06/23/5-librerias-de-r-para-la-limpieza-y-exploracion-de-datos/>

Vallverdú, J. (2013). *“Sentimental analysis” : ¿Qué es, como funciona, para qué sirve?* Recuperado de: <https://www.tecnonews.info/error/>

8. Anexos

8.1 Anexo 1: Código de análisis de sentimientos

```
#Lee el archivo
tuits <- read.csv("~/Desktop/TWEETS.csv", comment.char="#")
# Instala los paquetes:
install.packages("syuzhet")
install.packages("RColorBrewer")
install.packages("wordcloud")
install.packages("tm")

# Carga los paquetes
library(syuzhet)
library(RColorBrewer)
library(wordcloud)
library(tm)

texto_palabras <- get_tokens(tuits)
head(texto_palabras)
length(texto_palabras)
oraciones_vector <- get_sentences(text)
length(oraciones_vector)
sentimientos_df <- get_nrc_sentiment(texto_palabras,
lang="spanish")
head(sentimientos_df)
summary(sentimientos_df)
barplot(
  colSums(prop.table(sentimientos_df[, 1:8])),
  space = 0.2,
  horiz = FALSE,
  las = 1,
  cex.names = 0.7,
  col = brewer.pal(n = 8, name = "Set3"),
  main = "Tweets sobre la fusión de CaixaBank y Bankia",
  sub = "Análisis realizado por Santiago Sánchez",
```

```

xlab="Emociones", ylab = NULL)
barplot(colSums(prop.table(sentimientos_df[, 1:8])))
palabras_tristeza <- texto_palabras[sentimientos_df$sadness > 0]
palabras_tristeza_orden <-
sort(table(unlist(palabras_tristeza)), decreasing = TRUE)
head(palabras_tristeza_orden, n = 12)
length(palabras_tristeza_orden)
nube_emociones_vector <- c(
  paste(texto_palabras[sentimientos_df$sadness > 0], collapse = "
"),
  paste(texto_palabras[sentimientos_df$joy > 0], collapse = " "),
  paste(texto_palabras[sentimientos_df$anger > 0], collapse = "
"),
  paste(texto_palabras[sentimientos_df$fear > 0], collapse = "
"))
nube_corpus <- Corpus(VectorSource(nube_emociones_vector))

nube_tdm <- TermDocumentMatrix(nube_corpus)
nube_tdm <- as.matrix(nube_tdm)
head(nube_tdm)
colnames(nube_tdm) <- c('tristeza', 'felicidad', 'enfado',
'confianza')
head(nube_tdm)
set.seed(757) # puede ser cualquier número
comparison.cloud(nube_tdm, random.order = FALSE,
  colors = c("green", "red", "orange", "blue"),
  title.size = 1, max.words = 100, scale = c(1,
1), rot.per = 0.4)

sentimientos_valencia <- (sentimientos_df$negative *-1) +
sentimientos_df$positive
simple_plot(sentimientos_valencia)

install.packages("SnowballC")
install.packages("tm")

```

```

install.packages("twitterR")
install.packages("syuzhet")

library(SnowballC)
library(tm)
library(twitterR)
library(syuzhet)
library(NLP)

tweets <- read.csv("tweets.csv")

#Convirtiendo los tweets en un data frame
tweets.df <- twListToDF(tweets)
head(tweets.df)

#Quitando los links en los tweets
tweets <- gsub("http.*", "", tweets$tweet)
tweets <- gsub("https.*", "", tweets)

#Quitando los hashtags y usuarios en los tweets
tweets <- gsub("#\\w+", "", tweets)
tweets <- gsub("@\\w+", "", tweets)

#Quitando los signos de puntuación, números y textos con números
tweets <- gsub("[[:punct:]]", "", tweets)
tweets <- gsub("\\w*[0-9]+\\w*\\s*", "", tweets)

#Transformamos la base de textos importados en un vector para
#poder utilizar la función get_nrc_sentiment
palabra.df <- as.vector(tweets)

#Aplicamos la función indicando el vector y el idioma y creamos
#un nuevo data frame llamado emocion.df

```

```

emocion.df <- get_nrc_sentiment(char_v = palabra.df, language =
"spanish")

#Unimos emocion.df con el vector tweets.df para ver como
#trabajó la función get_nrc_sentiment cada uno de los tweets
emocion.df2 <- cbind(tweets, emocion.df)
head(emocion.df2)

#Creamos un data frame en el cual las filas serán las emociones
#y las columnas los puntajes totales

#Empezamos transponiendo emocion.df
emocion.df3 <- data.frame(t(emocion.df))

#Sumamos los puntajes de cada uno de los tweets para cada emocion
emocion.df3 <- data.frame(rowSums(emocion.df3))

#Nombramos la columna de puntajes como cuenta
names(emocion.df3)[1] <- "cuenta"

#Dado que las emociones son los nombres de las filas y no una
variable

#transformamos el data frame para incluirlas dentro
emocion.df3 <- cbind("sentimiento" = rownames(emocion.df3),
emocion.df3)

#Quitamos el nombre de las filas
rownames(emocion.df3) <- NULL

#Verificamos el data frame
print(emocion.df3)

install.packages("ggplot2")
library(ggplot2)

```

#Primer gráfico: se detallaran las 8 emociones con sus puntajes respectivos

```
sentimientos1 <- ggplot(emocion.df3[1:8,],
                        aes(x = sentimiento,
                            y = cuenta, fill = sentimiento)) +
  geom_bar(stat = "identity") +
  labs(title = "Análisis de sentimiento \n Ocho emociones",
        x = "Sentimiento", y = "Frecuencia") +
  geom_text(aes(label = cuenta),
            vjust = 1.5, color = "black",
            size = 5) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text = element_text(size=12),
        axis.title = element_text(size=14,face = "bold"),
        title = element_text(size=20,face = "bold"),
        legend.position = "none")
print(sentimientos1)
```

#Segundo gráfico: se detallan los puntajes para las valoraciones
#positiva y negativa

```
sentimientos2 <- ggplot(emocion.df3[9:10,],
                        aes(x = sentimiento,
                            y = cuenta, fill = sentimiento)) +
  geom_bar(stat = "identity") +
  labs(title = "Análisis de sentimiento \n Valoración positiva o
negativa",
        x = "Sentimiento", y = "Frecuencia") +
  geom_text(aes(label = cuenta),
            vjust = 1.5, color = "black",
            size = 5) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text = element_text(size=12),
        axis.title = element_text(size=14,face = "bold"),
```

```

        title = element_text(size=20,face = "bold"),
        legend.position = "none")
print(sentimientos2)

# Instalamos la librería que vamos a necesitar
library(tidyverse) # Importa varios paquetes diferentes. Los más
importaqntes: dplyr, magrittr, purrr y tidyr
library(rvest) # Para el scraping
library(tidytext) # Para la manipulación de textos
library(udpipe) # Parts of Speech tagging
library(data.table) # Para trabajar con data.tables
library(BTM) # Biterm Topic Modeling
library(widyr) # Representación bigramas
library(igraph)
library(gggraph)
library(quanteda)
library(topicmodels)
library(scales) # Útil para el formateo de campos numéricos

# Vamos a leer los datos que necesitamos
datos<-read_csv("TWEETS.csv")
# Vamos a quitar la primera columna que no nos hace falta
datos<-datos[,c(3)]
colnames(datos)[1]<-"textoComentario"

# Vamos a ver qué tal están los tipos de datos
str(datos)
# Solo la estructura de datos parece que nos falta por convertir
datos<-as.data.frame(datos)

#####
#####

#####
#####

```

```

# 1. Procesamiento léxico y sintáctico

#####
#####

#####
#####

# Nos creamos un lexicon de stopwords en español

lexiconSW<-stopwords("es")

lexiconSW<-
append(lexiconSW,c("Caixabank","caixabank","@caixabank","Bankia"
,"bankia","@bankia","+"))

lexiconSW<-as.data.frame(lexiconSW)

names(lexiconSW)<-"palabra"

lexiconSW$palabra<-as.character(lexiconSW$palabra)

# Preparamos el tipo de datos

datos$textoComentario<-as.character(datos$textoComentario)

# Creamos una estructura de datos para el análisis
df <- tibble::rowid_to_column(datos, "ID")

# Quitamos las stopwords: miramos el código para ello
lista_palabras <- df %>%
  dplyr::distinct(textoComentario, .keep_all = TRUE) %>%
  unnest_tokens(palabra, textoComentario, drop = FALSE) %>%
  distinct(ID, palabra, .keep_all = TRUE) %>%
  anti_join(lexiconSW) %>%
  filter(str_detect(palabra, "[^\\d]")) %>%
  group_by(palabra) %>%
  dplyr::mutate(word_total = n()) %>%
  ungroup()

# 2. Procesamiento semántico

# 2.1. Keywordcloud

# Creamos el corpus

corpus <- corpus(datos$textoComentario)

cdfm <- dfm(corpus, remove=c(stopwords("spanish"),

```



```

"Caixabank", "caixabank", "@caixabank", "Bankia", "bankia", "@bankia"
, "+"),

      verbose=TRUE,                                remove_punct=TRUE,
remove_numbers=TRUE)

cdfm <- dfm_trim(cdfm, min_docfreq = 2, verbose=TRUE)

# Vamos a hacer una nube de palabras
textplot_wordcloud(cdfm,
                   rot.per=0,
                   scale=c(5.5, 1.25),
                   max.words=50,
                   color    =    rev(RColorBrewer::brewer.pal(10,
"RdBu")))

# 2.2. Biterm Topic Modeling
# anotacion    <- udpipe(datos$textoComentario,
#                       "spanish",
#                       trace = 10)
# biterms <- as.data.table(anotacion)
# biterms <- biterms[, cooccurrence(x = lemma,
#                                   relevant = upos %in% c("NOUN",
"ADJ", "VERB") &
#                                   nchar(lemma) > 2 & !lemma
%in% stopwords("es"),
#                                   skipgram = 3),
#               by = list(doc_id)]

# 2.3. Análisis de bigramas
# Podemos visualizarlo también
pares_palabras <- df %>%
  unnest_tokens(palabra, textoComentario) %>%
  anti_join(lexiconSW) %>%
  pairwise_count(palabra, ID, sort = TRUE, upper = FALSE)
# Nos generamos el listado de bigramas

```

```

# listadoBigramas<-
title_word_pairs[which(title_word_pairs$n>10),]

set.seed(1234)

pares_palabras %>%
  filter(n >= 100) %>%
  graph_from_data_frame() %>%
  ggraph(layout = "fr") +
  geom_edge_link(aes(edge_alpha = n, edge_width = n), edge_colour
= "cyan4") +
  geom_node_point(size = 5) +
  geom_node_text(aes(label = name), repel = TRUE,
                 point.padding = unit(0.2, "lines")) +
  ggtitle('Bigram Graph: combination of 2 keywords')

# 2.4. Análisis cluster
# Función de análisis clúster
corpus <- corpus(datos$textoComentario)
cdfm <- dfm(corpus, remove=c(stopwords("spanish"),

"Caixabank", "caixabank", "@caixabank", "Bankia", "bankia", "@bankia"
, "+", "q"),
             verbose=TRUE,                      remove_punct=TRUE,
             remove_numbers=TRUE)
cdfm <- dfm_trim(cdfm, min_docfreq = 2, verbose=TRUE)

# Ahora lo exportamos a un formato para procesar los Topic Models.
dtm <- convert(cdfm, to="topicmodels")

# Seleccionando el número de topics óptimo
# vis<-numeroTopicsOptimo(dtm)

# vis

# Obtenemos el LDA con el número óptimo de topics que nos haya
salido

lda <- LDA(dtm, k = 10, method = "Gibbs",
           control = list(verbose=25L, seed = 123, burnin = 100,
iter = 500))

```

```

# Obtenemos la palabra más representativa de cada topic
terms(lda)
# Las top 10-palabras de cada topic
trms <- t(terms(lda, k=10))
# Algunos topics son fáciles de identificar
trms[6,]
trms[1,]
trms[9,]
trms[10,]
trms[4,]
trms[9,]
trms[1,]

# Sacamos los datos de cada topic
terminosTopic <- tidy(lda, matrix = "beta")
terminosTopic

# Sacamos los top 20 términos por cada topic
top_terms <- terminosTopic %>%
  group_by(topic) %>%
  top_n(8, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

# Los visualizamos
top_terms %>%
  mutate(term = reorder(term, beta)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()+
  labs(x = NULL, y = "Most important words per topic",

```

```
title = paste0("Topics and its descriptive words")
```

8.2 Anexo 2: Código de descarga de Tweets mediante python

```
"""  
Created on Tue may 27 15:06:13 2021  
  
@author: santiagosanchez  
"""  
  
from platform import python_version  
print(python_version())  
!pip install snsrape  
#!pip install vaderSentiment  
import pandas as pd  
import numpy as np  
import snsrape.modules.twitter as sntwitter  
import csv  
import matplotlib.pyplot as plt  
  
# Set maximum tweets to pull  
maxTweets = 20000  
  
# Set what keywords you want your twitter scraper to pull  
keyword = 'Caixabank Bankia'  
  
#Open/create a file to append data to  
csvFile = open('pfe_tweets_result.csv', 'a', newline='',  
encoding='utf8')  
  
#Use csv writer  
csvWriter = csv.writer(csvFile)  
csvWriter.writerow(['id','date','tweet',])  
  
# Write tweets into the csv file
```

```
for i,tweet in enumerate(sntwitter.TwitterSearchScrapер(keyword +
' lang:es since:2020-07-10 until:2021-03-30 -filter:links -
filter:replies').get_items()):
    if i > maxTweets :
        break
        csvWriter.writerow([tweet.id, tweet.date,
tweet.content])
csvFile.close()
```