

## **Uso de Twitter durante los últimos días de una campaña electoral. Elecciones presidenciales en EEUU 2020.**

**Nombre Estudiante: Antonio Sánchez Navarro**

Plan de Estudios del Estudiante: Máster de Ciencia de Datos (*Data Science*)

Área del trabajo final: M2.878 – TFM – Área 1

**Nombre Consultor/a: Josep Maria Grau Masot**

**Nombre Profesor/a responsable de la asignatura: Albert Solé Ribalta**

3 de enero de 2021

A mis padres, por el apoyo que siempre me han mostrado y por soportar con paciencia infinita mis repentinos cambios de humor.



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Uso de Twitter durante los últimos días de una campaña electoral. Elecciones presidenciales en EEUU 2020.</i>
<b>Nombre del autor:</b>	<i>Antonio Sánchez Navarro</i>
<b>Nombre del consultor/a:</b>	<i>Josep Maria Grau Masot</i>
<b>Nombre del PRA:</b>	<i>Albert Solé Ribalta</i>
<b>Fecha de entrega (mm/aaaa):</b>	01/2021 (Defensa pública, entrega final)
<b>Titulación::</b>	Máster de Ciencia de Datos ( <i>Data Science</i> )
<b>Área del Trabajo Final:</b>	M2.878 – TFM – Área 1 aula 1
<b>Idioma del trabajo:</b>	<i>Español</i>
<b>Palabras clave</b>	<i>Twitter, campaña electoral, 2020 Elections</i>
<p><b>Resumen del Trabajo (máximo 250 palabras):</b> <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i></p>	
<p>A día de hoy, es una evidencia que asistimos a un nuevo cambio en el paradigma de la comunicación en el que Twitter mantiene un lugar prominente dentro del ámbito político, tanto nacional como internacional. Las razones que justifican su enorme relevancia son sus posibilidades de interactividad, capacidad de alcance y rápida difusión en los mensajes.</p> <p>El objetivo principal de este trabajo es, a partir de los tweets obtenidos mediante un hashtag neutro y las menciones a los dos principales candidatos, hacer una detección de perfiles relevantes obtenidos en esta muestra y un análisis de los mismos, permitiendo determinar las posibles diferencias en el uso de estos mecanismos de participación en Twitter. El evento elegido consiste en las elecciones presidenciales de los Estados Unidos de América, con fecha el martes 3 de noviembre de 2020. Este objetivo permitirá realizar un análisis cualitativo de sus tweets, así como descubrir qué dicen los usuarios con más menciones a los candidatos. Se espera que del análisis se obtengan resultados como gráficos de palabras más frecuentes y al menos un <i>word cloud</i>, así como el análisis de sentimientos. Además, el <i>top</i> de palabras, la valoración de sentimientos según diccionarios y el contenido de algunos tweets representativos permitirá detectar diferencias, si las hubiera, entre los que utilizan el hashtag neutro y los que mencionan los nombres de los principales candidatos, así como diferencias entre las menciones a cada candidato.</p> <p>Para ello, se obtienen los tuits mediante una etiqueta (Hashtag) neutra y dos palabras clave tipo nombre de usuario en Twitter. En este punto toma especial relevancia el uso de librerías en R, como es el caso de <i>twitterR</i>.</p> <p>Una vez se dispone de un corpus de tweets asociado a cada uno de estos conjuntos de datos, a partir de las búsquedas detalladas en el párrafo anterior, se utilizan métricas que apuestan por el recuento de menciones en la detección de perfiles, para determinar su relevancia y sus mensajes en la red.</p> <p>La presentación de resultados, así como las conclusiones de este trabajo, se irán</p>	

publicando en próximas entregas, conforme el trabajo vaya tomando forma y se encuentre en una fase más madura.

**Abstract (in English, 250 words or less):**

Nowadays, it is evidence that we are witnessing a new change in the communication paradigm in which Twitter maintains a preeminent place within the political sphere, both nationally and internationally. The reasons that justify its enormous relevance are its possibilities for interactivity, outreach capacity and rapid dissemination in messages.

The main objective of this work is, from the tweets obtained through a neutral hashtag and the mentions of the two main candidates, to detect relevant profiles obtained in this sample and analyze them, allowing to determine the possible differences in the use of these participation mechanisms on Twitter. The chosen event consists of the presidential elections of the United States of America, dated Tuesday, November 3, 2020. This objective will allow a qualitative analysis of their tweets, as well as discovering what the users with the most mentions of the candidates say. The analysis is expected to yield results such as charts of the most frequent words and at least one word cloud, as well as sentiment analysis. In addition, the top of words, the evaluation of sentiments according to dictionaries and the content of some representative tweets will allow to detect differences, if any, between those who use the neutral hashtag and those who mention the names of the main candidates, as well as differences between the mentions to each candidate.

To do this, the tweets are obtained using a neutral hashtag and two keywords such as username on Twitter. At this point, the use of R libraries, such as `twitteR`, takes special relevance.

Once a corpus of tweets associated with each of these data sets is available, from the searches detailed in the previous paragraph, metrics are used that rely on the count of mentions in the detection of profiles, to determine their relevance and your messages on the web.

The presentation of results, as well as the conclusions of this work, will be published in future installments, as the work takes shape and is in a more mature phase.

## Índice

<b>1. Introducción</b> .....	<b>1</b>
1.1 Contexto y justificación del Trabajo .....	1
1.2 Objetivos del Trabajo .....	2
1.3. Enfoque y método seguido .....	3
1.4 Planificación del Trabajo .....	5
1.5 Breve resumen de productos obtenidos .....	7
1.6 Breve descripción de los otros capítulos de la memoria .....	7
1.7 Espacio de trabajo elegido .....	8
<b>2. Marco teórico</b> .....	<b>8</b>
2.1 Campañas electorales y estado de la comunicación política en Twitter.....	9
2.1.1 Actividad de los líderes políticos en Twitter.....	15
2.2 Temas abordados por los políticos en Twitter.....	17
2.3 Big Data en redes sociales y Twitter.....	20
2.3.1 Minería de datos.....	20
2.3.2 Clasificación automática de textos.....	20
2.3.3 Tratamiento de los “issues” a partir de los “hashtags”.....	22
2.4.4 Metodologías observadas.....	23
<b>3. Diseño e implementación</b> .....	<b>24</b>
3.1 Radiografía de tweets.....	25
3.2 Análisis de contenidos.....	34
3.2.1 Palabras más frecuentes.....	35
3.2.2 <i>Word clouds</i> (nubes de palabras).....	37
3.2.3 Análisis de sentimientos.....	39
3.2.3.1 Análisis e interpretación de resultados.....	47
3.2.4 ¿Qué dicen los usuarios con más menciones en Twitter?.....	48
3.2.5 Estrategias seguidas por los principales candidatos.....	58
<b>4. Obtención de resultados y conclusiones</b> .....	<b>63</b>
4.1 Obtención de resultados.....	63
4.2 Conclusiones.....	64
4.3 Aspectos susceptibles de mejora.....	65
4.4 Líneas de trabajo futuras.....	66
<b>5. Bibliografía / Webgrafía</b> .....	<b>66</b>
<b>6. Anexo. Código R utilizado</b> .....	<b>70</b>

## Lista de figuras

Figura 1. Esquema elaboración del TFM .....	8
Figura 2. Características de Hardware. Espacio de trabajo elegido .....	8
Figura 3. Temas hablados en Twitter .....	17
Figura 4. Ataques a Hillary Clinton .....	18
Figura 5. Críticas al <i>Obamacare</i> .....	19
Figura 6. Situación económica .....	19
Figura 7. Distribución de los tweets capturados para el presente trabajo.....	26
Figura 8. Clasificación de tweets por grupos.....	26
Figura 9. Lugares de procedencia tweets asociados a “#Elections2020”.....	27
Figura 10. Lugares de procedencia tweets asociados a “@realDonaldTrump”.....	27
Figura 11. Lugares de procedencia tweets asociados a “@JoeBiden”.....	27
Figura 12. Distribución de tweets por hora para el conjunto de datos “#Elections2020”.....	28
Figura 13. Distribución de tweets por hora para el conjunto de datos “@realDonaldTrump”.....	28
Figura 14. Distribución de tweets por hora para el conjunto de datos “@JoeBiden”.....	29
Figuras 15 y 16. Top 20 hashtags más populares junto con “#Elections2020”.....	29-30
Figuras 17 y 18. Top 20 hashtags más populares junto con “@realDonaldTrump”.....	30-31
Figuras 19 y 20. Top 20 hashtags más populares junto con “JoeBiden”.....	31-32
Figura 21. Top 10 recuento de “screen_name” conjunto de datos “#Elections2020”.....	32
Figura 22. Top 10 recuento de “screen_name” conjunto de datos “@realDonaldTrump”.....	33
Figura 23. Top 10 recuento de “screen_name” conjunto de datos “@JoeBiden”.....	33
Figura 24. Top 10 mentions en el conjunto de datos “@realDonaldTrump”.....	34
Figura 25. Top 10 mentions en el conjunto de datos “@JoeBiden”.....	34
Figura 26. Palabras más frecuentes encontradas en el grupo “#Elections2020”.....	36
Figura 27. Palabras más frecuentes encontradas en el grupo “@realDonaldTrump”.....	36

Figura 28. Palabras más frecuentes encontradas en el grupo “@JoeBiden”	36
Figura 29. Nube de palabras asociada al grupo #Elections2020	37
Figura 30. Nube de palabras asociada al grupo @realDonaldTrump	38
Figura 31. Nube de palabras asociada al grupo @JoeBiden	38
Figura 32. Nube de palabras Trump vs Biden	39
Figura 33 Top 15 palabras agrupadas por sentimiento binario conjunto de datos “#Elections2020”	40
Figura 34. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@realDonaldTrump”	41
Figura 35. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@JoeBiden”	41
Figura 36. Top 15 palabras agrupadas por tipo de sentimiento conjunto de datos “#Elections2020”	42
Figura 37. Top 15 palabras agrupadas por tipo de sentimiento conjunto de datos “@realDonaldTrump”	42
Figura 38. Top 15 palabras agrupadas por tipo de sentimiento conjunto de datos “@JoeBiden”	43
Figura 39. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “#Elections2020”. Léxico “nrc”	43
Figura 40. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@realDonaldTrump”. Léxico “nrc”	44
Figura 41. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@JoeBiden”. Léxico “nrc”	44
Figura 42. Puntuación sentimientos encontrados conjunto de datos “#Elections2020”	45
Figura 43. Puntuación sentimientos encontrados conjunto de datos “@realDonaldTrump”	45
Figura 44. Puntuación sentimientos encontrados conjunto de datos “@JoeBiden”	45
Figura 45. Puntuación sentimientos TOP 20 palabras encontradas en el conjunto de datos ‘#Elections2020’	46
Figura 46. Puntuación sentimientos TOP 20 palabras encontradas en el conjunto de datos ‘@realDonaldTrump’	46
Figura 47. Puntuación sentimientos TOP 20 palabras encontradas en el conjunto de datos ‘@JoeBiden’	46
Figura 48. Palabras más frecuentes usuarios con más menciones de “#Elections2020”	48



Figura 49. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “#Elections2020”. Léxico “bing” .....	49
Figura 50. Palabras clasificadas por tipo de sentimiento. Usuarios con más menciones de “#Elections2020”. Léxico “nrc” .....	49
Figura 51. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “#Elections2020”. Léxico “nrc” .....	50
Figura 52. Puntuación sentimientos encontrados usuarios con más menciones a “#Elections2020” .....	50
Figura 53. Palabras más frecuentes usuarios con más menciones de “@realDonaldTrump” .....	51
Figura 54. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@realDonaldTrump”. Léxico “bing” .....	52
Figura 55. Palabras clasificadas por tipo de sentimiento. Usuarios con más menciones de “@realDonaldTrump”. Léxico “nrc” .....	52
Figura 56. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@realDonaldTrump”. Léxico “nrc” .....	53
Figura 57. Puntuación sentimientos encontrados usuarios con más menciones a “@realDonaldTrump” .....	53
Figura 58. Palabras más frecuentes usuarios con más menciones de “@JoeBiden” .....	54
Figura 59. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@JoeBiden”. Léxico “bing” .....	55
Figura 60. Palabras clasificadas por tipo de sentimiento. Usuarios con más menciones de “@JoeBiden”. Léxico “nrc” .....	55
Figura 61. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@JoeBiden”. Léxico “nrc” .....	56
Figura 62. Puntuación sentimientos encontrados usuarios con más menciones a “@JoeBiden” .....	56
Figura 63. Nube de palabras discurso de usuarios con más menciones a Trump y a Biden .....	58
Figura 64. Ataque personal a Joe Biden y elogio a su propia candidatura .....	59
Figura 65. Advertencia a los votantes de Joe Biden .....	59
Figura 66. Instando a la población de Arizona a votar y proclamando su eslogan .....	60
Figura 67. Acusando a Biden de político corrupto que va a dejar el país en manos de China .....	60
Figura 68. Crítica a Trump por su mala gestión sobre la pandemia .....	61

Figura 69. Compromiso y seriedad frente a la pandemia.....	62
Figura 70. Solicitando un donativo a sus votantes, junto a Kamila Harris.....	62
Figura 71. Ejemplo de mensaje positivo. Promesa de tiempos mejores que debe cumplirse.....	63

# 1. Introducción

Este primer capítulo de la memoria se corresponde con los contenidos de la PEC1 y está dedicado a la definición y planificación del TFM. Su principal propósito es definir claramente cuál es la temática del trabajo, justificar su interés y/o relevancia, qué se pretende conseguir una vez finalizado el TFM, definir los objetivos principales y una planificación temporal del proyecto. Además, incluye mi motivación personal como estudiante para hacer este trabajo final.

## 1.1 Contexto y justificación del Trabajo

En la mayoría de los procesos de toma de decisiones, las personas suelen hacer uso de la opinión de otros individuos para decantarse por una u otra opción. En un pasado no muy lejano, antes de la irrupción de Internet y de su actual omnipresencia en la mayoría de los aspectos de nuestras vidas, las principales fuentes de opinión eran el conocimiento y la experiencia de las personas más cercanas de nuestro círculo de relaciones y amistades, junto con los medios de comunicación de masas. Si bien estas fuentes de información tradicionales no han desaparecido, la implantación de Internet y la llegada de la llamada Web 2.0<sup>1</sup> han significado un profundo cambio en la manera en que las personas buscan opiniones que les ayuden durante el proceso de toma de decisiones. Internet se ha convertido en un océano inmenso en donde millones de personas expresan su opinión sobre cualquier tema y en cualquier momento. En este sentido, en la última década ha surgido un interés creciente por el estudio de las opiniones políticas utilizando los datos a gran escala que se producen en las aplicaciones de red social, como es el caso de Twitter. Las redes sociales han potenciado y motivado la participación de los movimientos y candidatos políticos en Internet, especialmente en época de campañas electorales. Más allá de los típicos anuncios en la televisión, los partidos intentan ahora captar el voto por otros muchos canales, entre ellos, las redes sociales. Por eso, los políticos se están viendo obligados a diversificar estrategias en diferentes plataformas, tal y como explica la profesora de los Estudios de Ciencias de la Información y de la Comunicación de la Universitat Oberta de Catalunya (UOC), Silvia Martínez en un debate organizado para estudiar el papel de las redes sociales como medios de comunicación. Según esta profesora, las experiencias en elecciones ya pasadas están cambiando algunas de las políticas y condiciones de uso de las redes sociales en campaña. Las redes sociales donde más se van a comentar las elecciones en el nuevo paradigma, y donde los políticos deberán poner todo su ingenio, serán Facebook y Twitter, que tendrán un importante papel como difusoras de información. Sin embargo, expertos de la UOC consideran que su poder será cada vez más quirúrgico, en el sentido de que los partidos políticos utilizarán ambas plataformas para programar campañas muy pensadas, muy bien segmentadas y dirigidas a perfiles muy concretos. En este sentido, los líderes de opinión serán los verdaderos difusores de las políticas.

Por todo lo expuesto anteriormente, realizaré una captura lo bastante representativa de tuits en el entorno de desarrollo integrado RStudio<sup>2</sup> que harán referencia a los

---

<sup>1</sup> Concepto atribuido a Tim O'Reilly y nombrado en la conferencia sobre la Web 2.0 de O'Reilly Media en 2004. El término establece una primera época de Internet en donde la comunicación era unidireccional, ya que los usuarios eran objetos pasivos que recibían la información publicada por los administradores de los sitios web sin posibilidad de interactuar. En la Web 2.0 aparecen los blogs, foros de debate, wikis y las redes sociales, dando lugar a una comunicación bidireccional en donde todos los actores de la red aportan contenido por igual.

<sup>2</sup> **RStudio** es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que

últimos días de las elecciones presidenciales del 3 de noviembre de 2020 en EE.UU. y obtendré una radiografía y un análisis de contenidos de los mismos. En concreto, el periodo estudiado abarca del 30 de octubre al 2 de noviembre de 2020. Ello me permitirá, dado que la muestra obtenida facilitará la comparación entre los tuits que usan el hashtag neutro y los que interpelan directamente a perfiles de candidatos, responder a preguntas como qué uso se hace de los tuits en cada caso y si se observan diferencias significativas en el tono utilizado. Podrían observarse estas diferencias, ya que el tono suele ser de talante menos comedido cuando se interpela a líderes políticos que cuando nos referimos a un evento o a un concepto abstracto. Mi hipótesis de partida es que existen diferencias significativas en el contenido de las menciones para cada conjunto de datos, especialmente entre los perfiles relevantes que más mencionan a cada uno de los principales candidatos. A partir del análisis de los resultados obtenidos, esta hipótesis será corroborada o desmentida en el apartado de conclusiones del trabajo. La motivación personal para realizar esta tarea se alimenta, además, de una ilusión tremenda. La razón es que emprender un trabajo de estas características implica estar finalizando unos estudios tan apasionantes como arduos y difíciles. La sensación es parecida a la de ver la luz al final del túnel tras haber andado durante mucho tiempo.

Para terminar esta sección, considero oportuno señalar que decidí hacer este TFM por interés personal, ya que fue la opción que me pareció más atractiva entre todas las ofertadas.

## 1.2 Objetivos del Trabajo

Si bien en la primera PEC los objetivos eran muy orientativos y genéricos, en las sucesivas PEC's se han ido refinando. En esta sección establecemos un primer listado de objetivos principales y secundarios.

La presente investigación persigue un doble objetivo, que podemos identificar con los objetivos principales del proyecto:

- Obtención, limpieza y radiografía de los mensajes generados y difundidos en Twitter durante los últimos días de campaña electoral correspondiente a las elecciones presidenciales del 3 de noviembre en EE. UU..
- Análisis de contenido en las publicaciones correspondientes tanto a los que más mencionan a los principales candidatos como a los que más utilizan el hashtag neutro que alude a la campaña electoral. Detección de lo que se dice por parte de los que más mencionan cada conjunto de datos.

En cuanto a objetivos más específicos o secundarios, se pueden listar los siguientes:

- Lugares de procedencia de tuits, así como usuarios con más menciones para cada uno de los conjuntos de tuits obtenidos. Tratamiento análogo a las principales menciones detectadas.

---

apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.

RStudio está disponible para Windows, Mac y Linux o para navegadores conectados a RStudio Server o RStudio Server Pro (Debian / Ubuntu, RedHat / CentOS, y SUSE Linux). RStudio tiene la misión de proporcionar el entorno informático estadístico R. Permite un análisis y desarrollo para que cualquiera pueda analizar los datos con R.

- Distribución de tuits en el tiempo durante el día de estudio 30 de octubre de 2020.
- Determinación de las palabras clave en las menciones a la campaña electoral y a los principales candidatos durante el periodo objeto de estudio. Se trata de determinar qué dicen los que mencionan a los principales candidatos y al hashtag neutro durante los últimos días de campaña electoral y qué asuntos han concentrado su atención.
- Obtención de, al menos, un Word Cloud y análisis de sentimientos del conjunto de tuits, tanto de las menciones a la campaña electoral como de las menciones a los principales candidatos.
- Análisis cualitativo (palabras más repetidas y recuento de repeticiones) de los mensajes publicados tanto en el conjunto de menciones a la campaña electoral como en el de menciones a los principales candidatos. Permite obtener aspectos sobre el contenido de los mensajes.
- Evaluación de las formas, estilo y mensaje usados en Twitter por los que más mencionan a los candidatos principales. Junto con lo aprendido en la detección de palabras más frecuentes y en el análisis de sentimientos, se trata de detectar qué dicen en Twitter los usuarios con más menciones a cada candidato. Se intentará evidenciar si existen diferencias significativas entre ambos conjuntos de menciones, así como si están alineados con algún tipo de estrategia electoral. Para ello, se seleccionarán varios tuits publicados correspondientes a menciones a cada uno de los candidatos que sean lo bastante representativos como para evidenciar las intenciones y el mensaje subyacente de quienes más mencionan.

### 1.3 Enfoque y método seguido

La estrategia elegida para llevar a cabo el trabajo es la de desarrollar un producto nuevo a partir de la recolecta de tuits procedentes de la red social. El trabajo a realizar se descompondrá en partes más pequeñas o etapas que se irán atacando una a una, marcando unos objetivos que se desean satisfacer en cada hito o entrega parcial. Para cada una de estas etapas, se reservarán “cojines de tiempo” o ciertas holguras que permitan recuperar posibles desviaciones que se puedan presentar. Cada una de las entregas parciales será, salvo imprevistos y tanto si se trata de capítulos de la memoria o partes del producto, una parte de la entrega final del Trabajo en su versión definitiva. Se establecerá, por tanto, un timing, una programación y una prioridad para cada fase.

Conviene señalar también que el TFM que nos ocupa es de temática investigadora. Es por ello que toma una gran importancia la definición de la problemática que se investiga, las fuentes de información proporcionadas y la metodología de investigación y su aplicación, que se basará en métodos tanto cuantitativos como cualitativos de análisis de datos.

Esta investigación toma como punto de partida la revisión de la bibliografía más destacada publicada hasta el momento y relacionada con el tema de estudio, pero además aporta un importante trabajo exploratorio de rastreo y seguimiento de tuits en la red social Twitter. Las razones que han llevado a elegir Twitter como red social para extraer tuits con el software R<sup>3</sup> son múltiples, si bien podemos destacar:

---

<sup>3</sup> R es un entorno y lenguaje de programación con un enfoque al análisis estadístico. Nació como una reimplementación de software libre del lenguaje S, adicionado con soporte para ámbito estático. Se trata de uno de los lenguajes de programación más utilizados en investigación científica, siendo además muy popular en los campos de aprendizaje automático

- Uso gratuito. Permite analizar tuits y aprender a codificar al mismo tiempo.
- Infinitas posibilidades de análisis. Permite realizar análisis a medida según lo que se desee analizar en lugar de depender de un único informe para todos.
- Permite analizar cualquier cuenta de Twitter que se desee, incluso sin disponer de los detalles de inicio de sesión. Esto representa una ventaja respecto a otras muchas herramientas de análisis.

En cuanto a la metodología empleada, hay una parte cualitativa muy importante a partir del análisis de sentimientos y la obtención de las palabras más frecuentes. En este sentido, para determinar qué se dice en las menciones a los principales candidatos electorales en Twitter y hallar diferencias significativas si las hubiera, se cuantifica la presencia de palabras asociadas a diccionarios de sentimientos de dichos candidatos y se hace un recuento binario de sentimientos positivos y negativos, realizando por tanto un análisis métrico en el conjunto de tuits con menciones a cada candidato. Ello permitirá entender los componentes de los mensajes, así como su orientación y su significado.

El proyecto se llevará a cabo utilizando el entorno de desarrollo RStudio para la captura y procesamiento de tuits y el procesador de textos Microsoft Word de Office para la redacción de la memoria. Los pasos o etapas de la metodología a seguir se citan a continuación:

- Obtención de un acceso a la API de Twitter. Para ello, si no se dispone de una cuenta de Twitter, conviene crear una. Una vez aceptada nuestra solicitud por parte de Twitter, se obtienen cuatro credenciales que se deben mantener a salvo:
  - Consumer key
  - Consumer Secret
  - AccessToken
  - Access Secret
- Extracción de los tuits a partir de la página web de Twitter con R y el paquete *twitteR*. Se obtienen tuits comprendidos entre el 30 de octubre y el 2 de noviembre de 2020.
- Limpieza del texto con los paquetes *stringr*, *tidytext* o *syuzhet*, eliminando los signos de puntuación, números, hipervínculos y palabras vacías.
- Construcción de una matriz de documentos de términos.
- Análisis de sentimientos con los diccionarios de palabras “bing” y “nrc”, así como el uso de las librerías “dplyr”, “tidytext”, “tidyverse”, “reshape2” y “RColorBrewer”.

En cuanto al análisis de tuits, además de la extracción y limpieza de tuits y el análisis de sentimientos ya comentados, se tendrán en consideración los siguientes pasos:

- Radiografía de tweets
- Palabras frecuentes y nube de palabras.

---

(machine learning), minería de datos, investigación biomédica, bioinformática y matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y graficación. R es parte del sistema GNU y se distribuye bajo la licencia GNU GPL. Está disponible para los sistemas operativos Windows, Macintosh, Unix y GNU/Linux.

- Análisis de contenido de los tweets

Y en cuanto a las herramientas utilizadas, se destacan las siguientes:

- IDE de desarrollo en lenguaje R RStudio.
- API de Twitter.
- Paquetes y librerías de R (además de la ya mencionada *twitteR*):
  - *topicmodels*
  - *sentiment140*
  - *igraph*

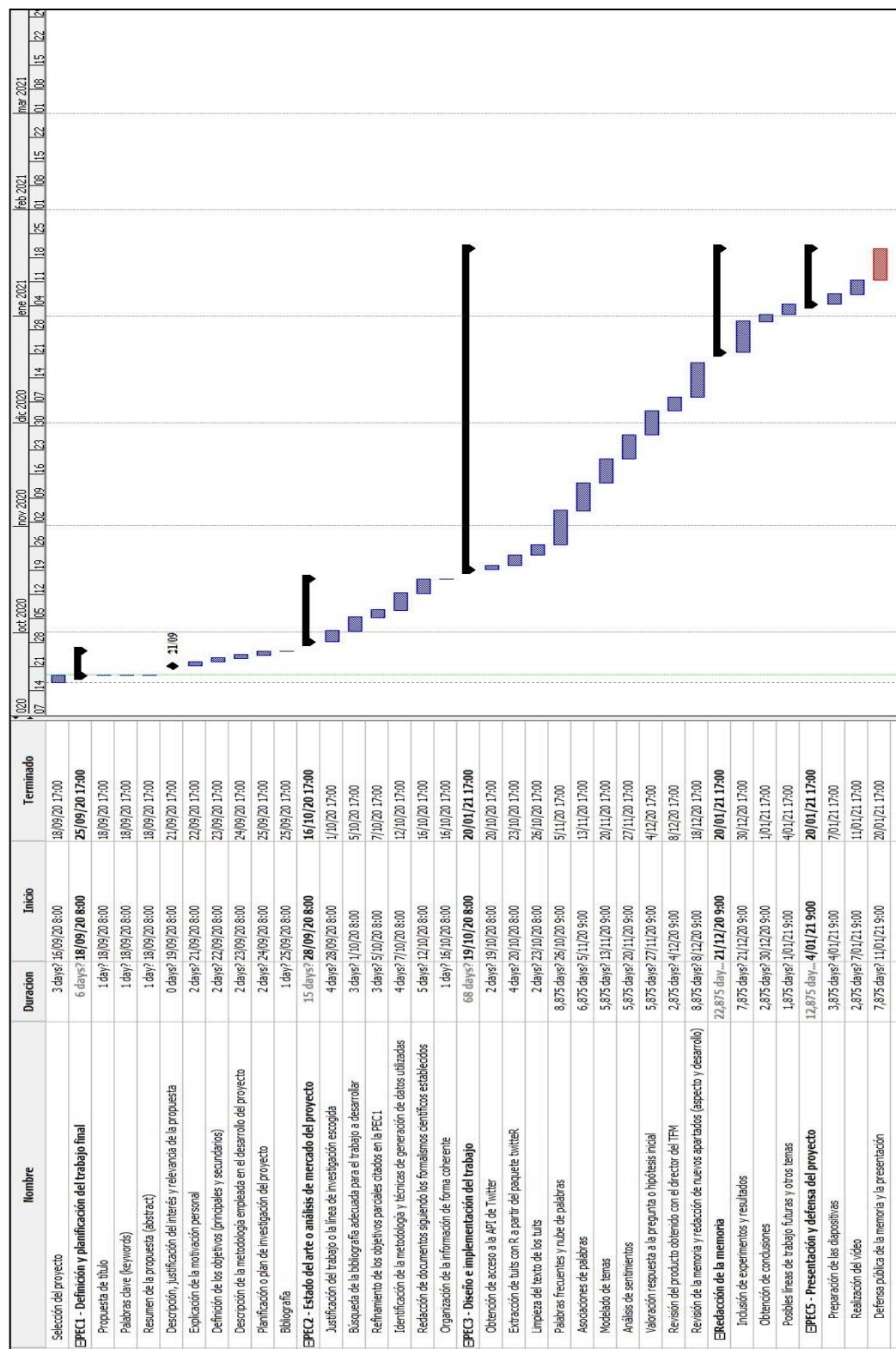
## 1.4 Planificación del Trabajo

En cuanto a los recursos necesarios para realizar este trabajo, conviene destacar los recursos que la UOC pone a disposición de los estudiantes. Son ejemplos tanto la Biblioteca como el software ofimático, así como los enlaces a internet, libros y webgrafía que facilita la propia asignatura. Resultan también imprescindibles las herramientas que he citado al final del punto anterior.

Una descripción detallada de las actividades a realizar y su planificación temporal contemplando los hitos a lograr, basándonos en tareas, actividades y volúmenes de trabajo, es la que se muestra en la siguiente tabla de hitos:

Nombre	Duración	Inicio	Terminado
Selección del proyecto	3 days?	16/09/20 8:00	18/09/20 17:00
<b>PEC1 - Definición y planificación del trabajo final</b>	<b>6 days?</b>	<b>18/09/20 8:00</b>	<b>25/09/20 17:00</b>
Propuesta de título	1 day?	18/09/20 8:00	18/09/20 17:00
Palabras clave (keywords)	1 day?	18/09/20 8:00	18/09/20 17:00
Resumen de la propuesta (abstract)	1 day?	18/09/20 8:00	18/09/20 17:00
Descripción, justificación del interés y relevancia de la propuesta	0 days?	19/09/20 8:00	21/09/20 17:00
Explicación de la motivación personal	2 days?	21/09/20 8:00	22/09/20 17:00
Definición de los objetivos (principales y secundarios)	2 days?	22/09/20 8:00	23/09/20 17:00
Descripción de la metodología empleada en el desarrollo del proyecto	2 days?	23/09/20 8:00	24/09/20 17:00
Planificación o plan de investigación del proyecto	2 days?	24/09/20 8:00	25/09/20 17:00
Bibliografía	1 day?	25/09/20 8:00	25/09/20 17:00
<b>PEC2 - Estado del arte o análisis de mercado del proyecto</b>	<b>15 days?</b>	<b>28/09/20 8:00</b>	<b>16/10/20 17:00</b>
Justificación del trabajo o la línea de investigación escogida	4 days?	28/09/20 8:00	1/10/20 17:00
Búsqueda de la bibliografía adecuada para el trabajo a desarrollar	3 days?	1/10/20 8:00	5/10/20 17:00
Refinamiento de los objetivos parciales citados en la PEC1	3 days?	5/10/20 8:00	7/10/20 17:00
Identificación de la metodología y técnicas de generación de datos utilizadas	4 days?	7/10/20 8:00	12/10/20 17:00
Redacción de documentos siguiendo los formalismos científicos establecidos	5 days?	12/10/20 8:00	16/10/20 17:00
Organización de la información de forma coherente	1 day?	16/10/20 8:00	16/10/20 17:00
<b>PEC3 - Diseño e implementación del trabajo</b>	<b>68 days?</b>	<b>19/10/20 8:00</b>	<b>20/01/21 17:00</b>
Obtención de acceso a la API de Twitter	2 days?	19/10/20 8:00	20/10/20 17:00
Extracción de tuits con R a partir del paquete <i>twitteR</i>	4 days?	20/10/20 8:00	23/10/20 17:00
Limpieza del texto de los tuits	2 days?	23/10/20 8:00	26/10/20 17:00
Palabras frecuentes y nube de palabras	8,875 days?	26/10/20 9:00	5/11/20 17:00
Asociaciones de palabras	6,875 days?	5/11/20 9:00	13/11/20 17:00
Modelado de temas	5,875 days?	13/11/20 9:00	20/11/20 17:00
Análisis de sentimientos	5,875 days?	20/11/20 9:00	27/11/20 17:00
Valoración respuesta a la pregunta o hipótesis inicial	5,875 days?	27/11/20 9:00	4/12/20 17:00
Revisión del producto obtenido con el director del TFM	2,875 days?	4/12/20 9:00	8/12/20 17:00
Revisión de la memoria y redacción de nuevos apartados (aspecto y desarrollo)	8,875 days?	8/12/20 9:00	18/12/20 17:00
<b>Redacción de la memoria</b>	<b>22,875 day...</b>	<b>21/12/20 9:00</b>	<b>20/01/21 17:00</b>
Inclusión de experimentos y resultados	7,875 days?	21/12/20 9:00	30/12/20 17:00
Obtención de conclusiones	2,875 days?	30/12/20 9:00	1/01/21 17:00
Posibles líneas de trabajo futuras y otros temas	1,875 days?	1/01/21 9:00	4/01/21 17:00
<b>PEC5 - Presentación y defensa del proyecto</b>	<b>12,875 day...</b>	<b>4/01/21 9:00</b>	<b>20/01/21 17:00</b>

Por otra parte, la herramienta ProjectLibre<sup>4</sup> nos proporciona el diagrama de Gantt que se detalla a continuación:



<sup>4</sup> **ProjectLibre** es un software de administración de proyectos de código abierto, similar a Microsoft Project. ProjectLibre corre sobre la plataforma Java, lo que permite ejecutarlo en varios sistemas operativos. ProjectLibre fue lanzado en agosto de 2012 y ya ha sido bajado desde más de 175 países. Ha sido votado como proyecto de código abierto del mes de octubre de 2012<sup>2</sup> por la comunidad de SourceForge.



## 1.5 Breve resumen de productos obtenidos

Los productos que se esperan obtener son un banco de tuits correspondientes al seguimiento de una campaña electoral en unas fechas determinadas y una memoria que explique todos los aspectos clave del proyecto. Los tuits serán limpiados y refinados para su análisis posterior, lo que nos permitirá obtener una radiografía de los mismos y un análisis de sus contenidos. Se intentará llevar a cabo también un análisis de sentimientos tanto del grueso de tuits de la campaña electoral como de los principales candidatos electorales. En cuanto a la memoria, contendrá aspectos realizados en PECs anteriores y añadirá otros nuevos, como resultados, discusión de los mismos, conclusiones y líneas de trabajo futuras. Esta memoria, por tratarse de un trabajo final de investigación, contemplará dos partes: una parte introductoria que explicará los elementos clave de investigación (justificación del trabajo, temática elegida, objetivos, metodología y planificación), y el desarrollo de una investigación que tratará todo o una parte del trabajo realizado y su producto final.

## 1.6 Breve descripción de los otros capítulos de la memoria

El resto de los capítulos o etapas de la memoria quedan descritos como sigue:

- **Estado del arte o análisis de mercado del proyecto.** En esta etapa se completa la parte del estado del arte de la memoria, de modo que se explica qué investigación o solución se ha desarrollado hasta el momento para resolver la problemática en la que se trabaja. Esto es lo que se conoce como “Estado del arte” o “Estado de la cuestión”. Conviene señalar que esta etapa puede inducir también a la ampliación o modificación de la explicación de la temática elegida y de la problemática o pregunta a resolver, ya que se pueden detectar matices que se desconocían en el momento de escribir la actividad anterior. Esto no debe representar ningún problema, ya que siempre es posible corregir o modificar algunos detalles realizados anteriormente. Esta etapa podría contemplar varias entregas parciales antes de la definitiva para garantizar que el trabajo se realice correctamente. Se corresponde, además, con la PEC2 del TFM.
- **Diseño e implementación del trabajo.** Llegado el momento de diseñar y desarrollar/implementar el/los producto/s, en esta fase se realizan las tareas definitivas mencionadas en la planificación del trabajo (PECs 1 y 2). Esta etapa empieza en la PEC3 y también asume que puede tener varias entregas parciales antes de la entrega definitiva para garantizar la correcta realización del trabajo.
- **Redacción de la memoria.** Etapa crucial del trabajo en la que se elabora la versión definitiva del documento que recoge todos los aspectos clave del proyecto a desarrollar. Recoge aspectos como el planteamiento del trabajo, la ejecución del mismo, los resultados obtenidos, las propuestas de mejora y la bibliografía utilizada. Esta etapa se corresponde con la PEC4.
- **Presentación y defensa del proyecto.** Dado el carácter académico del TFM, además de la memoria se precisa una evaluación diferente a la de otras asignaturas del su plan de estudios. En esta actividad se realiza una presentación del TFM a través del campus, que debe consistir en una presentación de unas veinte diapositivas y quince minutos de duración máxima. Esta actividad se corresponde con la PEC5.

- **Defensa pública.** Como último paso del TFM, la defensa contempla también la resolución de unas preguntas formuladas por un tribunal. Esta defensa puede realizarse de forma síncrona en el marco del aula virtual, siendo el profesor del aula quien determine su formato llegado el momento oportuno.

Por todo lo expuesto, el proceso de elaboración de este TFM se esquematiza como sigue:

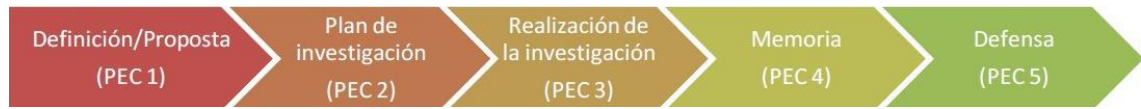


Figura 1. Esquema elaboración del TFM  
Fuente: Elaboración propia

## 1.7 Espacio de trabajo elegido

El espacio de trabajo se reduce a un PC de sobremesa con las siguientes características de Hardware:

- PC del fabricante Hewlett-Packard Company.
- Modelo CQ5315Es.
- Procesador Pentium(R) Dual-Core CPU E5300 @ 2.60 GHz.
- Memoria instalada (RAM): 4.00 GB.
- Sistema operativo Windows 10 Pro de 64 bits, procesador x64.
- Disco duro con sistema de archivos NTFS de 500 GB.

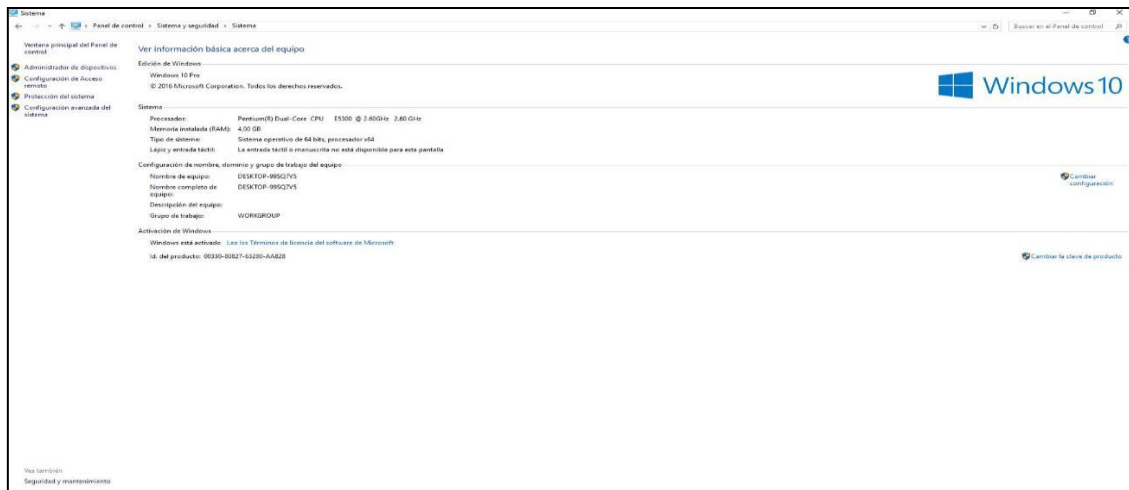


Figura 2. Características de Hardware. Espacio de trabajo elegido  
Fuente: Elaboración propia

## 2. Marco teórico

En este segundo apartado de la memoria, en estrecha correspondencia con los contenidos de la PEC2, me propongo conocer bien las problemáticas existentes en el mundo en cuanto a seguimiento y análisis de campañas electorales en Twitter u otras redes sociales, así como la fase en la que se encuentran. La metodología empleada para conseguir estos propósitos es el análisis de resultados procedentes de otros trabajos que ya han abordado esta misma temática o bien otras muy afines. Por

consiguiente, en esta parte de la memoria se completa el denominado Estado del arte<sup>5</sup> o Estado de la cuestión de la misma, explicando qué investigación (o soluciones) se han desarrollado hasta el momento para resolver la problemática del análisis de campañas electorales en redes sociales como Twitter y afines.

## **2.1 Campañas electorales y estado de la comunicación política en Twitter**

A partir de mediados de los 90 del siglo pasado, es una constante en círculos académicos y profesionales la reflexión acerca del poder que ha ido adquiriendo internet como herramienta de comunicación política y electoral (Selnow, 1998; Farrell, Kolodny & Medvic, 2001; Webster, 2001; Stein, 2003; Klotz, 2004; Hendricks & Kaid, 2010). De hecho, ya no se concibe una campaña sin el uso planificado y sistemático de este medio y ningún candidato desearía quedarse al margen de las nuevas posibilidades comunicativas que le puede brindar la Red.

Entre todas estas posibilidades comunicativas, las que más impacto están cobrando en los últimos años son las que proporcionan las redes sociales cuando se trata de interactuar con los ciudadanos, crear comunidades y establecer un modelo de comunicación bidireccional en el que los que lanzan los mensajes ya no sólo son los políticos, sino que los internautas hacen oír también su voz.

De todas estas redes sociales, Twitter es la que actualmente ha adquirido más relevancia entre la clase política y periodística. Se ha convertido “en uno de los mecanismos de comunicación más poderosos de la historia” (Piscitelli, 2011: 15). Además, “ha cambiado la red y ha completado el giro social que iniciaron los blogs a finales de los años noventa” (Orihuela, 2011: 21). No obstante, no todo son alabanzas a la Web 2.0 y a Twitter en particular. Marc Esteve del Valle sostiene que, junto a hallazgos interesantes, Twitter también propicia momentos desagradables debido a insultos, simplismos y abundantes ridiculizaciones (El País, 2020). Además, numerosos estudios muestran que esta red social acaba siendo una cámara de resonancia que amplifica la polarización política y la intolerancia entre posicionamientos ideológicos encontrados (por ejemplo, Connover et al, 2011; Garimella et al, 2018). Por su parte, Christian Fuchs afirma que “tenemos que cambiar las redes para salvar la democracia” (El Mundo, 2018), ya que nos encontramos ante “un intento sin precedentes de manipular el proceso democrático”.

---

<sup>5</sup> El Estado del Arte (o de la Cuestión) es una parte introductoria de cualquier TFG, TFM o tesis doctoral. Su ubicación habitual dentro de la memoria es el segundo capítulo (después de la introducción). No obstante, este puede estar fraccionado en más de un capítulo si el trabajo alberga varias temáticas o si la temática requiere la fragmentación.

Su contenido consiste en hacer un repaso de cómo los investigadores han abordado, hasta el momento, la problemática que la tesis quiere resolver. Por lo tanto, se puede ver como una fotografía del estado actual de un ámbito de conocimiento. Como se puede ver, el Estado del Arte requiere una búsqueda exhaustiva de la bibliografía existente.

El estado del arte se puede ver como la aplicación de la metodología “survey” (a menudo traducida incorrectamente como “encuesta”) que a su vez utiliza “document-based resarcí” como técnica de generación de datos.

Twitter fue creada en 2006 por Jack Dorsey<sup>6</sup>, si bien no se hizo popular hasta marzo de 2007. Con más de 200 millones de usuarios en todo el mundo, esta red de microblogging da cabida cada día a cientos de millones de mensajes, condensados en un máximo de 280 caracteres. La brevedad, junto con la rapidez de emisión y recepción de estos mensajes, es una de las cualidades principales de esta red, a la que muchos le vaticinaron poco éxito en sus inicios y que, no obstante, ha alcanzado ya cotas de popularidad e impacto informativo hasta ahora desconocidos en otros medios y soportes de comunicación (Islas, 2010).

Seminarios, congresos y encuentros profesionales están debatiendo mucho en los últimos meses acerca del impacto de Twitter en la comunicación política., hablándose de sus potencialidades y también de sus riesgos. Se ha analizado el uso que Obama llevó a cabo de esta red de forma pionera en la campaña presidencial de 2008 (Harfoush, 2010; Hendricks & Denton 2010; Beas, 2011) y su posterior empleo en otras elecciones (Ammann, 2010; Williamson, 2010; Jungherr, 2010; Maarek, 2011; Holotescu, Gutu, Grosseck & Bran, 2011). A partir de este debate, he logrado condensar en forma de decálogo las principales razones por las que conviene que los políticos formen parte de la comunidad tuitera, si no lo hacen ya, además de analizar las potencialidades y beneficios que esta red puede aportar a la comunicación política y electoral, así como al desarrollo del proceso democrático. Veamos cuáles son estas razones:

- Twitter aporta una imagen de modernidad. Aunque parezca un argumento superficial, muchas veces los políticos se suman a avances tecnológicos por el mero hecho de no quedarse atrás con respecto a lo que se considera innovador en la sociedad o lo que ya se está haciendo en otros países, en otros sectores o por parte de la competencia. “Estamos porque hay que estar” es la frase que a menudo se escucha entre consultores y, sobre todo, entre políticos, y más aún en los momentos iniciales de cualquiera de estas nuevas herramientas, cuando todavía no se tiene una perspectiva precisa de lo que pueden aportar realmente. En este contexto juega un papel determinante el efecto contagio que se produce en muchos países del mundo con respecto a las técnicas y procedimientos de comunicación política que se ensayan con carácter pionero en Estados Unidos y que, poco después, acaban llegando al resto de naciones, un efecto que ha sido denominado académicamente como de ‘americanización’ de las campañas electorales (Mancini, 1995; Thurber & Nelson, 1995; Swanson & Mancini, 1996; Plasser, 2000; Bowler & Farrell, 2000). El tema es que, si algo se prueba con éxito en América, lo más probable es que, poco tiempo después, acabe aplicándose también en otros países y eso es lo que ha pasado con Twitter y, en general, con todo lo que incumbe a Internet.

---

<sup>6</sup> **Jack Patrick Dorsey** (San Luis, Misuri, 19 de noviembre de 1976) es un desarrollador de software y empresario estadounidense. Se le conoce por ser uno de los cofundadores de Twitter, así como el fundador y director ejecutivo de Square, una empresa de pagos mediante teléfonos móviles. El 5 de octubre de 2015 fue nombrado CEO de Twitter por el Consejo de Administración de la compañía.

En 2008 fue nombrado en la TR35 del Instituto Tecnológico de Massachusetts como uno de los 35 principales innovadores en el mundo menor de 35 años. En 2012, el periódico The Wall Street Journal le otorgó con el premio de “Innovador del Año” de tecnología. Actualmente vive en San Francisco, Estados Unidos de América.

- Twitter permite la comunicación con el ciudadano. Quizá la principal virtud de Twitter es que favorece la comunicación, el contacto directo y el diálogo con los ciudadanos. Tradicionalmente, la comunicación política consistía en que el partido o el candidato elaboraban sus mensajes y los difundían a través de los medios, como la prensa, la radio o la televisión. Este modelo se sustentaba en un tipo de comunicación básicamente unidireccional, vertical y rígido, donde el que hablaba era el político y la audiencia escuchaba o leía los mensajes lanzados y actuaba en consecuencia, pero sin tener apenas oportunidad de participar. Twitter, junto con otros medios de Internet, ha roto estas barreras al permitir que la comunicación sea bidireccional y que se produzca un *feedback* entre el político y los ciudadanos en otro tiempo inexistente o que sólo se daba en contadas ocasiones. En palabras de Orihuela, Twitter “más allá de la tecnología, son las personas que lo utilizan, los mensajes que se comparten, las conversaciones que se mantienen y las relaciones que se establecen entre sus usuarios” (2011: 33). En este contexto, se puede decir que las redes sociales y Twitter de una forma muy destacada, han venido a “democratizar” las vías de comunicación entre los ciudadanos y los políticos. En definitiva, Twitter supone pasar del monólogo al diálogo, de modo que en el diálogo las dos partes participantes tienen las mismas posibilidades de interactuar.
  
- Los usuarios de Twitter son “líderes de opinión” en sus entornos. Según análisis y encuestas publicadas en distintos países (Guerrero, M.V., 2017; Márquez, L., 2017), los *tuiteros* pueden ser considerados como líderes de opinión o “influentials” , ya que son gente informada, involucrada política y socialmente, preferentemente de perfil urbano y con un buen nivel formativo y estatus socioeconómico. Gente, en definitiva, que disfruta compartiendo información, que se organiza y que se activa cuando se trata de generar debate y que trata de ejercer influencia sobre sus círculos. Además, por encima de todo, esta gente concibe la participación en la política como algo continuo en el tiempo y no sólo con el hecho de introducir una papeleta en una urna el día de las elecciones (Franco Álvarez & García Martul, 2008). El perfil de este tipo de gente cada vez es más decisivo en los resultados finales de las elecciones y resulta muy interesante para los políticos, de modo que deben tener presente que al interactuar con estos usuarios no sólo pueden estar dirigiéndose a ellos, sino que la influencia puede propagarse como una bola de nieve también a quienes estén en sus círculos más cercanos. Sin embargo, hay líneas de investigación contradictorias respecto a esta afirmación, como las que protagoniza Dani Gayo Avello. En su artículo de 2011 *Limits of Electoral Predictions Using Twitter*, consigue probar que los analistas habían sobreestimado la victoria de Obama en las elecciones de 2008 hasta tal punto que éstos predijeron erróneamente la victoria incluso en Texas. Si bien los perfiles antes citados pueden ser decisivos en el devenir de las elecciones, Dani Gayo Avello y otros perfiles afines señalan que no todas las investigaciones consiguen predecir correctamente el resultado de unas elecciones.
  
- Twitter es una herramienta de comunicación interna y genera comunidad. Los partidos políticos pueden utilizar Twitter como vía de comunicación interna con sus propios militantes, simpatizantes y seguidores, que ven cómo con esta herramienta asumen más protagonismo en la acción política y que los partidos les necesitan y cuentan con ellos para el desarrollo de las campañas. Esto les motiva y les impulsa aún más en su labor, a la vez que refuerza su sentimiento de identificación con el partido y de involucrarse

activamente a la hora de conseguir sus objetivos. En este sentido, Twitter puede servir como canal para incentivar aportaciones económicas para las campañas políticas, tal y como ya hizo Obama en 2008. Además, también es especialmente útil para diseminar y difundir los mensajes clave durante la misma y generar opinión. Por último, Twitter sirve para mantener informados a los propios militantes y para que éstos, a su vez, actúen como correa de transmisión de los mensajes, aumentando así su difusión y eficacia. Este proceso, además, no debe ser exclusivamente unidireccional, ya que un candidato o un partido pueden también retuitear los comentarios de sus seguidores.

- Twitter es el medio más pegado a la actualidad. Twitter tiene un componente informativo que le confiere mucho valor en términos de estrategia política y electoral. Mucha gente reconoce enterarse de las noticias antes por Twitter que por cualquier otro medio de comunicación. De hecho, según Orihuela “todas las noticias de alcance de los últimos años han aparecido inicialmente en Twitter” (2011: 97). Cualquier ciudadano puede convertirse en narrador de hechos noticiosos de los que está siendo testigo a través de Twitter y, cada vez con más frecuencia, consiguen adelantarse a los propios medios de comunicación a la hora de dar las exclusivas. Twitter permite saltarse las barreras que imponen los productos informativos tradicionales, ya que ya no es necesario comprar un periódico o disponer de un receptor de radio o de televisión para conocer las noticias. De hecho, ni siquiera precisa un ordenador. El uso de esta red social se hace preferentemente a través de teléfonos móviles, que permiten consultar en cualquier momento y lugar. En este sentido, los españoles encabezan el acceso a redes sociales a través del móvil en Europa (*El País*, 12 de mayo de 2010).
- Twitter es una fuente de información para los periodistas y una vía para mejorar la relación con ellos. Twitter es un cauce para ganar presencia en los medios de comunicación tradicionales y generar noticias, factor que ayuda a tenerlo presente en las estrategias de comunicación política. Su gran repercusión en la actualidad hace que los medios presten mucha atención a lo que se dice en Twitter, incluso creando secciones nuevas para hacer este seguimiento, produciendo un efecto “altavoz” o “rebote”, ya que esos medios acaban publicando o emitiendo lo que sucede en la Red. Cada vez es más frecuente ver noticias basadas en las “declaraciones” de un político a través de su cuenta de Twitter. Esta dimensión pública de los tuits, en contraste con la privacidad de los mensajes que caracteriza a otras redes sociales como Facebook, hace que todo lo que se diga en Twitter se transmita como “si se hablara en voz alta para que todo el mundo lo oiga” (Congosto, Fernández & Moro, 2011: 12). En consecuencia, los políticos han de tener presente que lo que escriben en este medio puede ser publicado o emitido por los periodistas y que no debe haber conflictos entre su opinión personal y la función representativa que ostentan. Si en una rueda de prensa ante los medios deben medir sus declaraciones, lo mismo deben hacer en Twitter.

Este medio exige ser especialmente cuidadoso a la hora de enzarzarse en discusiones. Los insultos del presidente de Panamá, Ricardo Martinelli, a un *tuitero* que le había criticado acabaron siendo noticia en todo el mundo (*La Estrella*, 19 de julio de 2011). Por su parte, el candidato Stuart MacLenman del Partido Laborista de Escocia arruinó su carrera política al

publicar varios mensajes con lenguaje ofensivo en abril de 2010, hecho que dio origen a la expresión “suicidio en Twitter” (Orihuela, 2011: 259).

No se debe olvidar tampoco la posibilidad que brinda este medio de mejorar la relación con los periodistas. Esto puede incluso reportar beneficios, porque fomentar el diálogo fluido entre políticos y periodistas a través de la red les puede conferir una imagen de mayor transparencia y cercanía. Según Michael Dolley, asesor de campaña de David Cameron, “Twitter todavía alimenta sólo al villaje político: los periodistas siguen a los políticos, y los periodistas se siguen entre ellos” (ABC, 2 de junio de 2011).

- Twitter ayuda a los políticos a pensar y hablar en “titulares” y, por tanto, a ser mejores portavoces. Una de las reglas de oro que imperan en la comunicación política actual es la de transmitir los mensajes con brevedad. Las frases deben ser breves, concisas y directas y en esto Twitter es un excelente campo de pruebas para los políticos. La necesidad de hablar siguiendo la clásica estructura de sujeto, verbo y predicado, sin florituras dialécticas y sin rodeos innecesarios, evitando así lenguajes enrevesados o demasiado burocráticos, es algo que pone en práctica Twitter de forma bastante garantizada, ya que la limitación de los 280 caracteres por mensaje es una barrera infranqueable que obliga a los políticos a hablar de esta manera.

Twitter representa también un último cambio en cuanto a la forma de construir el mensaje político, que tiene que ver con el peso que se concede a las imágenes y a las palabras a la hora de dirigirse a los ciudadanos para pedirles el voto. Con frecuencia, los consultores políticos apelan a las imágenes, sobre todo por su poder emocional, como la vía más efectiva para llegar a la gente e influir en su comportamiento, entendiendo que éste toma sus decisiones más con el corazón que con la cabeza. Así, la apelación a los sentimientos se ha convertido en habitual sobre todo en el medio televisivo, en donde las imágenes han sustituido a las palabras como principal vehículo de la comunicación política (Sartori, 1998).

- Twitter humaniza a los políticos y aumenta la empatía hacia ellos. En la comunicación política actual se fomenta el conocimiento de los líderes no sólo como políticos, sino también como personas. Esto se ha traspasado a Twitter, donde los políticos más seguidos no se limitan a hablar sólo de política, sino que dejan entrever opiniones y aspectos de su vida personal que acaba siendo de interés de los *tuiteros* (Pearson & O’Connell, 2010). Por citar algunos ejemplos, la presidenta argentina Cristina Fernández de Kirchner dió a conocer a través de Twitter en julio de 2011 que iba a ser abuela, hecho que acabó siendo *trending topic*<sup>7</sup> mundial en muy pocas horas. El venezolano Hugo Chávez fue detallando también en Twitter el

---

<sup>7</sup> Un **trending topic** (tendencia, tema de tendencia o tema del momento, en español y TT de forma abreviada) es una de las palabras o frases más repetidas en un momento concreto en una red social.

Acuñaada por primera vez por Twitter, quien en la página de inicio muestra los diez más relevantes, pudiendo el usuario escoger el ámbito geográfico que prefiera, mundial o localizado, o personalizadas, en función además de a quién sigue el propio usuario. La gran repercusión que están teniendo en la prensa ha provocado que esta expresión sea utilizada también para denominar un tema de gran interés, esté o no siendo comentado en la red social.

desarrollo de su tratamiento contra el cáncer. Otros recomiendan libros o películas, hablan de lugares o monumentos que han visitado, comentan sus aficiones o hacen bromas y chistes, tal y como recomendaba el PSOE en el decálogo sobre el uso de Twitter difundido entre sus candidatos en las elecciones municipales y autonómicas de 2011 (*El País*, 25 de abril de 2011). Así, Twitter favorece que los políticos publiciten esta serie de detalles que contribuyen a humanizarles y a parecer más cercanos, algo especialmente importante en una época en la que están tan mal valorados por la opinión pública y en la que hay una creciente desafección hacia ellos.

- Twitter es un termómetro social. Si bien no puede sustituir de momento a las empresas demoscópicas, puede ofrecer un conocimiento bastante certero de lo que piensa la gente sobre determinados temas, convirtiéndose así en un “gigantesco estudio de mercado en tiempo real y a escala global” (Orihuela, 2011: 36). En este sentido, se está hablando mucho en los últimos meses del poder “predictor” de Twitter con respecto al comportamiento de la opinión pública. Más concretamente, en el ámbito de las campañas electorales se han llevado a cabo diversos estudios empíricos que afirman que la circulación de mensajes y menciones en Twitter puede predecir los resultados de unas elecciones con una fiabilidad similar a la de los sondeos tradicionales, si bien esta línea de trabajo es aún muy incipiente y conviene esperar a futuros análisis para corroborar este supuesto poder (Livne, Simmons, Adar & Adamic, 2011; Tumasjan, Sprenger, Sandner & Welpe, 2010; Gayo-Avello, Metaxas & Mustafaraj, 2011). Otros autores, en cambio, sostienen todo lo contrario (Gayo-Avello, 2012). La “voz de la calle” puede reflejarse en Twitter en forma de *hashtags* o *trending topics* (TT), que pueden mostrar a los partidos hacia dónde se dirige la opinión pública (Mañana, 2011). Un ejemplo representativo del altavoz mediático que supone Twitter de ese sentir popular lo encontramos en las revueltas de Túnez, Egipto, Siria o Libia, en las protestas contra Ahmadinejad tras las elecciones iraníes de 2009 o contra Putin en las rusas de 2011, en el movimiento del 15M en España, en las revueltas de la comunidad universitaria de Chile, los “indignados” de Israel o en los disturbios de Londres en agosto de 2011 (Chatillard, 2011).

Por otra parte, Twitter permite a los políticos ser más cercanos a las demandas reales de la gente y escuchar realmente qué es lo que piensan de ellos y de su trabajo. De hecho, cada vez es más frecuente ver a políticos que *tuitean* propuestas o planes que tienen pensado ejecutar para que los ciudadanos les envíen sus puntos de vista y sugerencias, creando una gran lluvia de ideas y poder así contar con la participación de la opinión pública en la toma de decisiones.

- Puede ser determinante en el triunfo de unas elecciones. Estudios realizados hasta la fecha han comprobado que los efectos de Internet sobre el voto no están todavía lo bastante constatados, por lo que se precisan más análisis empíricos en el futuro para terminar de probar estos efectos. Se sospecha que puede llegar el día, no muy lejano en el tiempo, en el que la influencia de las redes sociales se convierta en decisiva para ganar unas elecciones. Parafraseando a Capdevilla, “de su impacto más profundo en política sólo estamos ante la punta del iceberg” (2009: 35). Entre tanto, si la victoria o la derrota se dirime en unos pocos puntos porcentuales, cualquier esfuerzo que se haga en la campaña resulta decisivo. Si con las redes sociales se consigue movilizar a una parte del electorado para que acuda a las urnas a prestar su voto, sin duda este hecho será muy importante, teniendo en cuenta además que la participación de los políticos en Internet



es mucho más barata que las costosas campañas publicitarias en televisión.

A modo de conclusión y para cerrar este último punto, si bien Twitter no es garantía fiable para ganar unas elecciones y podría cambiar o incluso desaparecer con el tiempo, sus aportaciones han sido más que suficientes para justificar su empleo en el ámbito de la comunicación política y electoral y para ganarse un espacio muy destacado en la historia de esta disciplina.

### **2.1.1 Actividad de los líderes políticos en Twitter**

Twitter se ha incorporado como un medio imprescindible en las estrategias comunicativas de los actores políticos, siendo su empleo especialmente intenso en los períodos de campaña electoral. Desde las elecciones norteamericanas de 2008, marcadas por el innovador uso de las redes sociales por parte de Barack Obama, la influencia de estas plataformas digitales sobre la comunicación política no ha parado de crecer en todas las democracias avanzadas (Bimber, 2014).

En esta sección, a modo de ejemplo, me planteo como objetivo analizar cuáles son las principales funciones atribuidas a Twitter por parte de los líderes políticos en el contexto electoral. Pretendo averiguar qué hacen los políticos en esta plataforma y qué implicaciones tiene lo que hacen para la comunicación política. Para ello, he tomado en consideración un único caso. Se trata de la campaña de las elecciones de 2016 en España, en la que se tomó una muestra formada por 9.042 tweets de los cuatro partidos más votados y sus respectivos candidatos que, en conjunto, representaban el 89,95 % de los votos. La metodología que se empleó para el estudio de esta campaña se basó en el análisis de contenido cuantitativo. Los resultados obtenidos son aplicables más allá del entorno español y conviene señalar, antes de enumerar las diferentes características de la utilización de Twitter, que todas ellas se ven condicionadas por tres factores determinantes:

- La trayectoria histórica del partido;
- La posición que ocupa el partido en el eje gobierno-oposición;
- La dimensión individual o colectiva del actor político que emite los mensajes.

En primer lugar, los resultados obtenidos dan a entender que la función predominante atribuida a Twitter por la totalidad de líderes políticos (con la excepción de Pablo Iglesias en el caso español) es la difusión de las propuestas políticas que integran su programa electoral. La posibilidad de dirigirse directamente a la ciudadanía, sin la intermediación de los periodistas y los medios convencionales, impulsando una automediación, está en la base del predominio de este tipo de uso de Twitter. En este sentido, el uso que se hace de Twitter para difundir las propuestas políticas se basa en una doble estrategia:

- Publicación de mensajes con las medidas planteadas en un tono positivo. En el caso de los partidos vinculados a la nueva política o de nueva hornada, esto se suplementa con la distribución de tuits que ensalzan sus valores y su ideología para remarcar la novedad de su oferta política, basada en el cambio y la regeneración democrática.
- Empleo de la crítica al adversario, como complemento de la estrategia anterior. Esta crítica se basa en una personalización negativa, ya que se focalizan los reproches en los líderes de los partidos rivales. Esta dinámica,

además, se ve condicionada por la posición del partido y por la dimensión individual o colectiva del actor político. Así, las formaciones de la oposición dan un uso más intenso de reprobación de los competidores que el partido que ostenta el Gobierno. Igualmente, las acusaciones a los contrincantes se concentran más en los perfiles de los partidos que en los líderes políticos.

Otra función destacada es la participación y la movilización, cuya actividad depende, en gran medida, de la trayectoria histórica de los líderes políticos. En contra de lo que cabría esperar, los partidos establecidos hacen un mayor uso que los emergentes. Además, mientras los primeros piden explícitamente el voto en sus tuits, apelando a la lógica de la persuasión, los segundos animan a los ciudadanos a tomar parte activa en la campaña, potenciando la lógica de la movilización. Otro factor que también aparece como determinante es la dimensión individual o colectiva, ya que los líderes políticos emplean más esta función que los partidos.

Los resultados demuestran un bajo nivel de personalización atribuido a Twitter por parte de los actores políticos. Nos estamos refiriendo a compartir cuestiones de carácter personal, tanto públicas como de la intimidad más estricta. Se podría concluir que ni partidos políticos ni candidatos son proclives a compartir en esta red social imágenes en las que puedan mostrar facetas íntimas. Los líderes políticos sí que lo hacen en alguna ocasión puntual, pero no deja de tratarse de episodios anecdóticos (López-Rubadán et al., 2016). Esta circunstancia no significa que no recurran a otras vías para mostrar su lado más humano, pero en dosis muy limitadas. Algunos candidatos aprovechan su timeline, por ejemplo, para compartir con sus seguidores sus preferencias culturales y recomendar libros, obras de teatro o películas. En definitiva, cuestiones personales, de acuerdo, pero que no les comprometen en absoluto. Se observa que los partidos de reciente aparición recurren más a la personalización en Twitter que los partidos establecidos. En el caso español, este patrón lo rompe de nuevo el candidato de *Podemos*, que apuesta por dar a una parte de sus tuits una función de revelación de aspectos vinculados a su esfera privada, como gustos, aficiones, sentimientos o emociones. Con ello, intenta humanizar su figura y acercarse a la ciudadanía (Enli; Skogerbø, 2013). También utiliza el recurso del humor mediante bromas y memes para potenciar esa conexión. Podría decirse que la función de personalización en Twitter adopta las características de la política pop (Mazzoleni, Slardini, 2009). Ello implica que algunos líderes políticos incorporan a sus repertorios comunicativos estilos, narrativas y puestas en escena propias del mundo del espectáculo y el entretenimiento. Estos resultados indican que este fenómeno, más propio de la televisión, se extiende también al entorno de las redes sociales, saliendo de su hábitat natural. Si bien en el caso español sólo afecta a Pablo Iglesias, esta tendencia podría ser precursora y generar una notable influencia en el futuro en la construcción de la imagen del líder político en los medios sociales. Se trata de un aspecto que se tendrá que comprobar a base de nuevos estudios.

Por último, se observa una tendencia incipiente hacia la hibridación entre los nuevos medios digitales y los convencionales en el empleo de esta plataforma. Así, mediante sus perfiles en Twitter los líderes políticos difunden su participación en programas televisivos, como entrevistas o debates, o sus apariciones en las noticias para maximizar su impacto. Con ello, articulan una estrategia comunicativa que combina y mezcla de forma complementaria los viejos y los nuevos medios. Esta mayor focalización en la hibridación entre viejos y nuevos medios se da en partidos emergentes o de menor trayectoria histórica y se puede explicar por varios motivos:

- Lograr un mayor impacto y proyección en sus apariciones en los medios convencionales, especialmente la televisión.

- Conseguir un elevado número de retuits y “me gusta” que garanticen un doble efecto: una mejor difusión del mensaje y, gracias a ello, convertirse en sujetos de interés para los medios al acumular un elevado volumen de audiencia.
- Porque los partidos emergentes tienen en las redes sociales su hábitat natural, aunque son conscientes de la importancia de los medios convencionales en la sociedad actual y de la necesidad de conectar viejos y nuevos medios (Casero-Ripollés; Feenstra; Tormey, 2016).
- Aprovecharse de la capacidad de legitimación de los medios tradicionales y aparecer ante la ciudadanía como opciones políticas válidas, aceptables y fidedignas.

## 2.2 Temas abordados por los políticos en Twitter

El uso generalizado de Twitter en la comunicación política ha traído consigo una variedad temática destacada en los tuits. Se abordan cuestiones de política, problemas de orden político, asuntos propios de la campaña electoral, temas vinculados a la vida íntima de los candidatos, etc. Si bien las posibilidades son enormes, se observa una cierta preferencia por ciertos asuntos en la mayoría de los casos. En un análisis realizado sobre una muestra integrada por 2.310 tuits publicados por los líderes de *Podemos* (Pablo Iglesias, España), el *Movimiento 5 Stelle* (Beppe Grillo, Italia), el *Frente Nacional*, (Marine Le Pen, Francia) y el *UKIP* (Nigel Farage y Paul Nuttall, Reino Unido) durante tres periodos temporales aleatorios (Alonso-Muñoz y Casero-Ripollés, 2018), se concluye que los líderes conservadores otorgan gran importancia en sus mensajes a los asuntos relacionados con la inmigración, el terrorismo y los asuntos exteriores, mientras que los líderes progresistas se decantan más por comentar asuntos relacionados con las políticas sociales, la corrupción o la regeneración democrática.

En un contexto más local, un análisis de la firma PiperLab para Expansión sobre un total de 287.076 tuits entre el 30 de septiembre y el 29 de octubre de 2019 para la campaña electoral del 10 N, señala que la mayoría de los comentarios son relativos a la campaña electoral, si bien el peso de este tema varía mucho entre los diferentes líderes políticos españoles. Véase, a modo de resumen, el siguiente gráfico:

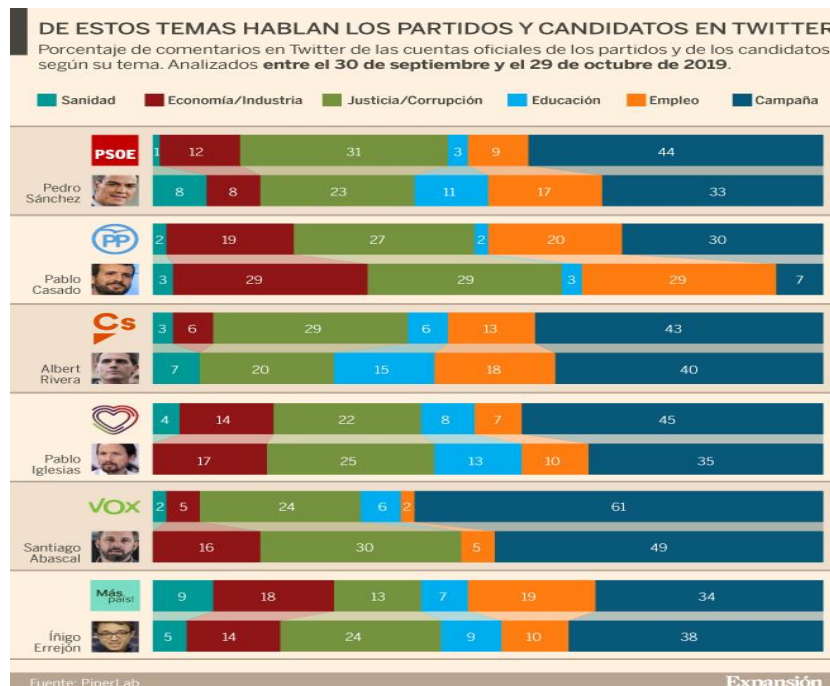


Figura 3. Temas hablados en Twitter . Fuente: PiperLab para Expansión

Vox es la formación que dedica más cobertura a este tema (el 61% de sus comentarios), seguido por Unidas Podemos (un 45%) y el PSOE (un 44%). El PP, en cambio, es el que menos espacio dedica (un 30%), al igual que su líder, Pablo Casado (un 7%). El segundo de los grandes asuntos más mencionados es la justicia y corrupción, salvo para Más País (que es el empleo). En cuanto al empleo, PP y Más País son los que le conceden más importancia, y a la economía, Ciudadanos y también el partido de Íñigo Errejón. A la sanidad el partido que le otorga más peso en su conversación es Unidas Podemos. Por su parte, Pablo Casado apenas da importancia a la campaña y centra sus mensajes en tres grandes puntos: la economía, la justicia y el empleo. Centrándonos en temas más concretos, la crisis de Cataluña gana presencia ya que el análisis se llevó a cabo durante la sentencia del “procés”. Sobre este asunto, Albert Rivera ha sido el gran protagonista de las conversaciones en Twitter. Además, mientras que Íñigo Errejón es el que más referencia hace al feminismo, Santiago Abascal es el que más habla de los inmigrantes, refiriéndose sobre todo a la eliminación de la inmigración ilegal. En cuanto a la exhumación del dictador Francisco Franco, tanto Pedro Sánchez como Íñigo Errejón son los que más se manifestaron a su favor.

Para terminar esta sección, en un análisis del uso que hizo Donald Trump de Twitter en el último mes de su campaña electoral (08/10/2016 – 08/11/2016) (Fernández, 2018) a partir de 526 tuits de la cuenta @realDonaldTrump, se concluye que Donald Trump hizo comunicación política en Twitter, que sus mensajes fueron básicamente negativos y centrados en ataques y que sus mensajes no aportaron veracidad. El hecho más destacado de sus tuits es que se componen, en gran medida, de ataques (342 tuits), siendo el mayor número de ellos dirigido a su entonces contrincante Hillary Clinton, que recibe 172 ataques. En otras ocasiones también ataca en sus mensajes a otros actores: a la prensa, a Obama o a los poderes establecidos asociados a la capital de Estados Unidos, a Washington. Esta suma de ataques supone el 69,4% de sus mensajes identificados, que convierte en la temática principal en la que basa la estrategia de contenidos en Twitter durante ese citado mes. En cuanto a Hillary Clinton, la relaciona con la corrupción, con actividades deshonestas y tilda de fracaso el papel de la candidata demócrata como Secretaria de Estado. También hace hincapié en el supuesto uso del servicio privado que hizo Hillary Clinton para la gestión de sus emails. Las siguientes figuras muestran ejemplos de tuits de crítica a Hillary Clinton:



Figura 4. Ataques a Hillary Clinton  
Fuente: <http://hdl.handle.net/10609/83110>

Llama también la atención que no dedique ni un solo mensaje a política exterior, tercer tema de interés para los estadounidenses, así como a educación. En cuanto al

sistema de salud, tema determinante para el voto, sus tuits se centran en la crítica al sistema conocido como *Obamacare*,<sup>8</sup> sin aportar nuevas alternativas.



Figura 5. Crítica al *Obamacare*  
Fuente: <http://hdl.handle.net/10609/83110>

En cuanto a la economía, publicó tuits en los que se aludía a una “desastrosa” situación económica y a la pérdida de empleos, de modo que votar a Hillary Clinton significaba continuar con ese desastre. Véase la siguiente muestra:

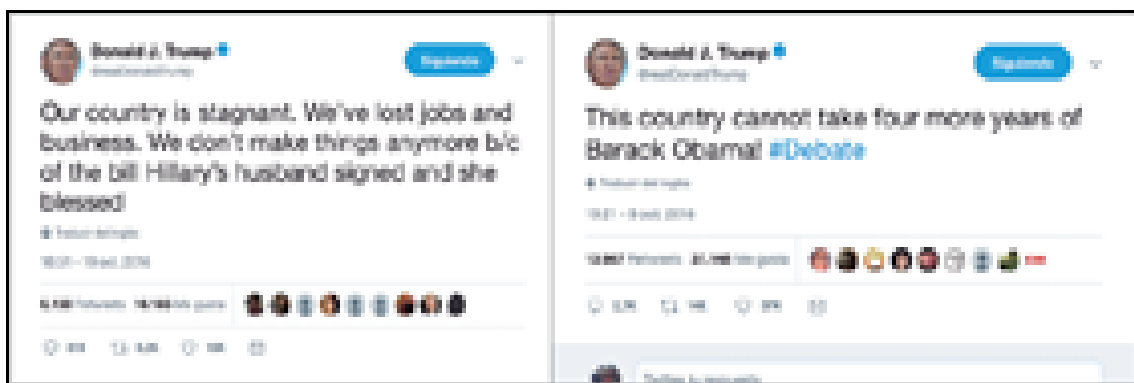


Figura 6. Situación económica  
Fuente: <http://hdl.handle.net/10609/83110>

No obstante, el análisis del contexto económico mostraba que el país estaba en situación de pleno empleo y en continuo crecimiento, con los niveles de PIB estatal y de renta per cápita más altos de años anteriores. Por tanto, son mensajes sobre economía que no aportan veracidad y que aprovechan el tema para transmitir una negatividad que no se corresponde con la realidad.

Por último, en sus mensajes más positivos da a conocer su eslogan *Make America Great Again*, anima a los ciudadanos al voto, agradece la participación en mítines o se pone al lado de la gente diciendo que está con ellos para hacer un país mejor.

*Get out and vote! I am your voice and I Will fight for you!  
We Will make America great again! (25/10/2016)*

<sup>8</sup> *ObamaCare* (Ley de Cuidados de la Salud Asequibles o *Affordable Care Act*) es una reforma de salud en E.E.U.U. que extiende y mejora el acceso a los cuidados de salud y disminuye el gasto mediante regulaciones e impuestos.

## **2.3 Big Data en redes sociales y Twitter**

En esta sección, intentaré concretar de qué modo otras investigaciones (o soluciones) han desarrollado hasta el momento la resolución de la problemática en la que estoy trabajando, esto es, el seguimiento de campañas electorales en Twitter. Tras unos apartados precedentes mucho más genéricos, en el presente pretendo añadir un plus de concreción. En particular, voy a centrarme en el estado actual del Big Data en estudios de redes sociales bajo un contexto político. Las técnicas de Big Data permiten encontrar, recoger y analizar elementos deliberativos propios de la discusión política en redes sociales y, en especial, en Twitter. Estos elementos pueden servir como recursos para acciones y decisiones públicas, en ocasiones más allá de las meramente electorales.

### **2.3.1 Minería de datos**

Como técnica de exploración, la Minería de Datos combina técnicas semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica para la obtención de información que no está representada explícitamente en los datos. La aplicación de las tecnologías de minería para extraer valor del caudal continuo de millones de tuits generados por los usuarios en Twitter está abriendo nuevas líneas de investigación, entre ellas la predicción de resultados electorales. Actualmente existen dos corrientes de opinión científica sobre la capacidad de predecir resultados electorales analizando la participación ciudadana en Twitter. Por una parte, hay una opinión favorable, cuyo caso más destacado es el estudio realizado por Tumasjan (Tumasjan, 2010) y en el otro extremo hay una corriente crítica desde los trabajos de Panagiotis (Panagiotis, 2011), Gayo-Avello (Gayo-Avello, 2011) y Jungherr (Jungherr, 2011). A día de hoy no hay un método validado y reproducido en distintos escenarios que permita predecir elecciones, si bien el análisis de elecciones puede contribuir a conocer mejor el entorno social en las campañas electorales y a aportar nueva luz hacia nuevos procesos de predicción de resultados.

Algún que otro trabajo de investigación ha aplicado técnicas de minería de datos para obtener clasificadores que permitan identificar automáticamente la categoría (Positiva, Negativa o Neutral) de las opiniones públicas manifestadas en Twitter en relación a temas políticos (Cumbicus, 2017). El objetivo principal de esa investigación fue utilizar minería de datos para verificar los patrones de los tweets recolectados y poder clasificarlos en sus respectivas clases.

Por su parte, Congosto M. L. y Aragón, P. (Congosto, M.L. ; Aragón, P., 2012) analizaron mediante técnicas de minería aspectos de la campaña de las elecciones generales del 20N en Twitter como la participación de los ciudadanos, el estado emocional de los políticos, la popularidad de los candidatos, las relaciones entre ellos y la predicción de resultados electorales. Como métodos de predicción, usaron tanto medidas basadas en el cómputo de menciones como medidas basadas en la polaridad política, obteniéndose mejores resultados con las medidas basadas en el cómputo de menciones.

### **2.3.2 Clasificación automática de textos**

En otro trabajo (González, 2015), se desarrollan soluciones específicas de clasificación automática de textos<sup>9</sup> en el ámbito de la política (se analizan mensajes de

---

<sup>9</sup> La clasificación automática de textos puede definirse como la acción ejecutada por un sistema artificial sobre un conjunto de elementos para ordenarlos en clases o categorías. Los elementos a clasificar pueden ser de cualquier tipo (TXT, Word, PDF, etc.).

Twitter de los principales partidos políticos y sus principales representantes, en plena campaña electoral). Adicionalmente, se predice el valor de abstencionismo electoral con un error diferencial respecto al valor real, demostrando la confiabilidad de la metodología propuesta. La clasificación automática de textos es un área de investigación que ha cobrado gran importancia en el estado del arte de los últimos años, debido a los grandes volúmenes de textos digitales que se almacenan en bases de datos tanto empresariales como en páginas web, comentarios en foros y redes sociales. Una de sus principales virtudes es el aumento de la rapidez con que se pueden procesar las opiniones políticas, lo que conlleva una reducción consustancial de costes frente a la tradicional clasificación manual.

Conviene citar que el *topic modeling* es una de las herramientas informáticas más usuales de la minería de textos. Esta técnica computacional usa modelos estadísticos de distribución (Latent Semantic Indexing (LSI) y Latent Dirichlet Allocation (LDA)) para reconstruir temas (*topics*) agrupando palabras similares que tienden a “coaparecer”. En su aplicación se utilizan lenguajes de código abierto como R o Python o software comerciales, siendo el Mallet el más utilizado actualmente. Otras de las técnicas de minería de texto que se aplican usualmente en Twitter son el *sentiment analysis* y el *opinion mining*. Lo que se busca en estas subdisciplinas es analizar el vocabulario de un texto con el fin de determinar sus cargas emocionales, haciendo uso de un ordenador que a través de *lexicones* procese, reconozca y evalúe dichos sentimientos (Leetaru, 2012). Mientras que el *opinion mining* intenta detectar polaridad política, el *sentiment analysis* se ocupa del reconocimiento de emociones, pero debido a que la identificación de sentimientos es a menudo explotada para la detección de la polaridad, los dos campos se suelen utilizar como sinónimos (Cambria, Schuller, Liu, Wang, Havasi, 2013). Aunque existen aplicaciones comerciales para hacer consultas (*MeaningCloud, Semantria, WordStat, etc.*) la mayoría utiliza los lenguajes colaborativos *Python* y *R*. En este tipo de técnicas es cada vez habitual el uso del aprendizaje automático o *machine learning*.<sup>10</sup> En ejemplos de *sentiment analysis* y *opinion mining* sobre política, Gruzd y Roy (2014) comprobaron que el 40% de los mensajes publicados por personas de izquierda tienden a ser hostiles o negativos.. En un nivel que combinan *data mining, text mining* y teoría de grafos aplicados a *social networks* se inscriben trabajos que buscan en Twitter no sólo identificar sentimientos u opiniones sino también comunidades y vectores de difusión de información, tratando de establecer una relación con los movimientos sociales. Puestos los ojos en la “conversación” de Twitter sobre determinados temas de agenda pública, se combinan

---

<sup>10</sup> Se conoce como *machine learning* al subcampo de las ciencias de la computación y rama de la inteligencia artificial, cuyo objetivo es hacer posible que las computadoras aprendan. Esta técnica utiliza algoritmos que razonan de manera similar al cerebro humano. Se trata de elaborar programas capaces de generalizar comportamientos a partir de la información suministrada en forma de ejemplos. El aprendizaje automático cuenta con una amplia gama de aplicaciones: motores de búsqueda, diagnósticos médicos, detección de fraude en el uso de tarjetas de crédito, análisis del mercado de valores, clasificación de secuencias de ADN, reconocimiento del habla y del lenguaje escrito, juegos y robótica, entre otros. Geoffrey Hinton acuñó el término *deep learning* a las nuevas arquitecturas de Redes Neuronales profundas que permiten a las computadoras “ver” y distinguir objetos y texto en imágenes y videos. Sobre los “peligros” que implica el *machine learning*, más de tres mil investigadores de IA y robótica, respaldados por Stephen Hawking, Elon Musk y Steve Wozniak, entre otros, firmaron en 2015 una carta abierta en la que se advertía de las amenazas que implica desarrollar armas autónomas que seleccionan y atacan objetivos sin intervención humana. Muchas universidades y plataformas ofrecen curso gratuitos sobre *machine learning* y sus aplicaciones, como el Knight Center for Journalism in the Americas de la Universidad de Austin, EEUU: [https://journalismcourses.org/inteligencia\\_artificial.html](https://journalismcourses.org/inteligencia_artificial.html)

elementos de *topic model*, *sentiment analysis* y teoría de grafos para identificar comunidades políticas.

### 2.3.3 Tratamiento de los “issues” a partir de los “hashtags”<sup>11</sup>

En Twitter, los *#hashtags* son muy utilizados por los usuarios con el fin de influir en los temas de conversación que acaecen en las redes. Dentro de la plataforma, tienen la ventaja de alentar una doble comunicación, ya que el tweet es leído tanto por los seguidores como por todos aquellos que están siguiendo un determinado *#hashtag*. Poseen la capacidad de instalar temas en la agenda de discusión de la plataforma, por ello cierta parte de la literatura sostiene que Twitter es, para el activismo digital, un medio democrático. Esta capacidad es aprovechada por el resto de los actores, entre los que se destacan los principales medios de comunicación y los institucionalizados (políticos e industrias).

En un artículo que analiza los *issues* del PSOE y podemos en Twitter durante la campaña electoral de mayo de 2015 en España (Machado y Capdevila, 2017), el punto de mira son los hashtags producidos por ambas cuentas para comparar el tratamiento realizado por partidos tradicionales y noveles. Para el análisis, se utilizan sendos programas informáticos: uno para la captura y almacenamiento de tuits y otro para procesar cuantitativamente los datos.

Por otra parte, otro artículo estudia el comportamiento de las etiquetas de campaña electoral en Twitter correspondientes a la segunda vuelta de las elecciones presidenciales en Ecuador del año 2017 (Puyosa, 2017). En el estudio, se analizan 10 etiquetas de ataque a los adversarios que fueron *trending topics*, capturando y analizando los datos con NodeXL, una aplicación para visualización y análisis de redes sociales con Microsoft Excel similar a otras aplicaciones como Pajek, UCInet y Gephi. El análisis fue capaz de verificar el rol central en el posicionamiento de las etiquetas que tuvieron las cuentas automatizadas.

Siguiendo en este marco de las investigaciones que abordan la relación entre Twitter y la Política, hay más estudios que indagan en las posibilidades que el seguimiento de hashtags ofrece a la hora de analizar la construcción de narrativa colectiva y comunidades en el proceso de transmisión de información en las redes. Así, Perez Lagos y Cabral Salles (2018) indagan en cómo se visibiliza a través de Twitter la vinculación del candidato presidencial Sebastián Piñera con el caso de los ‘Panama Papers’, considerado como un escándalo político. La base de estudio fue la construcción de un corpus de dos mil tuits que mencionaran algunas de las palabras clave: “Panama Papers”, “Papeles de Panamá”, “paraísos fiscales”, “Sebastián Piñera” y “Piñera”. A partir de estos datos se basaron en el análisis del discurso para examinar los tuits, encontrando tres tipos de actos recurrentes: desacreditar, enfrentar y tomar posición. Por su parte, Aguerre y Tarullo (2016) se proponen comparar los *trendingtopics* (TT) de Twitter en Argentina con las noticias que encabezan los rankings de lectura en los diarios en Argentina y los tuits más retuiteados de las cuentas oficiales de medios nacionales, en un trabajo que nos lleva a interrogarnos si los usuarios de Twitter utilizan las posibilidades que las plataformas digitales ofrecen respecto a una mayor participación ciudadana, en relación a asuntos de interés público y político. A partir de la intersección entre las noticias más leídas y los TT más importantes en Argentina, encuentran que un pequeño porcentaje de los TT están vinculados a la política, aunque estos temas son los más retuiteados, y los escándalos de corrupción son los mayores generadores de conversaciones en temas de política.

---

<sup>11</sup> *Hashtag* (#). Cadena de caracteres formada por una o varias palabras concatenadas y precedidas por un numeral.



Estos dos últimos estudios muestran el alcance que el seguimiento de hashtags tiene para analizar la construcción de narrativas colectivas y hacer visible la construcción de comunidades en el proceso de transmisión de información en las redes, en particular cuando se trata de escándalos políticos, que son los mayores generadores de conversaciones en Twitter. En este sentido, me resulta relevante señalar las constantes apariciones de hashtags vinculados a hechos de corrupción, y, por ende, la necesidad de seguir explorando en futuras investigaciones este método.

Podemos encontrar también estudios realizados a partir de las menciones a determinados sucesos en Twitter. En un artículo de Yeste y Franch (2018) que analiza el tratamiento de la prensa desde la cuenta de Twitter de Donald Trump del 20 de enero al 20 de julio de 2017, se estudia las menciones a los medios y el tono de dichas menciones, así como la presencia y el uso de las palabras clave relacionadas con los medios. Además, se analiza el tratamiento de Trump hacia la prensa en relación con el control político y la función del periodismo como perro guardián (*whatchdog*) de la democracia. Como conclusiones, se constata que Trump, de manera muy simple y directa, elogia a los medios afines y critica a los que le resultan hostiles, a los que muy a menudo califica de *fake news*. Por otra parte, un artículo de Azorín-Richarte, Orduna-Malea y Ontalba-Ruipérez (2015) identifica y describe, mediante un análisis cibernético de menciones de usuario, la intensidad y evolución de las relaciones establecidas entre compañías de un determinado sector industrial (tecnología) a través de sus correspondientes cuentas corporativas de Twitter, con el propósito de comprobar el valor que un análisis métrico de estas características puede tener a la hora de determinar la conectividad entre dichas compañías. Para ello contabiliza todo tipo de menciones entre las cuentas de Twitter de una muestra de 50 compañías internacionales durante un periodo de seis meses. Los resultados indican que el grado de interacción entre las 50 compañías internacionales es débil y el artículo concluye que las cuentas corporativas de las empresas tecnológicas en Twitter no son suficientes para analizar desde un punto de vista métrico la conectividad web establecida entre éstas, aunque sí útiles para conocer las políticas de comunicación oficiales entre ellas. Como último ejemplo, Tumasjan, Sprenger, Sandner y Welpe (2011) analizan en su trabajo las menciones a partidos y a políticos alemanes en un contexto electoral. Usando el software de análisis de texto LIWC, se realiza un análisis de contenido de más de 100.000 tuits que contienen menciones a un partido o a un político. Los resultados delatan a Twitter como plataforma de deliberación política, que el número de mensajes que mencionan un partido refleja el resultado de las elecciones y que el análisis del sentimiento político de los tuits demuestra su estrecha correspondencia con las posiciones de los partidos políticos y sus líderes.

### **2.3.4 Metodologías observadas**

Se observa, en la mayoría de los casos, la utilización de una metodología cuantitativa para extraer los datos cuantificables de las redes sociales de los políticos, obteniéndose así datos de valor como el número de seguidores, la cantidad diaria de publicaciones, los temas tratados en la campaña electoral, el énfasis que dan a sus publicaciones y la interacción entre el votante y el político. Las principales características del análisis de contenido cuantitativo son la objetividad, la sistematicidad y la cuantificación de los contenidos manifiestos de la comunicación (Cea, 2011). Es la técnica más adecuada para analizar contenidos textuales con un considerable volumen de información. La obtención de estos datos cuantificables suele realizarse a través de algún instrumento de análisis. Este análisis de contenido cuantitativo puede ir acompañado de un análisis métrico de los perfiles oficiales de los partidos políticos y candidatos. También pueden darse algunos casos en los que se realice un trabajo cuantitativo y cualitativo sobre los contenidos de los tuits (Machado y Capdevila, 2017). En estos casos, de lo cuantitativo se puede deducir el tipo de uso

que se está dando de Twitter, esto es, qué tipo de mensajes contenían los tuits; y de lo cualitativo, aspectos tan relevantes como la identificación de los *issues* o si detrás de cada perfil estaban los candidatos o personas de sus respectivos equipos.

En cuanto a la gestión de grandes volúmenes de datos (Big Data), sean estructurados o no estructurados, se aplican distintas técnicas computacionales para procesarlos en información. La generación de datos no contempla en sí misma conocimiento alguno. Los datos son la “materia prima bruta”, se convierten en información y luego en conocimiento a través del análisis (Molina, 2002). Para procesar los grandes volúmenes de datos que provienen de Twitter, existen un par de maneras: con aplicaciones comerciales o con software libre. Lo primero que hay que hacer es registrar una aplicación personal (API) en Twitter Developer, que nos habilita a hacer gathering (recolección) de datos en la red social a través de sus APIs. Twitter ofrece también una enorme cantidad de datos semiestructurados en formato JSON (javascript object notation), que pueden obtenerse tanto en flujo directo (streaming) como del archivo histórico a partir del uso de las APIs (Streaming y REST) que la red social ofrece a sus usuarios de forma gratuita. Por último, se observan dos formas de acercarse a los datos que ofrece Twitter: una más “impresionista”, que recoge manualmente una serie de tuits y sigue el movimiento de sus usuarios, y otra que utiliza el análisis de los grandes volúmenes de datos que ofrece la red. Al ser un campo en incipiente formación, que sobre todo se nutre de los avances entrecruzados de la investigación académica, periodística y corporativa, existe todavía a día de hoy una gran confusión alrededor de los conceptos fundamentales de análisis computacional de datos.

Para terminar esta sección, conviene citar que dada la magnitud y complejidad de los desarrollos, métodos y técnicas que actualmente se incluyen bajo la denominación Big Data, podemos dividir las tendencias actuales en Tecnologías Big Data, Herramientas Big Data y Bases de Datos a Gran Escala. Dentro del primer grupo se incluyen el análisis predictivo, las bases de datos NoSQL, la virtualización de datos y el *machine learning* o aprendizaje automático, entre otras. Dentro del segundo grupo, las herramientas con mayor capacidad en la actualidad son Apache Hadoop, Apache Spark y otras herramientas como Apache Avro, Hbase, Apache Hive o ElasticSearch. En cuanto a las Bases de Datos a Gran Escala, podemos citar los Sistemas de gestión de Bases de Datos tradicionales, los sistemas Data Warehouse y los Sistemas de Bases de Datos NoSQL.

### **3. Diseño e implementación**

En este tercer apartado de la memoria, en estrecha correspondencia con la PEC3 de la asignatura Trabajo Final de Máster (TFM), se diseña y se desarrolla el proyecto que se ha especificado en apartados anteriores. Dividiremos este apartado en dos secciones bien diferenciadas: radiografía de tweets y análisis de contenidos a partir de los tweets capturados. Esta última sección detallará el proceso de limpieza de datos realizado a los tuits, citará las palabras más frecuentes encontradas, representará las nubes de palabras y realizará el análisis de sentimientos, así como un análisis de lo que dicen los usuarios que más mencionan a los principales candidatos, para terminar descubriendo la estrategia seguida en Twitter por cada uno de estos candidatos. Se llevará a cabo un seguimiento de tweets correspondientes a las elecciones presidenciales norteamericanas del 3 de noviembre de 2020. Unas elecciones en las que Donald Trump y Joe Biden compiten por la presidencia del país y que se producen en unas circunstancias sin precedentes. Su resultado tendrá un impacto más notable que otras veces y afectará de manera especial a empresas y emprendedores de otros países con intereses en el mercado americano. Configuran el escenario dos candidatos diferentes y con visiones dispares y una nación inmersa en la mayor

recesión desde la Gran Depresión de hace 90 años, agravada con una crisis sanitaria y social asimétrica, que está afectando a todo el mundo, el Covid-19.

### 3.1 Radiografía de tweets

En esta investigación, se contemplan un total de 41.863 tweets. Si bien se programó la función *search\_tweets()* de R para obtener un número mayor, valores del parámetro *n* superiores a 18.000 dieron lugar a problemas de interrupción en la descarga que abortaron dicha operación. El entorno RStudio mostraba el mensaje *Warning message: Rate limit exceeded.* En cualquier caso, la cantidad obtenida es aceptable para el estudio que se ha llevado a cabo. Los tuits se obtuvieron programando la función de búsqueda antes mencionada del 19 de octubre al 2 de noviembre de 2020 (parámetros “since” y “until” de la función). No obstante, sea por restricciones de Twitter o de la misma función de búsqueda referida, los tuits obtenidos abarcaban el periodo del 30 de octubre al 2 de noviembre de 2020 (campo *created\_at* de los tuits). El estudio se centra, por consiguiente, en los últimos días de campaña electoral. Siguiendo las indicaciones de mi tutor de TFM, se han utilizado 2-3 hashtags o palabras neutras para llevar a cabo el estudio. No obstante, estas indicaciones no implican un número determinado de hashtags o palabras clave, ya que este número depende del objeto de estudio y de lo que se pretende estudiar del mismo. En este sentido, concretando las búsquedas realizadas, en el caso del hashtag elegido, “#Elections2020”, si bien había otras opciones (*#EleccionesUSA2020*, *#Elecciones2020*, *#USAelection2020*), la opción elegida ha facilitado la captura de tuits en lengua inglesa y ha resultado ser bastante genérica, lo que ha asegurado una adecuada cantidad de tuits. En cuanto a las palabras clave o perfiles de los principales candidatos, debe tenerse en cuenta que buena parte del desarrollo e implementación del trabajo gira en torno al nombre de perfil de dichos candidatos. Poner como palabras clave el nombre de perfil de los principales candidatos lo he considerado como una garantía para conseguir suficientes tuits, de modo que se asegurase llevar a cabo un análisis de palabras más frecuentes y de sentimientos lo bastante sólido y satisfactorio a partir de los perfiles que citan a cada uno de los candidatos. Esto permitirá hacer una comparativa entre los contenidos de los perfiles que directamente interpelan a los candidatos y los contenidos de los perfiles que incluyen el hashtag neutro. Todas estas elecciones han establecido, en buena medida, el marco principal de trabajo de esta investigación. Los tuits capturados se distribuyen de la siguiente manera:

- 13.877 corresponden a la búsqueda con la palabra clave o hashtag “#Elections2020”;
- 11.574 corresponden a la búsqueda con la palabra clave o perfil de usuario “@realDonaldTrump”;
- 16.412 corresponden a la búsqueda con la palabra clave o perfil de usuario “@JoeBiden”.

El siguiente gráfico nos permite entender mejor la distribución de tweets:

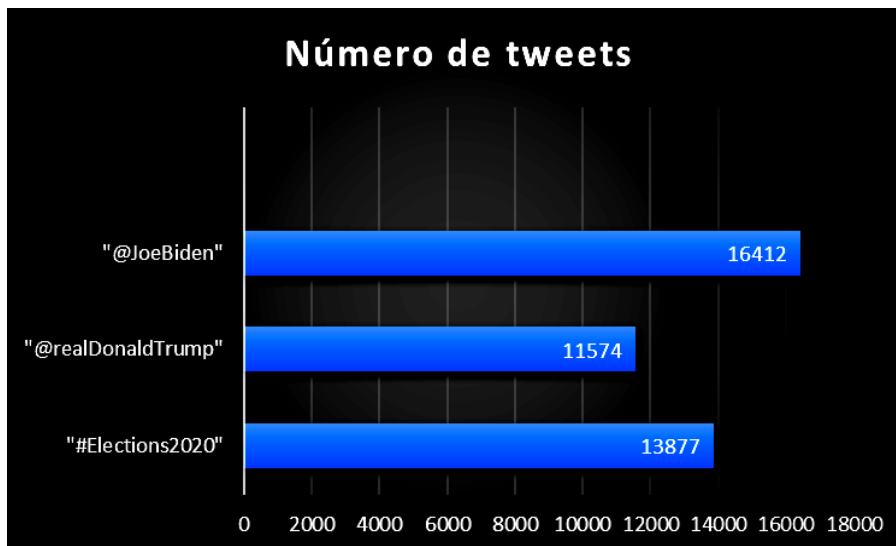


Figura 7. Distribución de los tweets capturados para el presente trabajo.  
Fuente: Elaboración propia.

Cada uno de los grupos establecidos, además, contempla tanto tweets originales como retweets. Conviene citar que se programó la función de búsqueda *search\_tweets()* para que sólo contemplara tweets en lengua inglesa y se excluyeran respuestas (parámetro *'-filter' = "replies"*), en el periodo que va del 30 de octubre al 2 de noviembre de 2020 (días previos que preceden a las elecciones, siendo de esperar que se intensificara la actividad en Twitter sobre este tema). El siguiente gráfico nos muestra la clasificación para cada uno de los grupos establecidos:

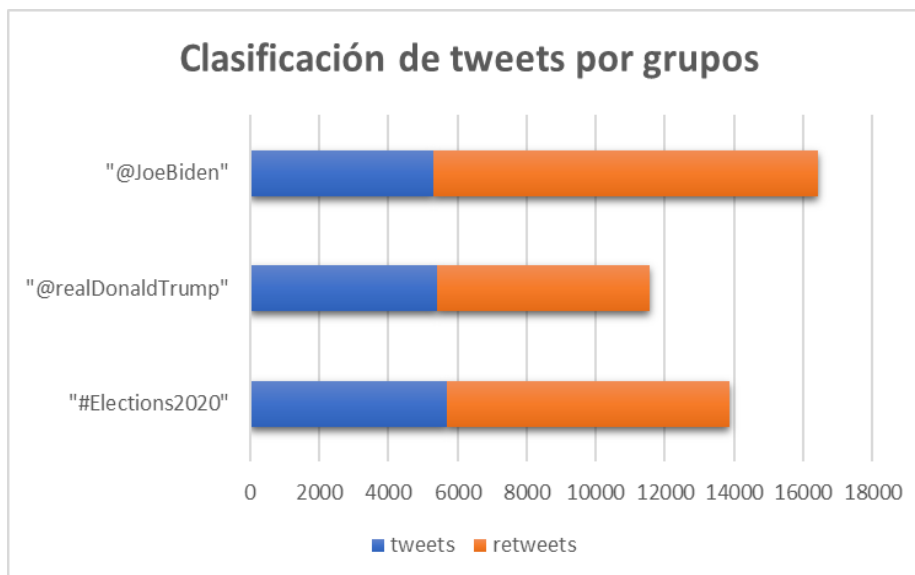


Figura 8. Clasificación de tweets por grupos.  
Fuente: Elaboración propia.

De la muestra obtenida, el candidato demócrata Joe Biden es citado más veces en los textos de Twitter que su oponente Donald Trump, al que supera en número de retuits que contienen su nombre de perfil. Del total de tweets capturados, se lleva la mayor parte. Se debe entender que los tuits que contienen el perfil de Biden, al haber sido más retuiteados, son los que han tenido una mayor difusión.

En cuanto a los lugares de procedencia de los tweets, para cada uno de los grupos podemos establecer unos rankings que se detallan a continuación:

“#Elections2020” :

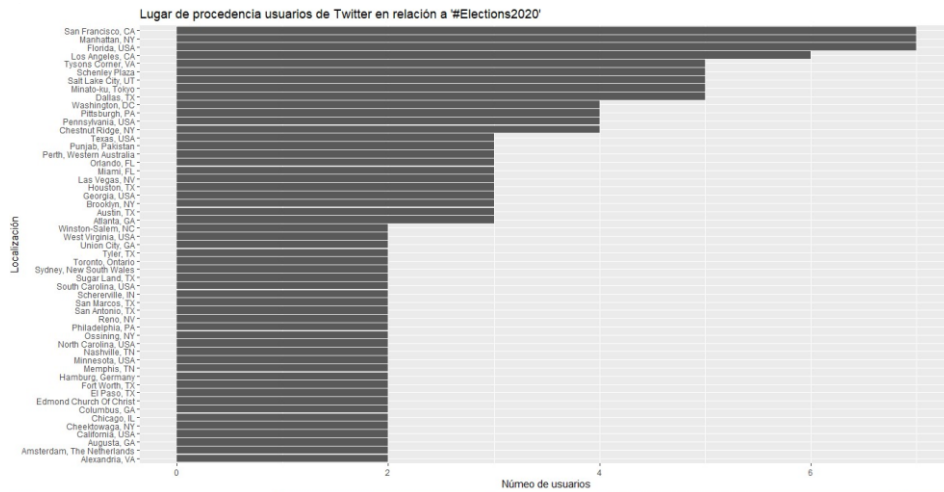


Figura 9. Lugares de procedencia tweets asociados a “Elections2020”  
Fuente: Elaboración propia

“@realDonaldTrump”

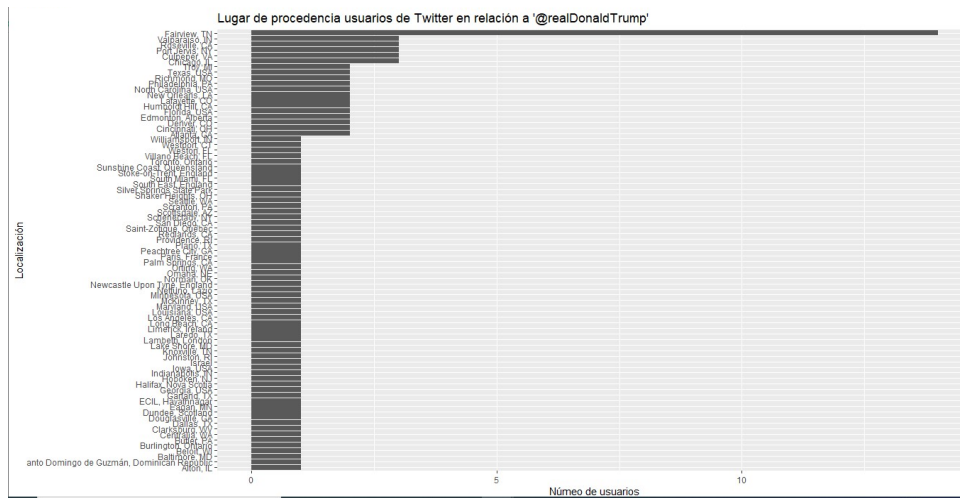


Figura 10. Lugares de procedencia tweets asociados a “@realDonaldTrump”  
Fuente: Elaboración propia

“@JoeBiden” :

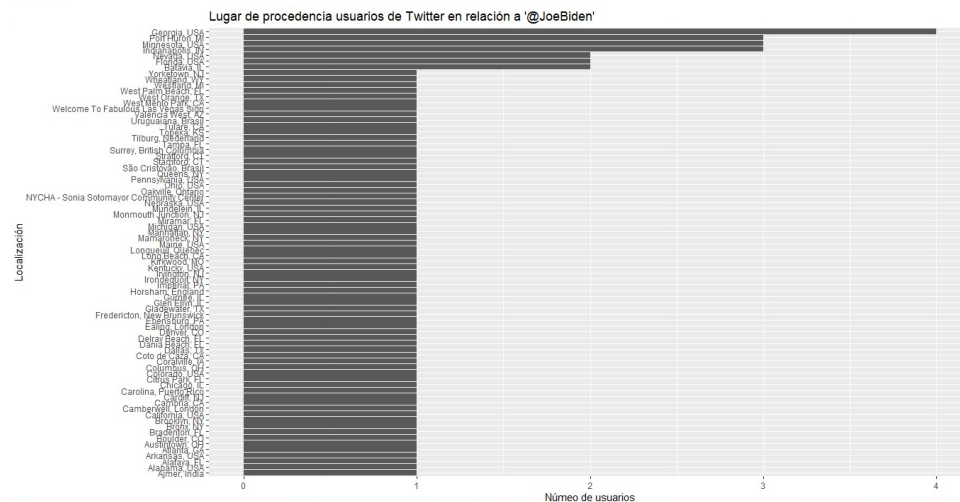


Figura 11. Lugares de procedencia tweets asociados a “@JoeBiden”  
Fuente: Elaboración propia

Notamos que los tres conjuntos de datos presentan diferentes lugares con el mismo número de tweets.

Otro aspecto interesante es indagar sobre las horas del día en que se han capturado más tuits asociados a este trabajo. Para simplificar el estudio y ante la imposibilidad de reproducir el resto de horas para todo el conjunto de días considerado, se muestran los gráficos para cada uno de los conjuntos de datos correspondientes al día 30 de octubre de 2020 entre las 21:00 y las 00:00 horas en los tres casos.

Para la designación “#Elections2020”, podemos apreciar que entre las 22 y las 23 horas se dio el máximo flujo de tweets y retweets.

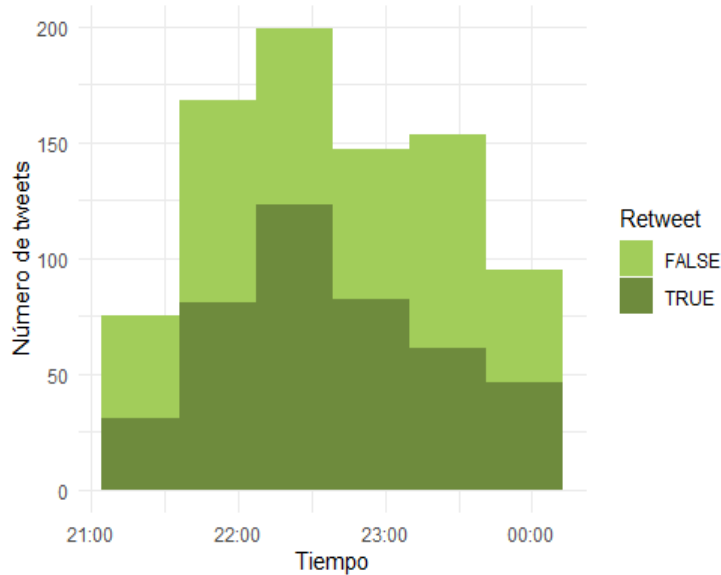


Figura 12. Distribución de tweets por hora para el conjunto de datos “#Elections2020”  
Fuente: Elaboración propia

Se advierte un flujo de datos con la designación “@realDonaldTrump” bastante uniforme a partir de las 14 horas en el mencionado día.

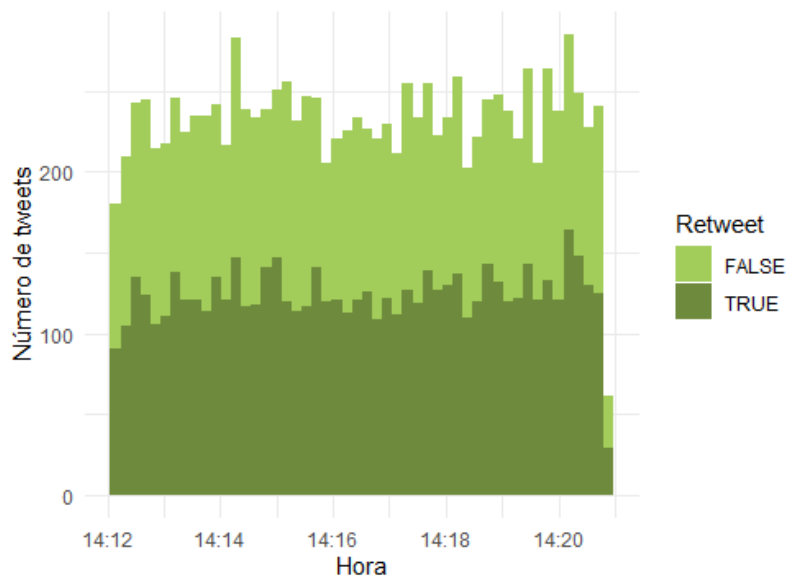


Figura 13. Distribución de tweets por hora para el conjunto de datos “@realDonaldTrump”  
Fuente: Elaboración propia

Se aprecia un flujo de datos con la designación “@JoeBiden” con un marcado repunte a las 14:25 horas del día 30 de octubre de 2020, para luego mantener un comportamiento más o menos uniforme durante los instantes en que se capturaron más tuits. Las razones de este incremento temporal del número de tweets son desconocidas.

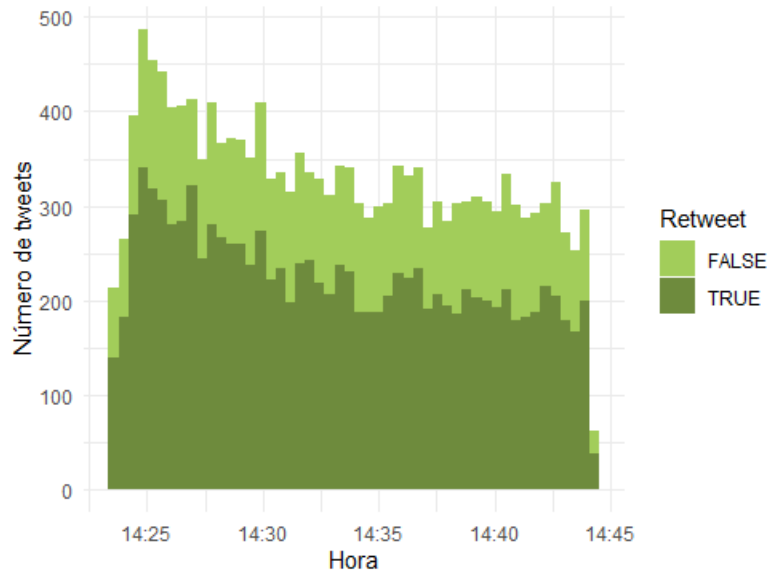


Figura 14. Distribución de tweets por hora para el conjunto de datos “@JoeBiden”  
Fuente: Elaboración propia

También es útil considerar un ranking de hashtags o temas más populares asociados a cada uno de los conjuntos de datos. Los hashtags brindan la oportunidad de encontrar tuits que hablen de un mismo tema. Existen miles de estadísticas de Twitter que se pueden obtener de una búsqueda en Twitter. Algunas estadísticas son sobre el hashtag, otras sobre los usuarios que han compartido contenido con ese hashtag. No sólo es importante analizar qué dice la gente y cuánto dice, sino también hay que saber quién es esa gente que comparte tuits con el hashtag. Podemos observar los resultados a partir de un diagrama de barras o bien a partir de una nube de palabras. Veamos cada uno de los conjuntos de datos:

“#Elections2020” :

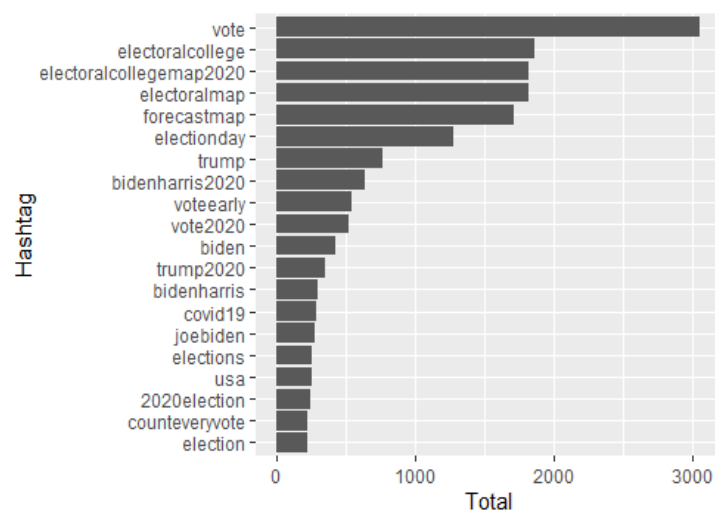


Figura 15. Top 20 hahstags más populares junto con “#Elections2020”  
Fuente: Elaboración propia

En este primer caso, estamos graficando hashtags en tuits que están descargados a partir del hashtag “#Elections2020”, con lo que la información no es muy relevante porque se encuentra enmarcada dentro del tema que fija el hashtag de inicio. No obstante, he decidido incluir el gráfico en esta sección de radiografía de tuits para tener una visión de los hashtags más populares asociados al elegido como objeto de estudio. Esta inclusión, a mi entender, ha valido la pena, ya que me llama la atención que forme parte del ranking el hashtag “covid19”, el único que no se relaciona directamente con el hashtag de las elecciones. De alguna forma, aporta información sobre el contexto en que se desarrolla la campaña electoral objeto de estudio.

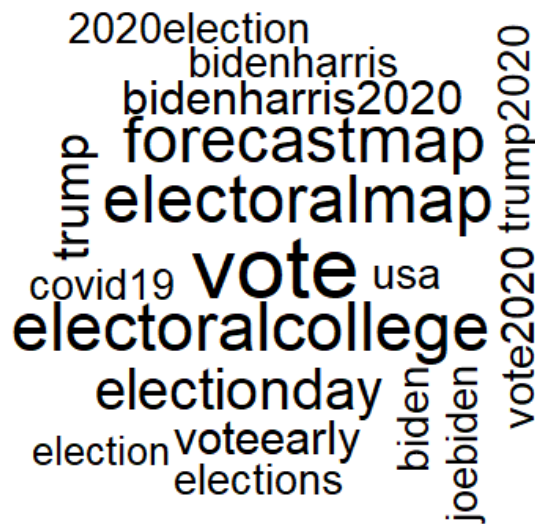


Figura 16. Top 20 hashtags más populares junto con “#Elections2020”  
Fuente: Elaboración propia

“@realDonaldTrump” :

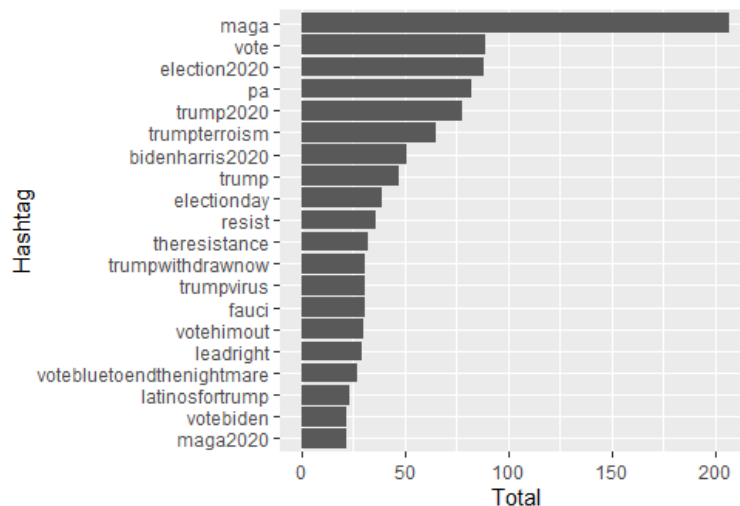


Figura 17. Top 20 hashtags más populares junto con “@realDonaldTrump”  
Fuente: Elaboración propia

Se observa que “maga” (abreviatura de *Make America Great Again*) es el hashtag más utilizado y que el número de tuits asociados a este hashtag es muy grande respecto al total. Esto da a entender que los que citan a Trump en sus tuits le dan apoyo de forma mayoritaria por no decir incondicional, ya que usan como hashtag más destacado el



eslogan de su campaña. Viendo este hashtag mayoritario, parece que el grueso de mensajes de los que mencionan a Trump es en tono de elogio.

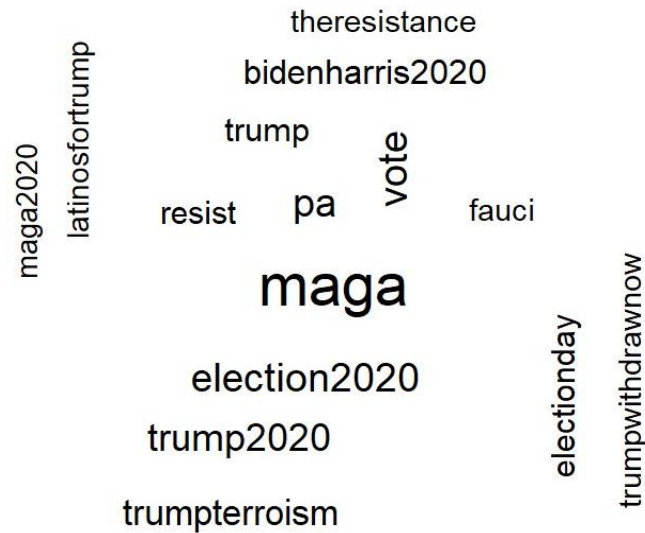


Figura 18. Top 20 hashtags más populares junto con “@realDonaldTrump”  
Fuente: Elaboración propia

“@JoeBiden” :

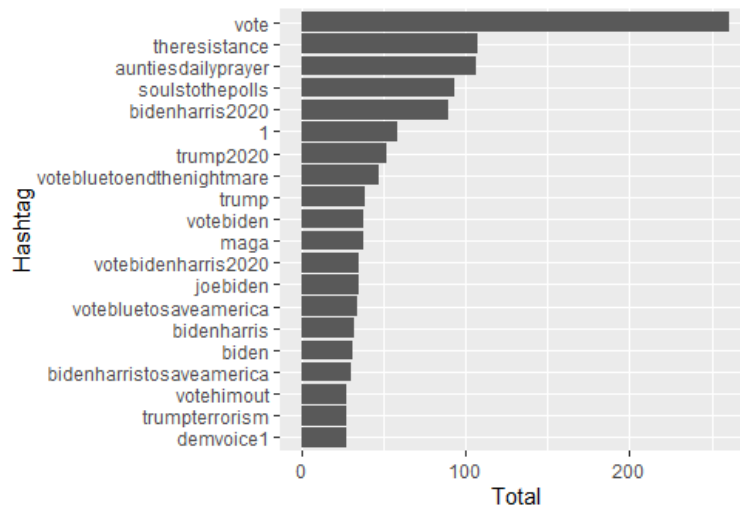


Figura 19. Top 20 hashtags más populares junto con “@JoeBiden”  
Fuente: Elaboración propia

De forma análoga al caso anterior, “vote” es el hashtag más utilizado, queda reforzado por “theresistance” y su número de tuits asociado es muy grande con respecto al total. Parece dar a entender que los que citan a Biden en sus tuits también lo apoyan de forma mayoritaria. Como curiosidad, he podido comprobar en Twitter que el “1” que aparece no corresponde realmente a un hashtag. Se trata, por tanto, de un dato de “ruido” que ha interferido en la búsqueda que se ha llevado a cabo.

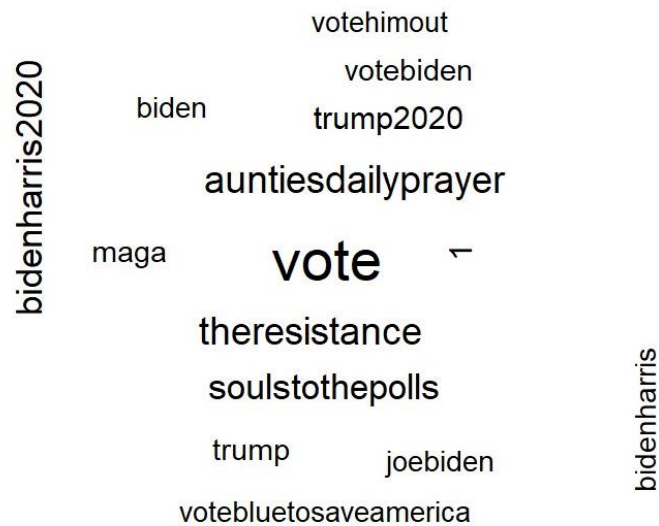


Figura 20. Top 20 hashtags más populares junto con “@JoeBiden”  
Fuente: Elaboración propia

Para terminar esta sección de diagnóstico de tweets, identificamos los usuarios de Twitter más activos para cada uno de los conjuntos de datos que forman esta investigación y graficamos los resultados obtenidos. La selección la obtenemos a partir del componente “screen\_name” de cada tuit, aplicando como criterio un recuento en orden decreciente mediante la función *count()* del lenguaje R. Como último paso, con la función *mutate()* añadimos el carácter “@” al campo “screen\_name” tratado para que tenga formato de nombre de usuario Twitter. Para los gráficos que se muestran a continuación, el eje horizontal corresponde al número de veces que se detecta un “screen\_name” y se escala en valor absoluto en base a este recuento (resultado de aplicar la función *count()* de R), mientras que en el eje vertical simplemente se citan estos “screen\_name”.

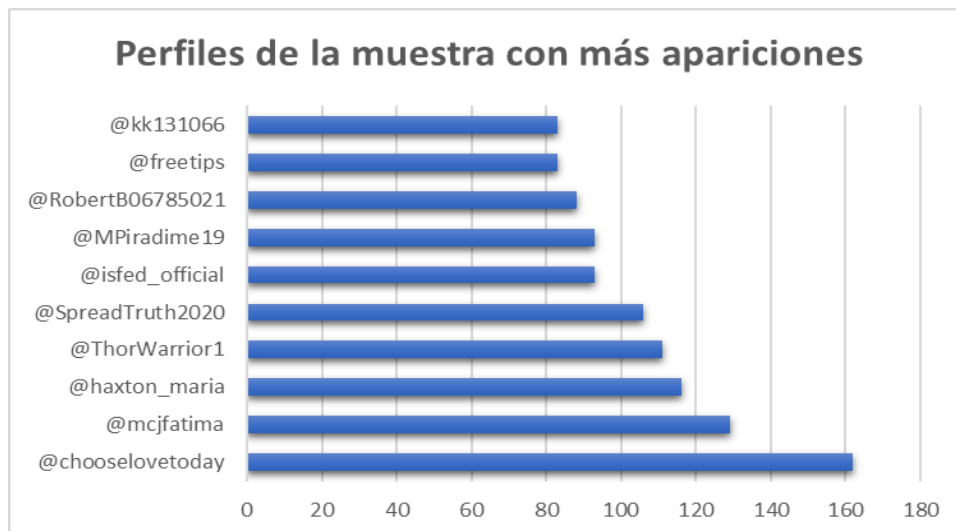


Figura 21. Top 10 recuento de “screen\_name” conjunto de datos “#Elections2020”  
Fuente: Elaboración propia

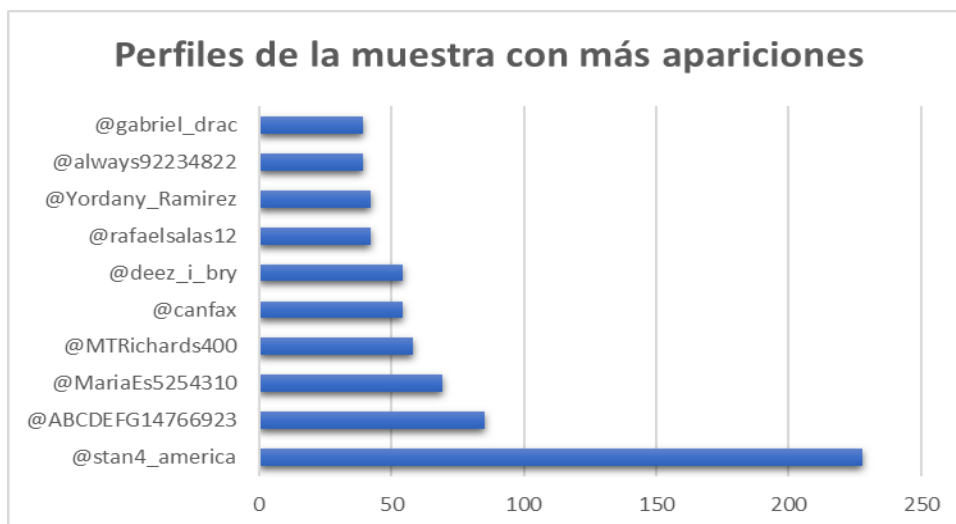


Figura 22. Top 10 recuento de “screen\_name” conjunto de datos “@realDonaldTrump”  
Fuente: Elaboración propia

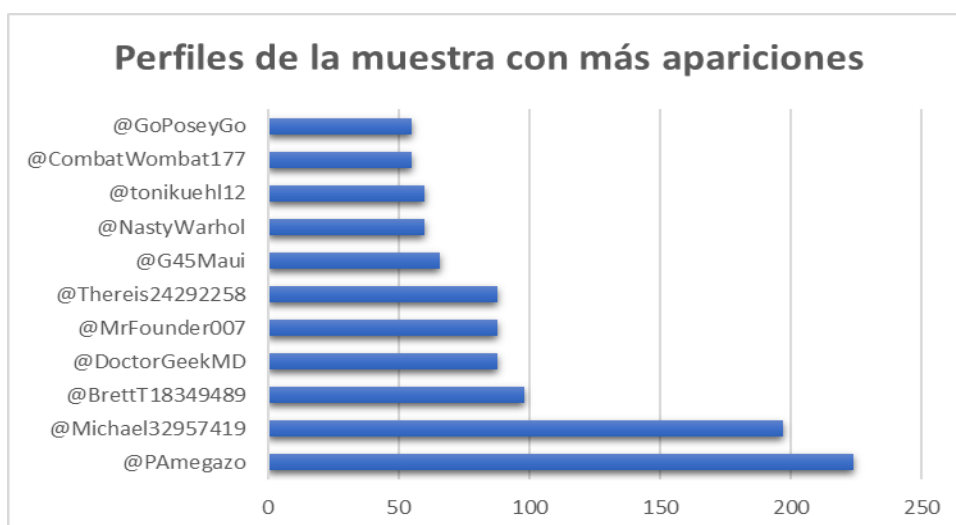


Figura 23. Top 10 recuento de “screen\_name” conjunto de datos “@JoeBiden”  
Fuente: Elaboración propia

Respecto a los gráficos anteriores, conviene citar también que la detección de perfiles se ha llevado a cabo excluyendo retuits mediante la opción *filter(is\_retweet==F)* para garantizar un análisis exclusivo de los usuarios que han realizado más menciones y no de los que han tenido más difusión siguiendo un criterio de recuento de retuits. De no haber procedido así, los resultados no se podrían considerar del todo válidos. Atendiendo a los nombres de algunos perfiles obtenidos, además, parece que algunos de ellos son *bots*<sup>12</sup> (programas que controlan cuentas de Twitter). Es el caso, por ejemplo, del perfil de usuario “@ABCDEFGH14766923”.

De forma análoga, habiendo retirado previamente la cita propia de cada banco de datos, resulta también de interés conocer los top 10 de perfiles más citados, más allá de las citas a los principales candidatos. Estas gráficas van a permitir conocer a quién

<sup>12</sup> Un **bot** (aféresis de *robot*) es un programa informático que efectúa automáticamente tareas reiterativas mediante Internet a través de una cadena de comandos o funciones autónomas previas para asignar un rol establecido.

más citan los que citan tanto a Donald Trump como a Joe Biden, así como disponer de un recuento de estos tuits. Con esto, doy por finalizada la radiografía de tweets.

“@realDonaldTrump” :



Figura 24. Top 10 mentions en el conjunto de datos “realDonaldTrump”  
Fuente: Elaboración propia

“@JoeBiden” :

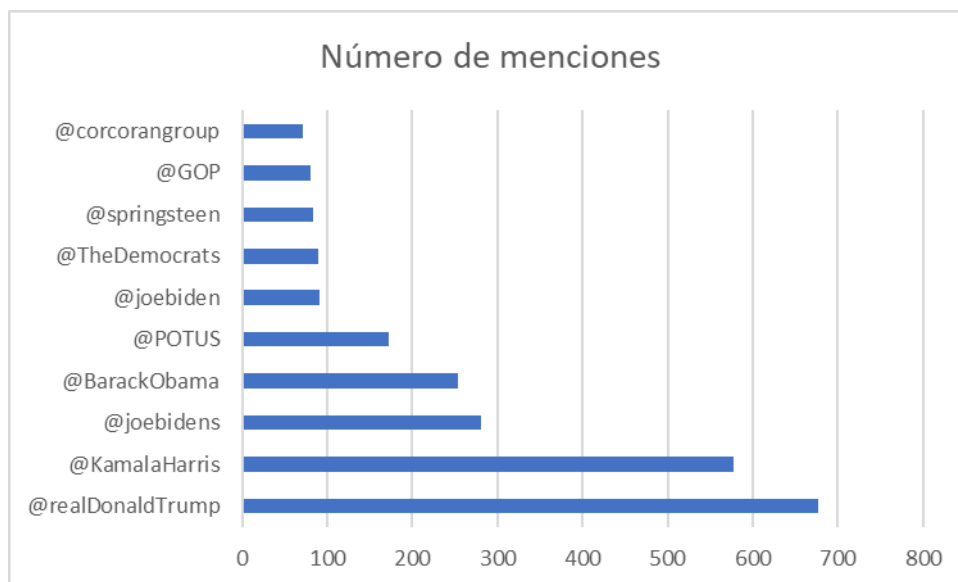


Figura 25. Top 10 mentions en el conjunto de datos “@JoeBiden”  
Fuente: Elaboración propia

Nótese que ambos conjuntos de datos mencionan el perfil de su candidato contrincante. El conjunto de datos del perfil Joe Biden, además, menciona a Kamila Harris, Barack Obama y Bruce Springsteen.

### 3.2 Análisis de contenidos

En esta sección se detalla el análisis realizado sobre el conjunto de tuits radiografiado en el apartado anterior. Conviene señalar que este análisis no se ha realizado sobre

los datos en bruto, sino que previamente se ha sometido a los tuits a un proceso de limpieza y filtro para garantizar la calidad de los datos y de los resultados obtenidos. Este proceso de limpieza de datos ha contemplado las siguientes acciones:

- Construcción de un corpus y especificación de una fuente de vectores de caracteres. Se actúa sobre el campo "text" de los tuits.
- Conversión a palabras en minúscula.
- Eliminación de URLs.
- Eliminación de todo aquello que no sean letras o espacios en inglés.
- Eliminación de palabras vacías (palabras comunes del vocabulario inglés, como por ejemplo, "the" o "it").
- Eliminación de espacios en blanco adicionales.
- Eliminación de emoticonos.
- Copia del corpus por seguridad y como medida de contingencia ante pérdida del corpus original.
- Sustitución de algunas palabras por otras.
- Creación de la matriz de documentos de términos.

Una vez los datos han sido preprocesados y se puede asegurar su limpieza y calidad (datos en un formato de texto ordenado de un token por fila), el análisis se ha concretado en cuestiones como las palabras más frecuentes de cada grupo de datos, los *wordclouds* o nubes de palabras, el análisis de sentimientos y el de contenidos de los usuarios que más mencionan cada grupo de datos, así como la detección de la estrategia seguida por cada uno de los principales candidatos, si la hubiera, a partir de sus mensajes publicados en Twitter. Vamos a ver cada una de ellas por separado.

### 3.2.1 Palabras más frecuentes

Se ha calculado el recuento de palabras para cada una de las colecciones de tuits y se han obtenido gráficos de barras con las 20 palabras más frecuentes. Los gráficos se han obtenido usando la función *ggplot* de R y contemplando en el recuento tanto tuits como retuits (los "replies" se excluyeron en la descarga). Para el primer gráfico, se ha vetado la palabra "elections2020" porque es la propia del hashtag y se da por supuesto que sea utilizada, ya que todos los tuits de este primer conjunto de datos contienen este hashtag. Como resultados interesantes, la palabra "vote" aparece en todos los conjuntos de datos y, además, la palabra más utilizada en los discursos con menciones a cada uno de los principales candidatos alude al oponente. Así, en el discurso de los que citan a Donald Trump la palabra más usada es "biden", mientras que en el discurso de los que citan a Joe Biden la palabra más utilizada es "donald". Conviene señalar que las palabras "president", "people" y "time" están presentes en los discursos de los que siguen a ambos candidatos. El discurso de los que mencionan al candidato demócrata Joe Biden parece ser más esperanzador y comprometido, atendiendo a palabras como "i'll", "let's" o "promise". No obstante, también nos encontramos con palabras como "crisis" y "chaos". Para salir de dudas, se tendría que comprobar si estas palabras van dirigidas contra el gobierno de Trump o su gestión. En el discurso de los que mencionan a Donald Trump, por el contrario, se hallan palabras asociadas a connotaciones más sombrías, como por ejemplo "black" y "crazy". A continuación, se muestran los resultados obtenidos en el recuento de palabras más frecuentes.

“#Elections2020” :

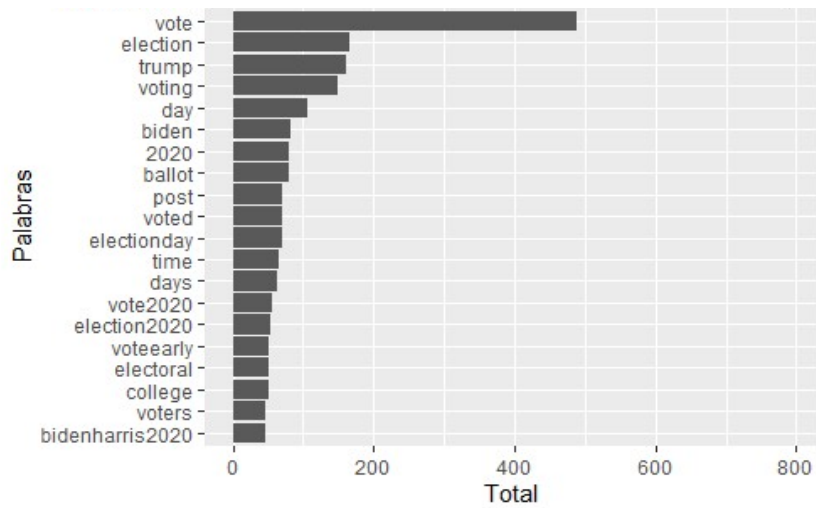


Figura 26. Palabras más frecuentes encontradas en el grupo #Elections2020  
Fuente: Elaboración propia

“@realDonaldTrump” :

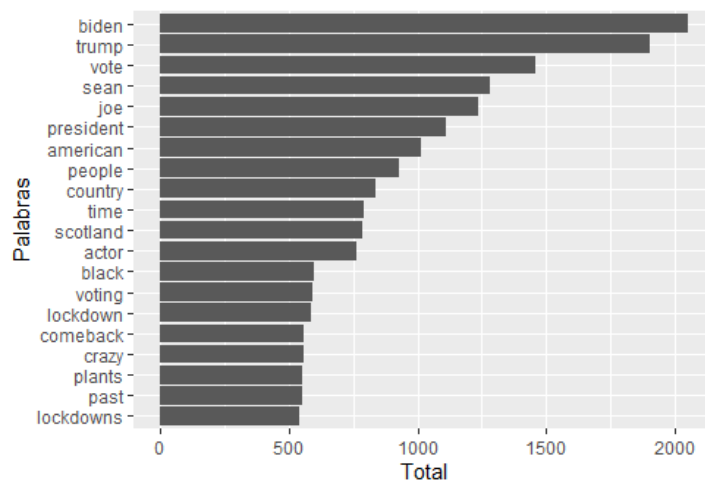


Figura 27. Palabras más frecuentes encontradas en el grupo @realDonaldTrump  
Fuente: Elaboración propia

“@JoeBiden” :

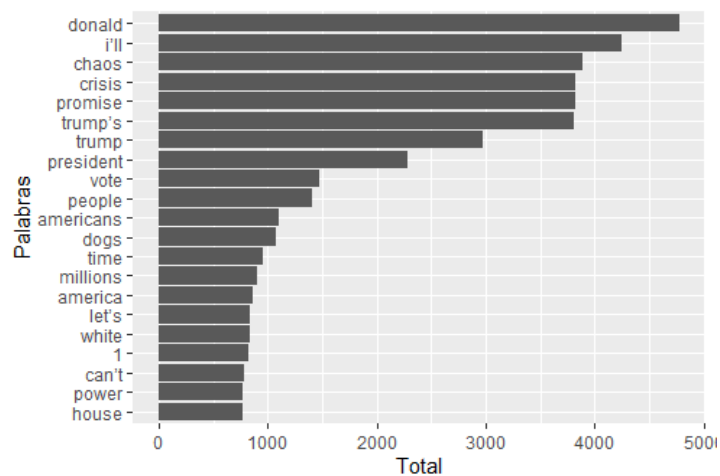


Figura 28. Palabras más frecuentes encontradas en el grupo @rJoeBiden  
Fuente: Elaboración propia

### 3.2.2 Word clouds (nubes de palabras)

Una nube de palabras es una representación visual de las palabras más frecuentes en un texto. En el ámbito web, se utiliza para hacer referencia a los términos que más se emplean en ella, algo que normalmente está conectado a lo que se conoce como hashtags o etiquetas. Conviene destacar que, para reforzar más o menos la presencia de los distintos términos, estos suelen tener un mayor o menor tamaño y colores distintos en función de la cantidad de veces que aparecen. Las nubes de palabras tienen como principal función dar a conocer los términos más empleados dentro de una web. Para el lector, es algo que ayuda a saber qué temas son los que más se cubren en los diferentes contenidos que haya en esta, como también acceder a todos ellos con tan sólo interactuar con dicha nube. Crear una nube de palabras adecuada supone aprovecharse de las siguientes ventajas:

- Es un indicador exacto para saber de qué trata un espacio online concreto.
- Se trata de información visual muy sencilla y clara.
- Es muy útil para crear relaciones entre conceptos.
- Bien trabajada, puede ser un aspecto importante para mejorar el posicionamiento de una web.

No obstante, si la nube de palabras no se ha creado bien puede presentar como desventajas dar lugar a contenidos duplicados y provocar la dispersión de la autoridad que tenga concedido un contenido web.

Siguiendo con nuestro estudio, si bien las palabras frecuentes y las nubes de palabras aportan la misma información y se podría prescindir de una de las dos, a continuación muestro las nubes de palabras correspondientes a los distintos conjuntos de datos obtenidas haciendo uso de la función *wordcloud2* de R.

“#Elections2020” :

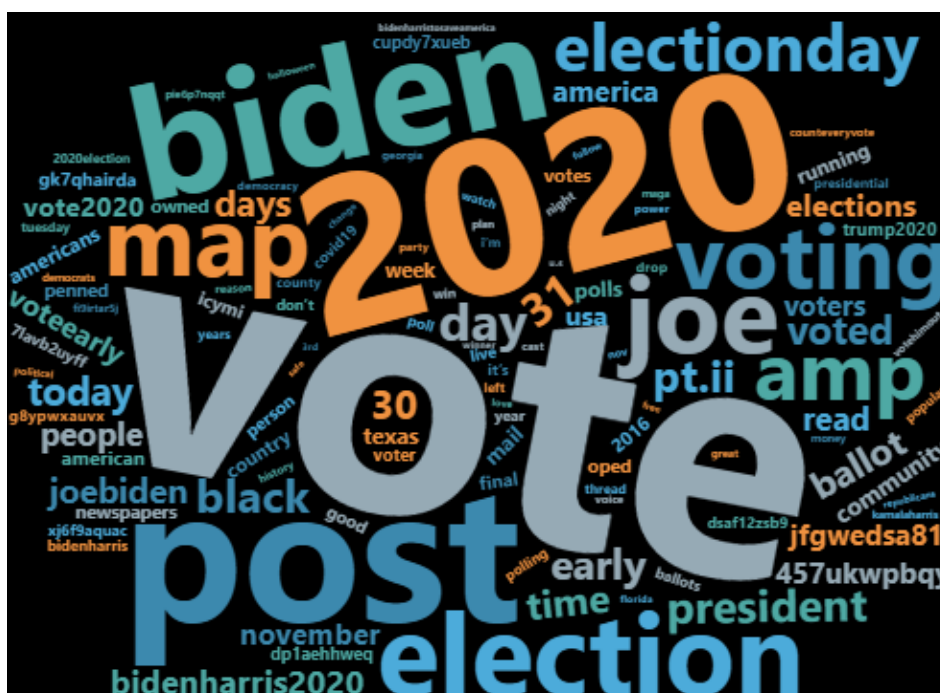


Figura 29. Nube de palabras asociada al grupo #Elections2020  
Fuente: Elaboración propia

“@realDonaldTrump” :

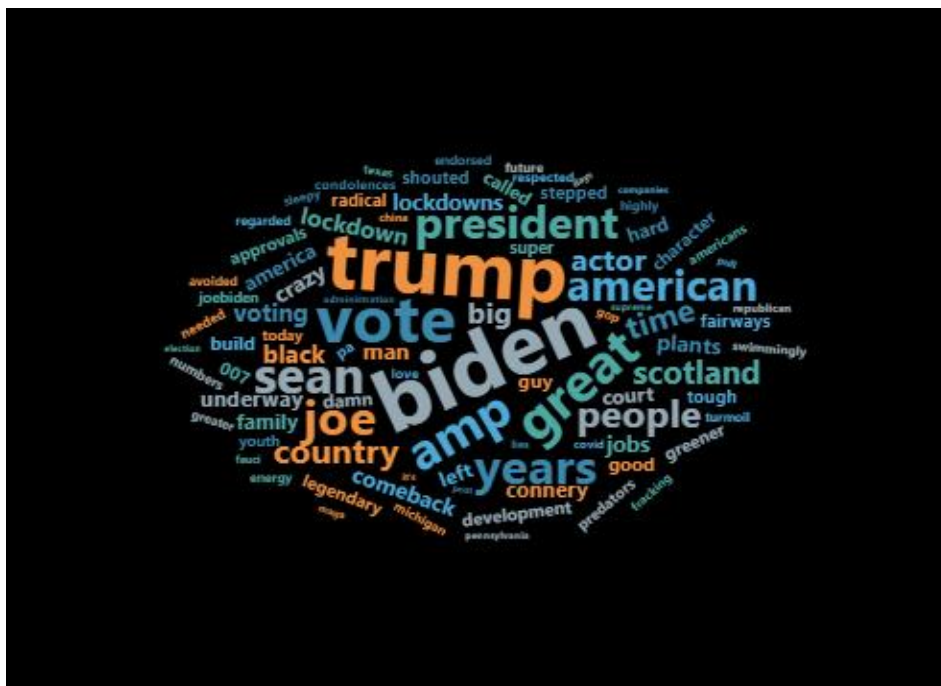


Figura 30. Nube de palabras asociada al grupo @realDonaldTrump  
Fuente: Elaboración propia

“@JoeBiden” :

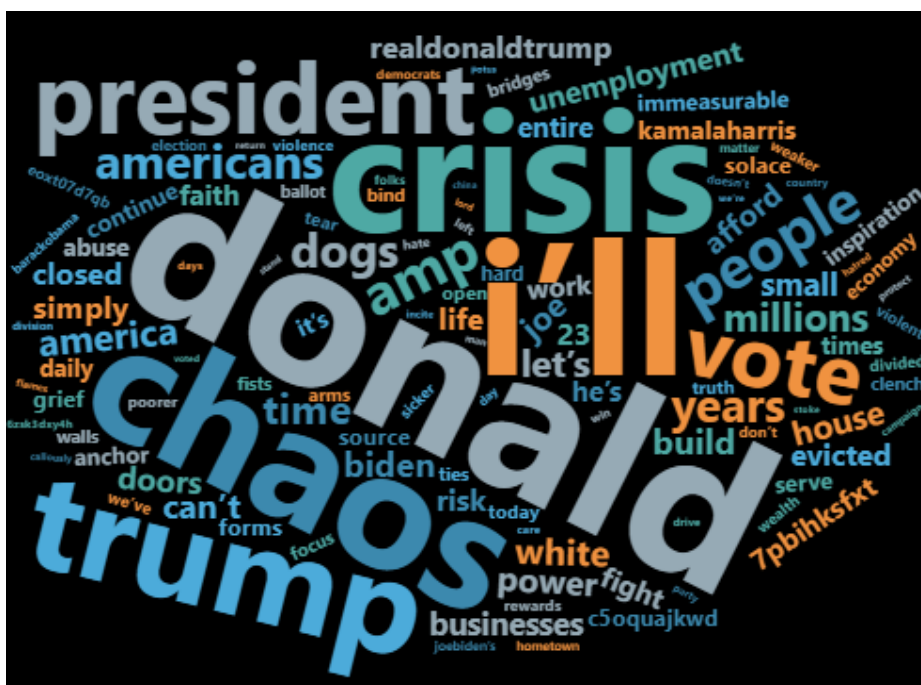


Figura 31. Nube de palabras asociada al grupo @JoeBiden  
Fuente: Elaboración propia

Llama la atención que la nube de Joe Biden está mucho más poblada de palabras que la de Donald Trump. Se han recogido muchos más tuits que contienen el nombre de perfil del candidato demócrata que del republicano, pero este hecho no justifica necesariamente que tengamos más palabras. Una posible explicación sería que los que mencionan a Joe Biden en Twitter tienden a repetir más veces las mismas





proporciona la función *get\_sentiments()*, que evalúa el sentimiento de cada palabra u oración. Este trabajo analizará cada uno de los conjuntos de datos obtenidos a partir de los léxicos de sentimientos “bing” y “nrc”.

En todos los conjuntos de datos, el diccionario otorga a la palabra “trump” un sentimiento positivo de elevada puntuación debido a su significado (triumfo). No obstante, es muy probable que se contemplen tuits que contengan esta palabra y que no tengan connotaciones positivas (mensajes de crítica, reproches, ataques viscerales o incluso descalificaciones). Véase la sección 3.2.4 de este mismo trabajo. A modo de conclusión, siempre hay que tener en cuenta el contexto. Además, cuando se analizan palabras individuales, no se puede diferenciar el doble sentido. Por ejemplo, el léxico “nrc” da a la palabra “hope” una clasificación positiva, pero si el tweet dijera “there is no hope” estaríamos en presencia de una expresión negativa.

Empezamos con el léxico “bing”. Mostramos, para cada conjunto de datos, el top 15 de palabras agrupadas por sentimiento en modo binario (negativo o positivo). En este primer análisis, por consiguiente, no se entra en detalles en cuanto a la especificación del sentimiento, dentro del grupo positivo o negativo.

“#Elections2020” :

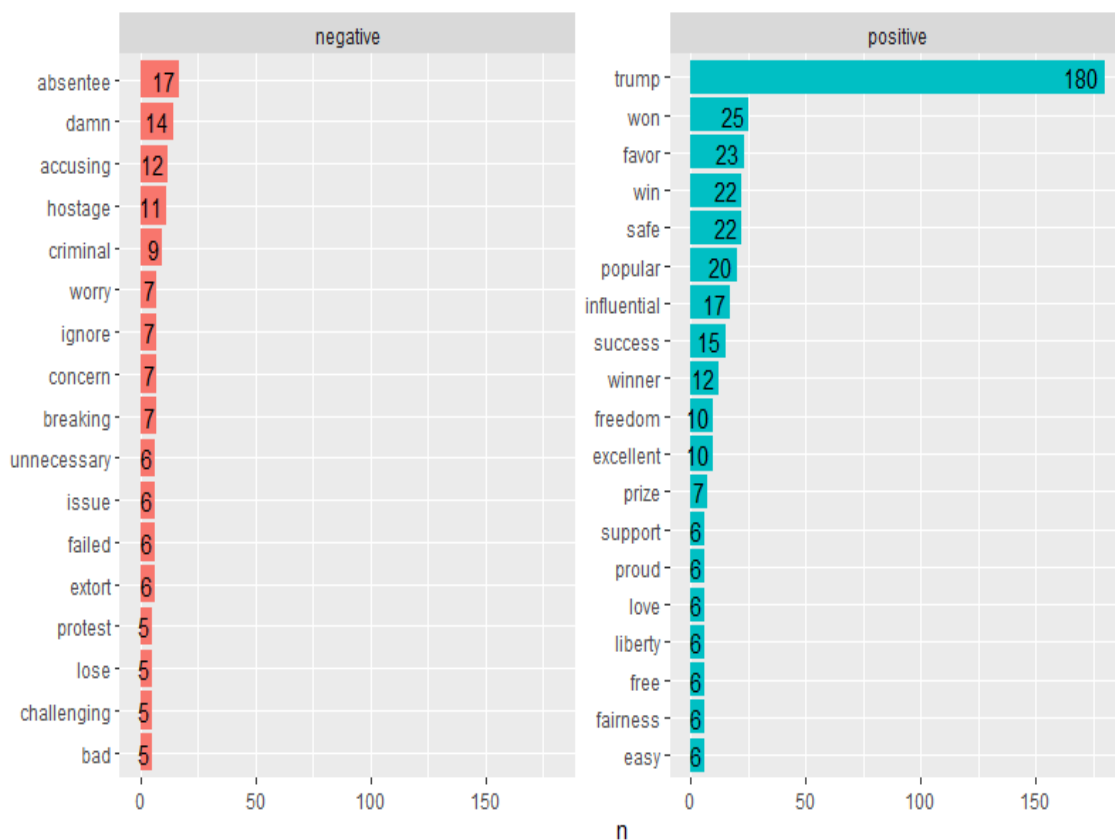


Figura 33. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “#Elections2020”  
Fuente: Elaboración propia

“@realDonaldTrump” :

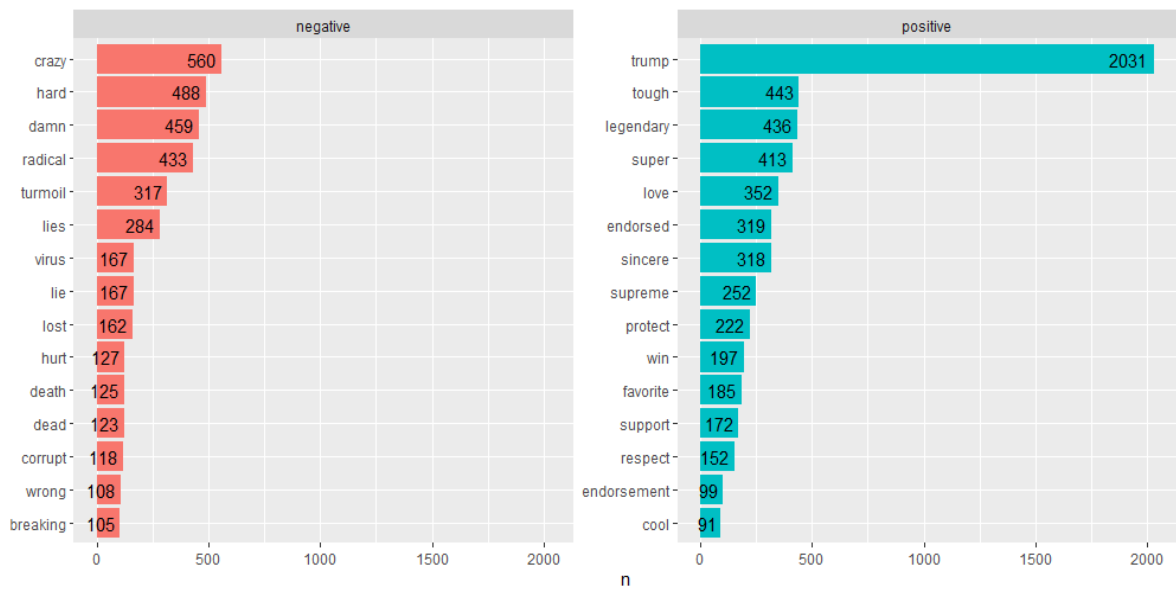


Figura 34. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@realDonaldTrump”  
Fuente: Elaboración propia

“@JoeBiden” :

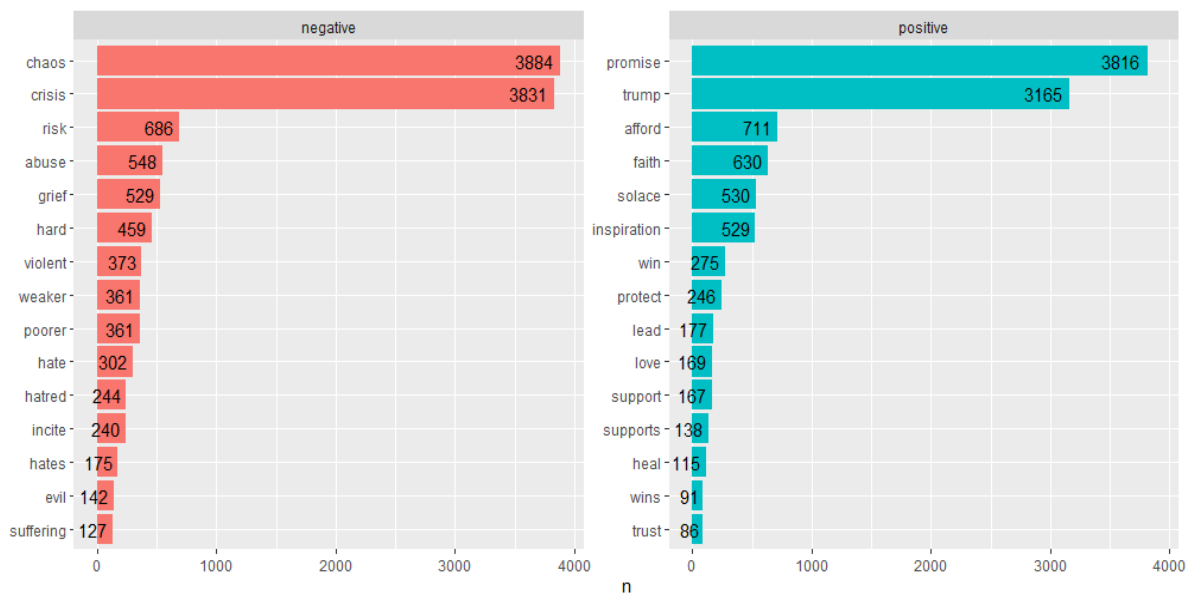


Figura 35. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@JoeBiden”  
Fuente: Elaboración propia

Vamos a centrarnos ahora en el léxico “nrc”. Para cada conjunto de datos se muestra un top 15 de palabras más repetidas según los distintos tipos de sentimiento que contempla este diccionario.

“#Elections2020” :

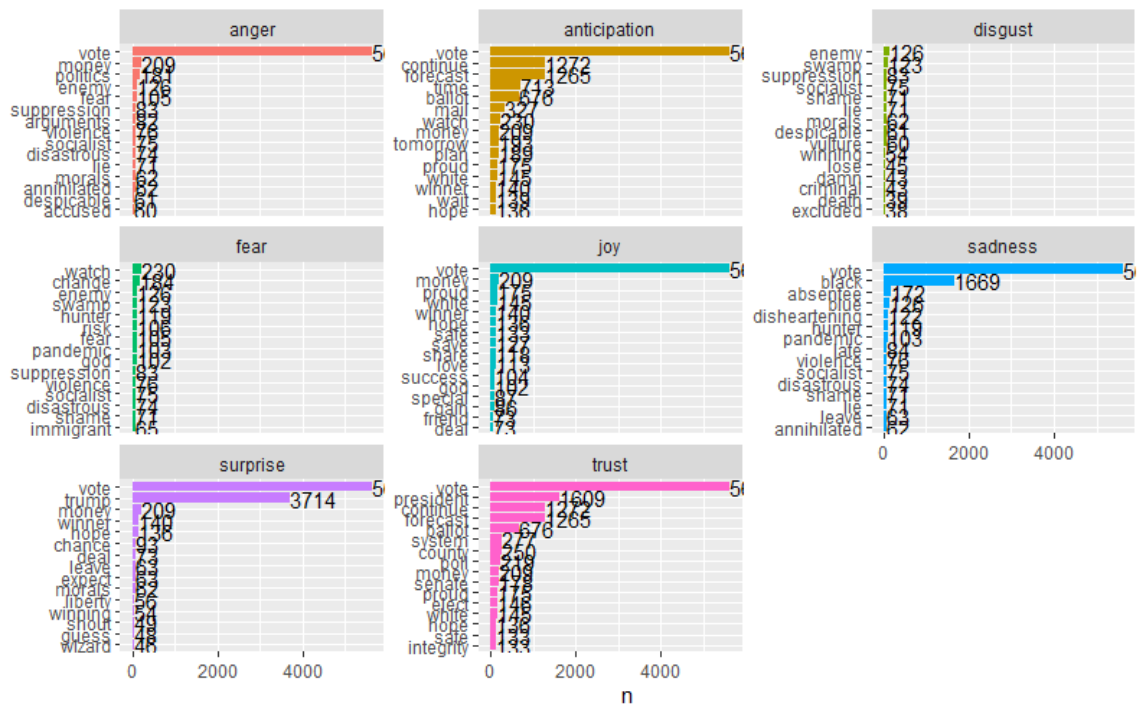


Figura 36. Top 15 palabras agrupadas por tipo de sentimiento conjunto de datos “#Elections2020”  
Fuente: Elaboración propia

“@realDonaldTrump” :

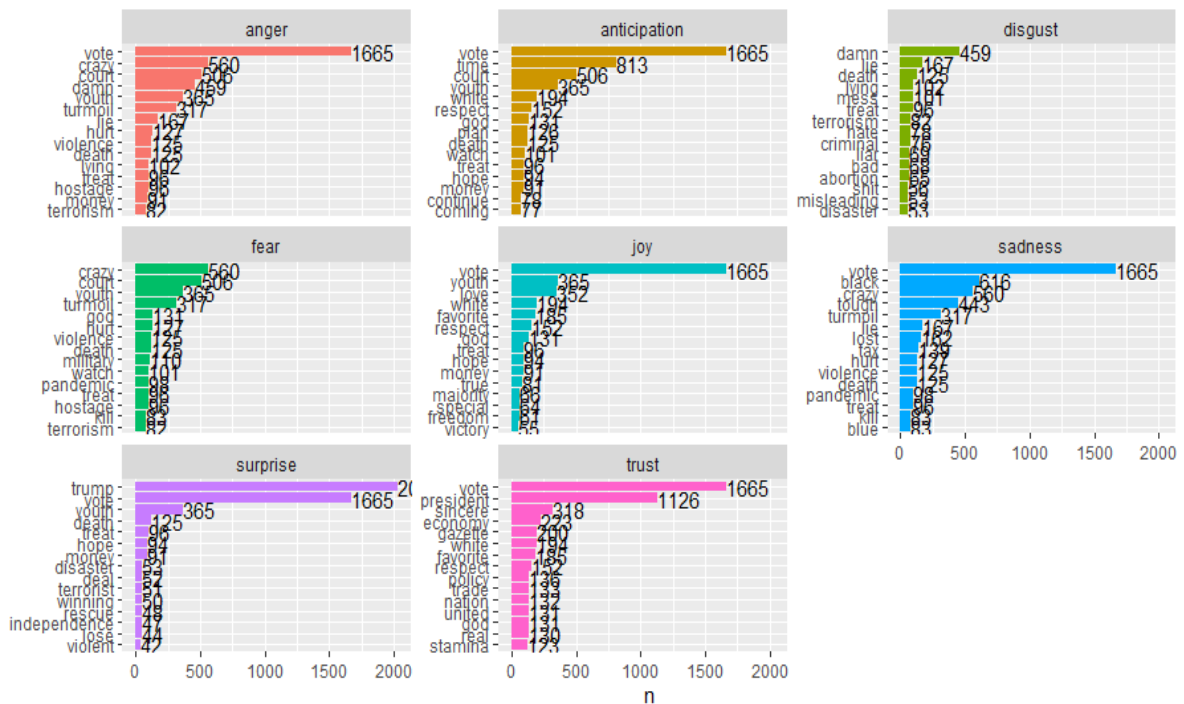


Figura 37. Top 15 palabras agrupadas por tipo de sentimiento conjunto de datos “@realDonaldTrump”  
Fuente: Elaboración propia

“@JoeBiden” :

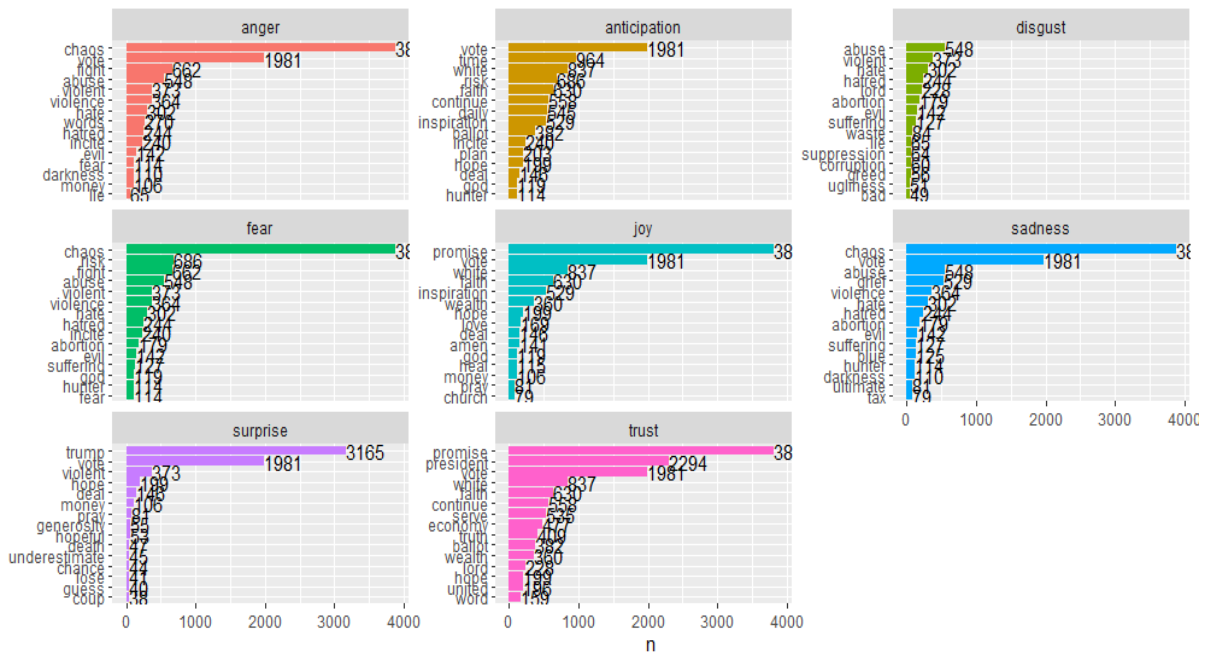


Figura 38. Top 15 palabras agrupadas por tipo de sentimiento conjunto de datos “@JoeBiden”  
Fuente: Elaboración propia

Seguindo con el léxico “nrc” y tal y como se hizo con el léxico “bing”, para cada conjunto de datos se muestra a continuación el top 15 de palabras agrupadas por sentimiento en modo binario (negativo o positivo).

“#Elections2020” :

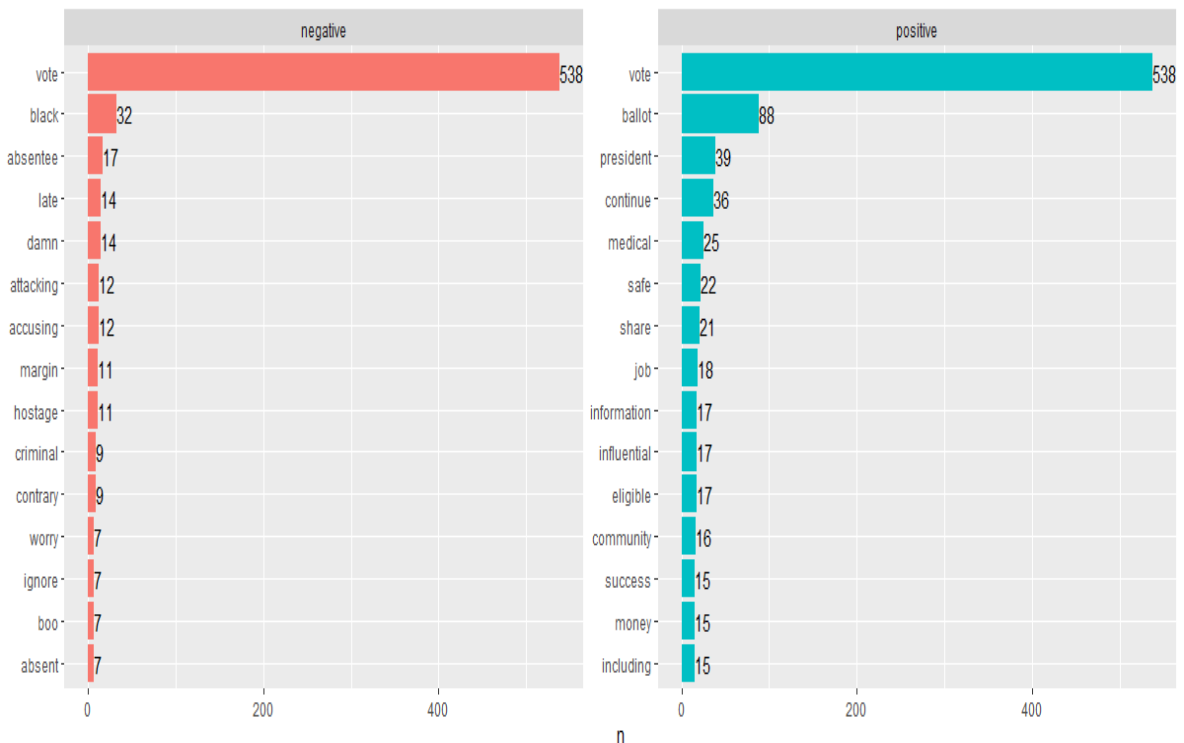


Figura 39. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “#Elections2020”  
Léxico “nrc”. Fuente: Elaboración propia

“@realDonaldTrump” :

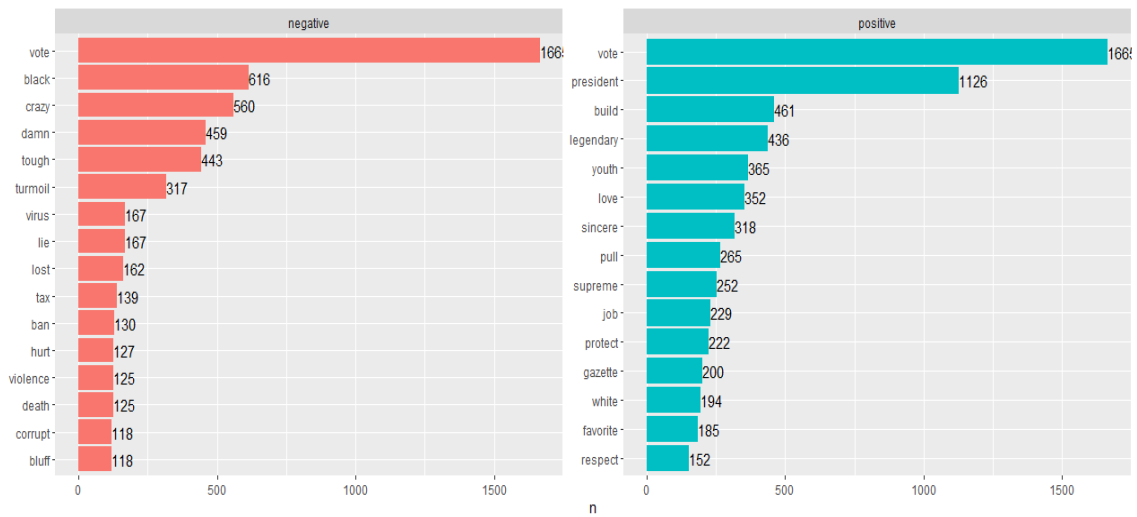


Figura 40. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@realDonaldTrump”  
Léxico “nrc”. Fuente: Elaboración propia

“@JoeBiden” :

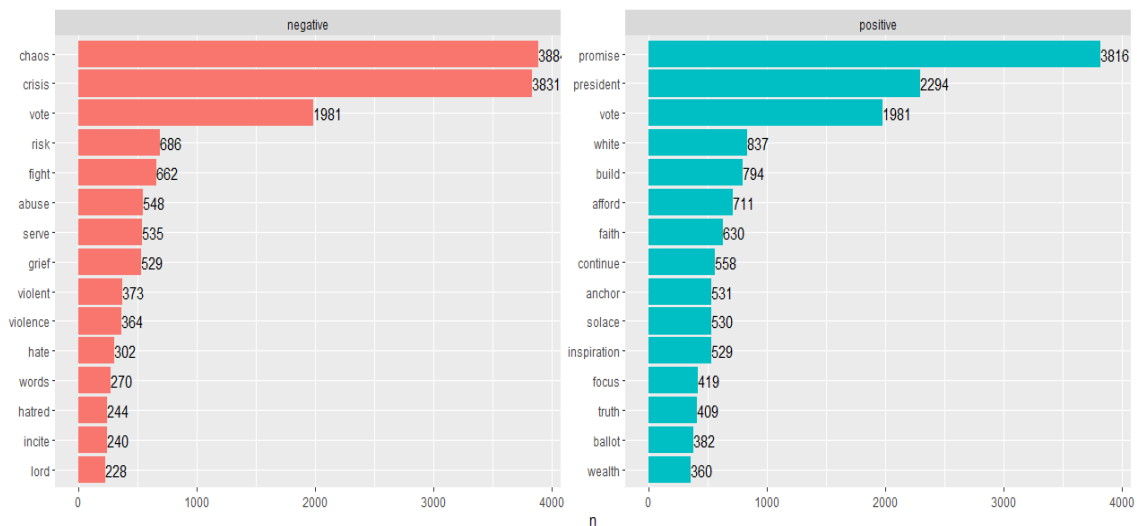
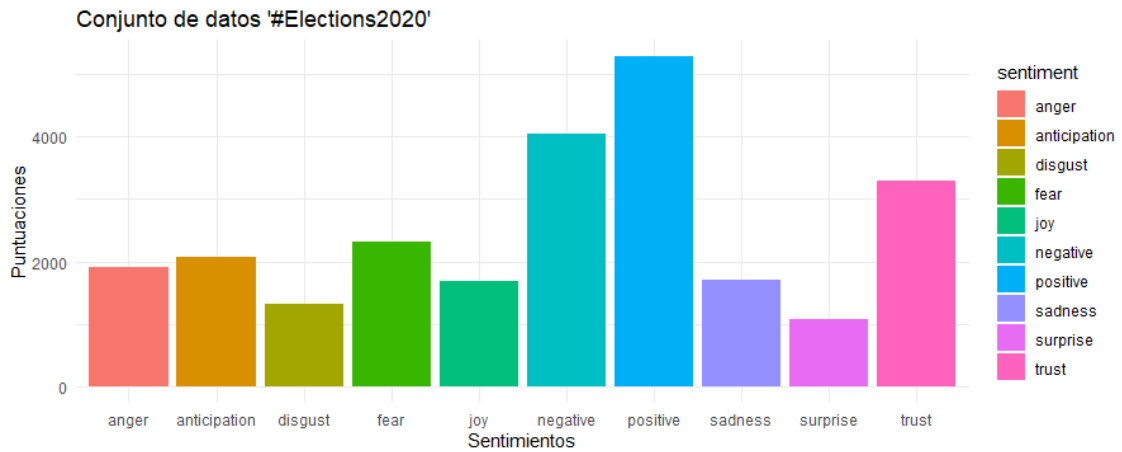
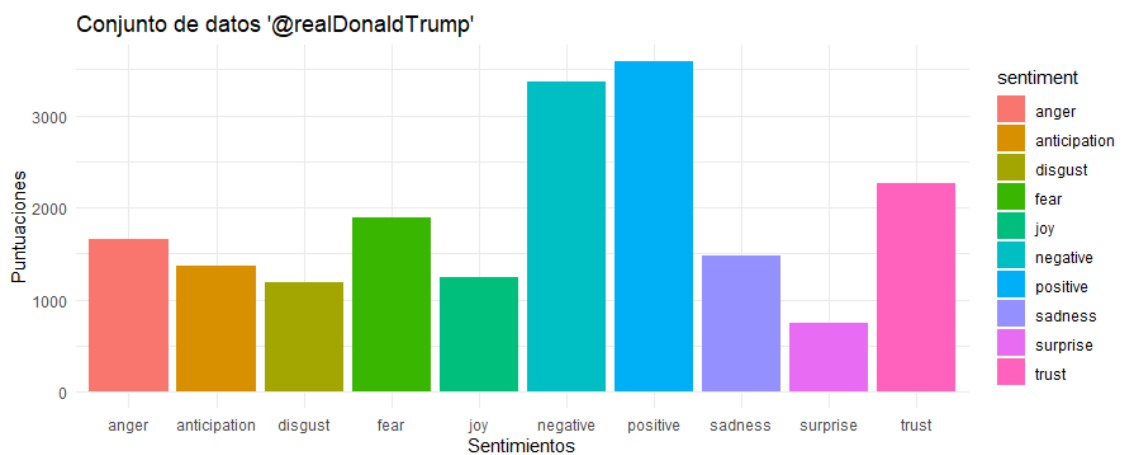


Figura 41. Top 15 palabras agrupadas por sentimiento binario conjunto de datos “@JoeBiden”  
Léxico “nrc”. Fuente: Elaboración propia

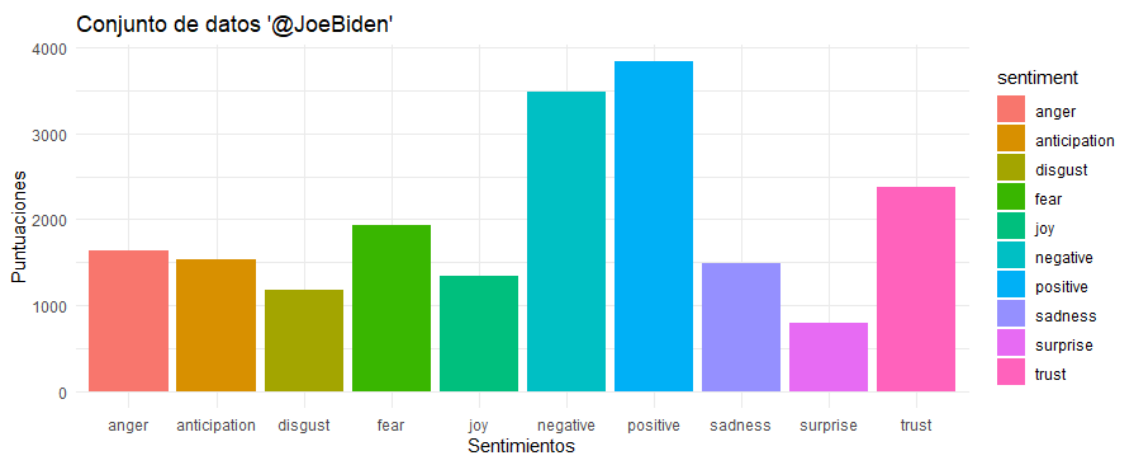
Por último, se muestran diagramas de barras correspondientes a los sentimientos encontrados por el diccionario “nrc” en cada uno de los conjuntos de datos. Cada sentimiento se distingue por un color y lleva asociada una barra que corresponde con su puntuación.



**Figura 42. Puntuación sentimientos encontrados conjunto de datos '#Elections2020'.**  
Fuente: Elaboración propia



**Figura 43. Puntuación sentimientos encontrados conjunto de datos '@realDonaldTrump'.**  
Fuente: Elaboración propia



**Figura 44. Puntuación sentimientos encontrados conjunto de datos '@JoeBiden'.**  
Fuente: Elaboración propia

También es interesante plantearse la cuestión a la inversa y obtener, para cada muestra, el sentimiento dominante de las palabras TOP halladas en la sección 3.2.1. Ello permitirá deducir el tono predominante de cada muestra.

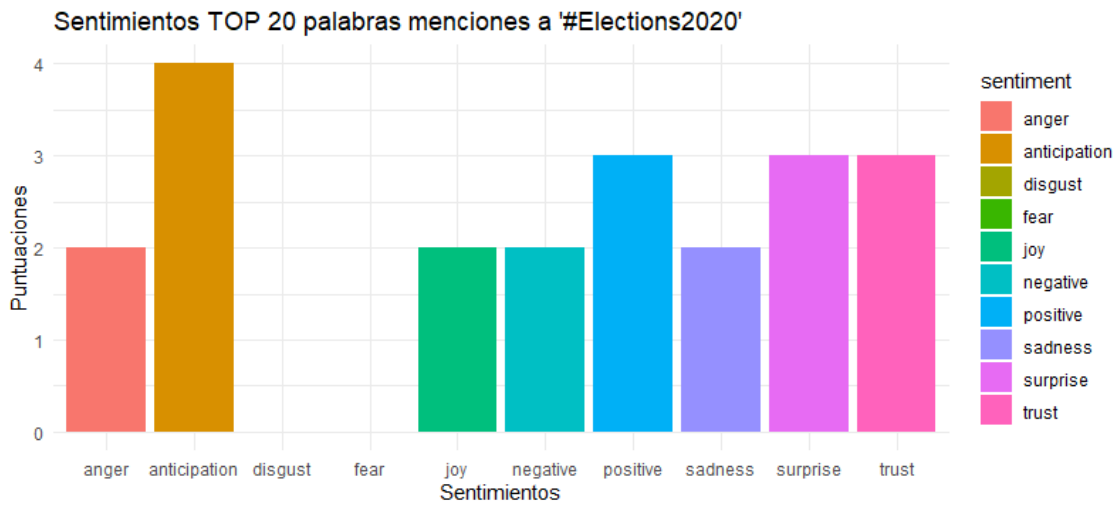


Figura 45. Puntuación sentimientos TOP 20 palabras encontradas en el conjunto de datos '#Elections2020'.  
Fuente: Elaboración propia

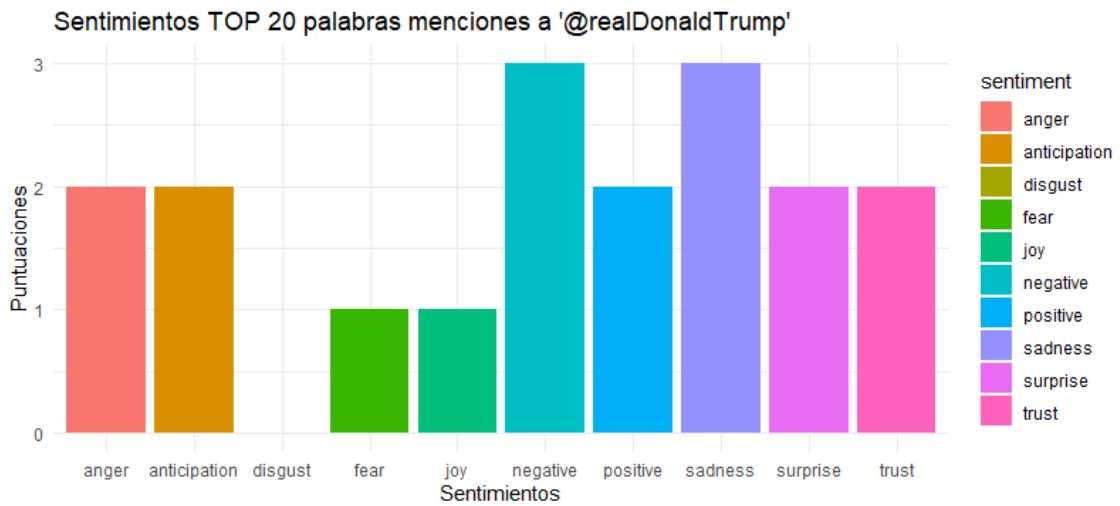


Figura 46. Puntuación sentimientos TOP 20 palabras encontradas en el conjunto de datos '@realDonaldTrump'.  
Fuente: Elaboración propia

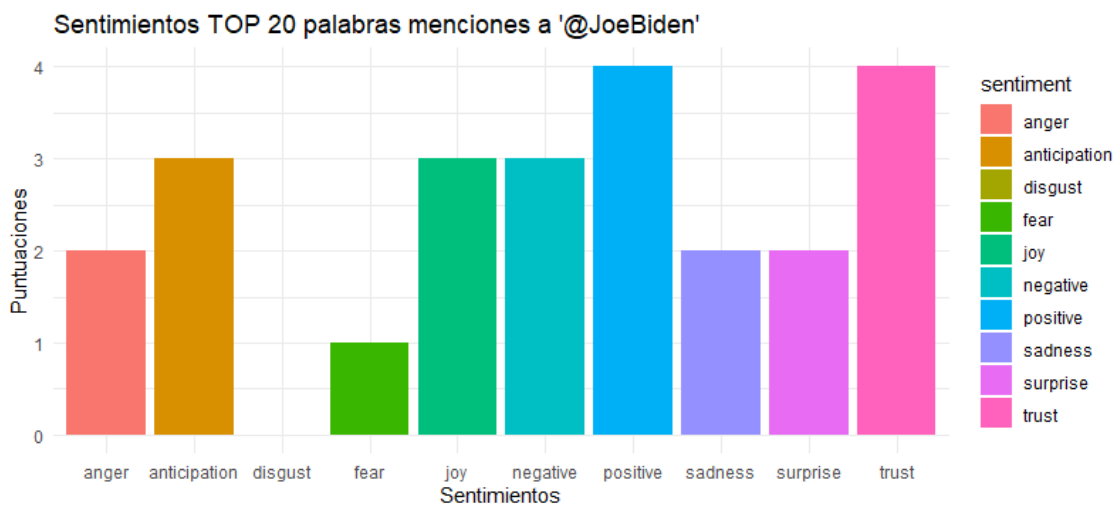


Figura 47. Puntuación sentimientos TOP 20 palabras encontradas en el conjunto de datos '@JoeBiden'.  
Fuente: Elaboración propia



### 3.2.3.1 Análisis e interpretación de resultados

Una vez obtenidos los resultados del análisis de sentimientos, conviene hacer un análisis e interpretación de los mismos. Empezando por el hashtag neutro “Elections2020”, el léxico “bing”, omitiendo la palabra “trump”, a la que concede la máxima puntuación positiva, da un global de 135 puntos negativos frente a 225 puntos positivos en su top 15 de palabras clasificadas como negativas o positivas. En cuanto al léxico “nrc”, prescindiendo de la palabra “vote”, a la que concede la misma puntuación negativa que positiva, da un global de 169 puntos negativos frente a 361 puntos positivos en su top 15 de palabras negativas o positivas. Este último léxico permite entrar en detalles de tipo de sentimiento y otorgarles una puntuación. En este sentido, se observan puntuaciones muy altas en sentimientos como “anticipation”, “joy” o “trust” y puntuaciones notablemente más bajas en sentimientos como “disgust” o “fear”. En cuanto al diagrama de barras de sentimientos encontrados, encontramos una brecha de más de mil puntos entre las puntuaciones de sentimientos positivos y negativos encontrados. Las puntuaciones “fear” y “joy” van un poco a la par y “trust” supera los 2000 puntos. Atendiendo a las puntuaciones que se han obtenido en los gráficos del hashtag neutro, parece que esta campaña electoral despierta más sentimientos positivos que negativos. La trascendencia de las mismas, por la situación económica y social del país, así como la perspectiva de un posible cambio de ciclo, pueden ser las razones de la sentencia anterior. No obstante, no podemos afirmarlo a ciencia cierta. Si bien hay palabras ubicadas en uno y otro sentido, en función del tono de la frase el sentido podría dar un vuelco. Este podría ser el caso de palabras como “chaos” y “crisis”, en principio negativas, pero que podrían ser usadas en críticas constructivas o incluso plegarias o llamamientos a una situación mejor. Esto también quedaría reflejado en las palabras frecuentes. El tono dominante en el TOP 20 de palabras asociado a este conjunto de datos es de anticipación y positivismo.

Siguiendo con las menciones al perfil de Donald Trump, el léxico “bing” da una puntuación global de sentimientos negativos de 2863, frente a una de 5102 para sentimientos positivos (3071 puntos si excluimos del análisis la palabra “trump”) en su top 15 de palabras. En cuanto al léxico “nrc”, en su top 15 de palabras excluyendo, la palabra “vote”, otorga una puntuación global de 3773 en sentimientos negativos y de 4757 puntos en sentimientos positivos. Su diagrama de barras correspondiente a sentimientos encontrados muestra una diferencia menos acusada entre sentimientos positivos y negativos, aunque el balance global vuelve a ser positivo. No obstante, el tono dominante asociado al TOP 20 de palabras correspondiente a este conjunto de datos es de tristeza y negatividad.

En cuanto a las menciones al perfil de Joe Biden, el léxico “bing”, en su top 15 de palabras, concede 10232 puntos a sentimientos negativos y 9545 puntos a sentimientos positivos. Por su parte, el léxico “nrc” da en su top 15 de palabras 14562 puntos a sentimientos negativos y 14765 puntos a sentimientos positivos. Su diagrama de barras correspondiente a sentimientos encontrados muestra un saldo a favor de sentimientos positivos algo mayor que en el diagrama de barras correspondiente a las menciones a Donald Trump. Ambos diagramas de barras de sentimientos encontrados muestran resultados muy parecidos, pero el conjunto formado por las menciones a Biden supera en puntuación al de Trump en sentimientos como “joy”, “trust” y “anticipation”. Las menciones a Trump superan ligeramente a las de Biden, no obstante, en sentimientos de disgusto. El tono dominante asociado al TOP 20 de palabras para este conjunto de datos es de confianza y positividad.

A modo de conclusión, si bien el contenido de los tuits con el hashtag de las elecciones muestra una tendencia global hacia sentimientos positivos, las menciones a cada uno de los principales candidatos observan diferencias poco significativas en



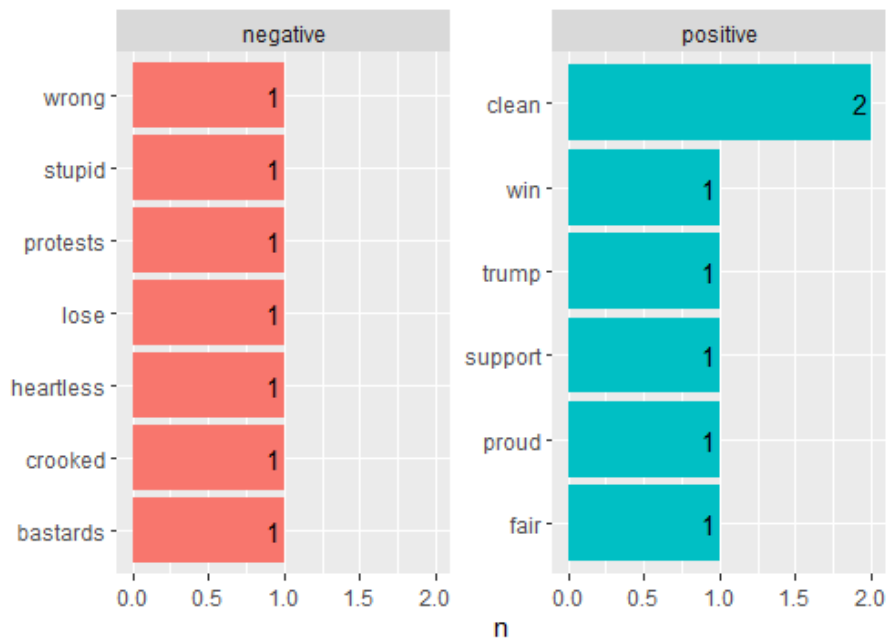


Figura 49. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “#Elections2020”. Léxico “bing”. Fuente: Elaboración propia

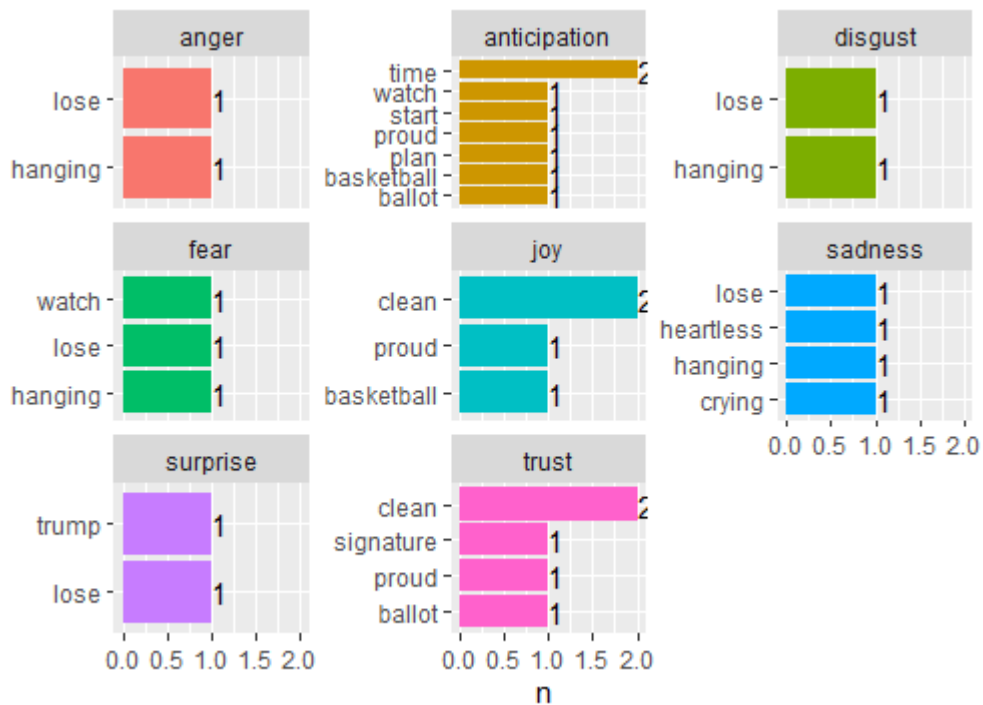


Figura 50. Palabras clasificadas por tipo de sentimiento. Usuarios con más menciones de “#Elections2020”. Léxico “nrc”. Fuente: Elaboración propia

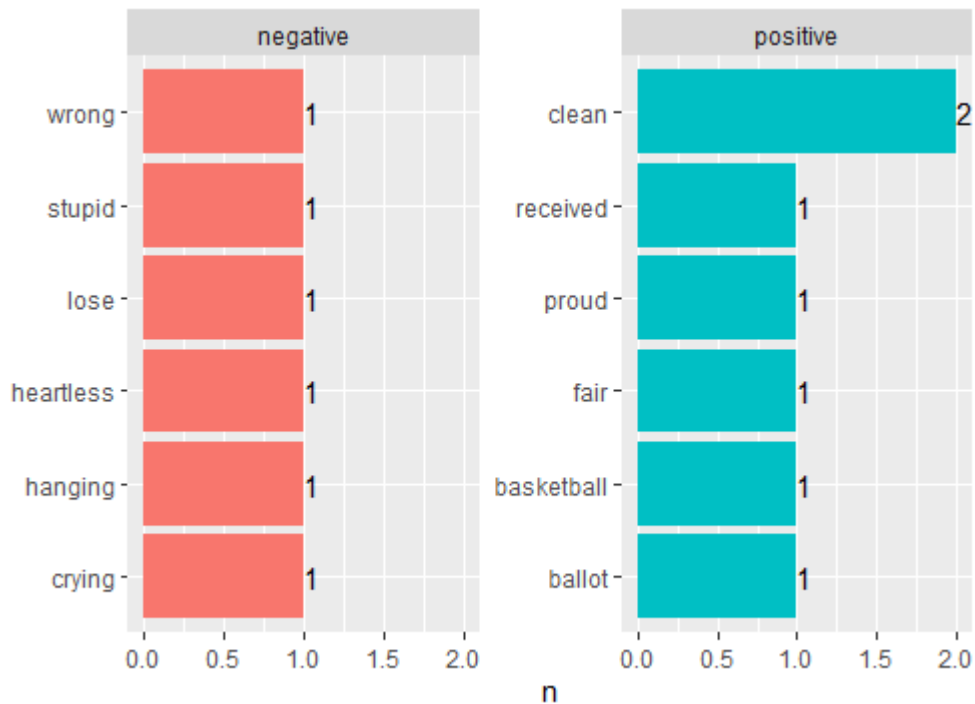


Figura 51. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de "#Elections2020". Léxico "nrc". Fuente: Elaboración propia

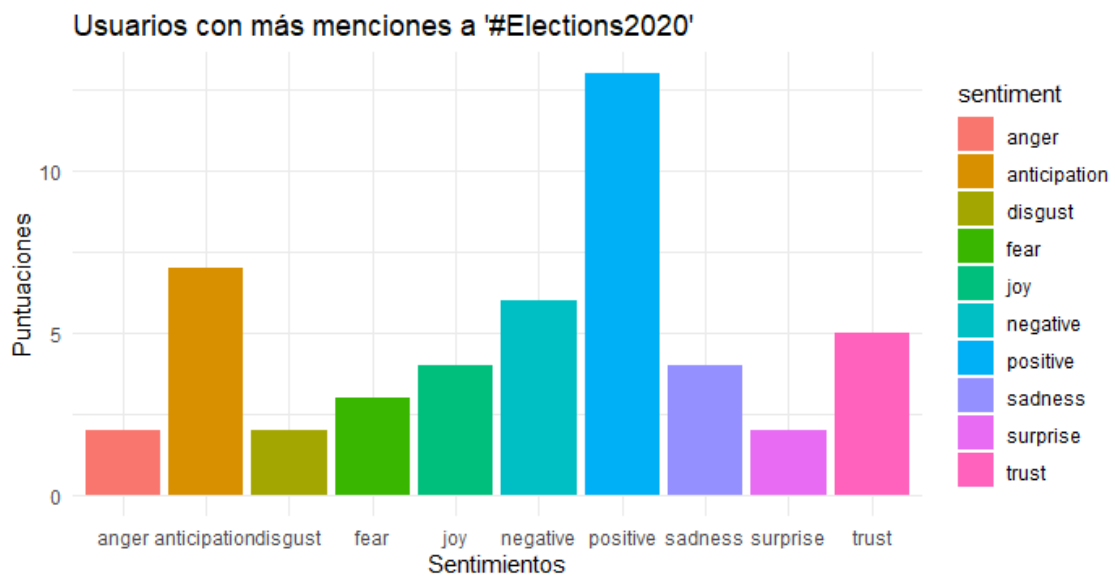


Figura 52. Puntuación sentimientos encontrados usuarios con más menciones a "#Elections2020". Fuente: Elaboración propia

Atendiendo a los resultados, los usuarios con más menciones al hashtag "#Elections2020" emplean, como es lógico, algunas palabras que ya aparecieron en el top 20 de palabras frecuentes para dicho hashtag (es el caso de "time", "trump", "election") y otras nuevas en un tono claramente beligerante ("ChristiansAgainstTrump" y "republicansagainstTrump". El análisis de sentimientos, no obstante, muestra un tono algo más positivo. Podemos mostrar algunos textos de hashtags representativos, así como menciones y expresiones:

**"Heartless bastards! You ALL!!! @realDonaldTrump @stevenmnuchin1 @WhiteHouse @SenateGOP #ChristiansAgainstTrump #DumpTrump2020 #Elections2020 #RepublicansAgainstTrump #StimulusNow"**

**"@realDonaldTrump NOTHING TO BE PROUD ABOUT!! CROOKED WAYS ARE YOUR SIGNATURE!! @realDonaldTrump @WhiteHouse #Elections2020 #ChristiansAgainstTrump #RepublicansAgainstTrump"**

**"All the pundits that have opinions about who will win and why, must say they were wrong, or right, when the race is called\n\nThat would be fair\n\nI honestly think they make their theories up.\n\n#Elections2020"**

**"So Democratic Governors are stopping Trump rallies because of covid but support protests and basketball championship parades?\n\nSee the stupid?\n\n#Elections2020"**

**"The hanging chad of 2020 will be, if the ballot was received in time, and what exactly is the definition of time! :))\n\n#Elections2020"**

Analizando contenidos, encontramos un insulto explícito que no se sabe bien a quién va dirigido; un reproche a los supuestos expertos en elecciones que vaticinan quién va a ganar, ya que podrían estar equivocados, así como una crítica a los gobernadores demócratas en cuanto a la temática covid: impiden mítines de Donald Trump pero apoyan marchas y desfiles supuestamente multitudinarias por el campeonato de baloncesto. Por último, hay quien escribe que una cuestión decisiva en las elecciones va a ser el recuento de votos a tiempo, así como la correcta definición de tiempo en este tipo de eventos.

Seguimos analizando el discurso de los que han mencionado más veces el perfil de usuario "@realDonaldTrump" (se trata de los mensajes de los usuarios "@stan4\_america" y "@ABCDEFGFG14766923").

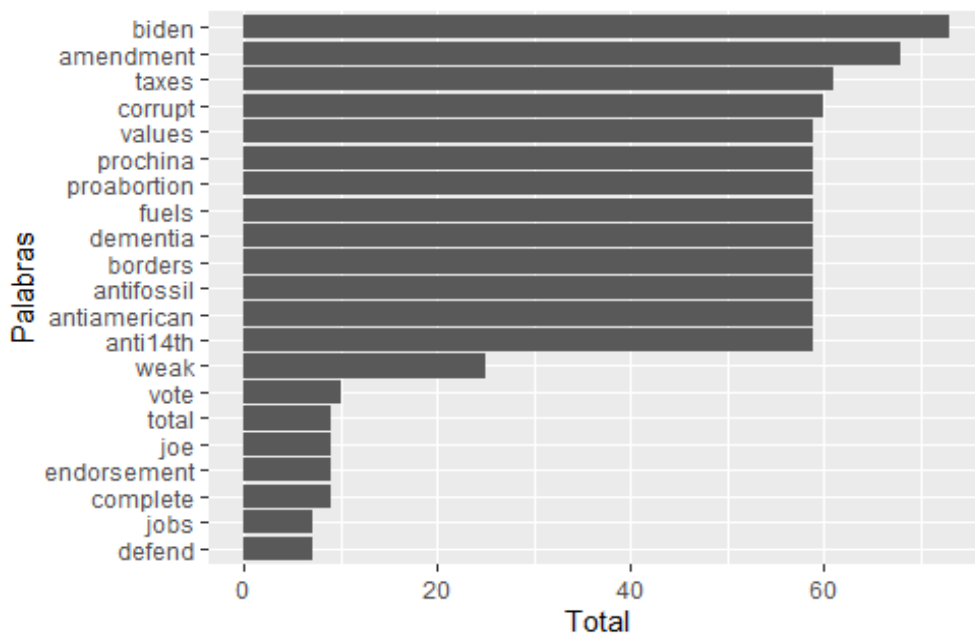


Figura 53. Palabras más frecuentes usuarios con más menciones de "@realDonaldTrump".  
Fuente: Elaboración propia

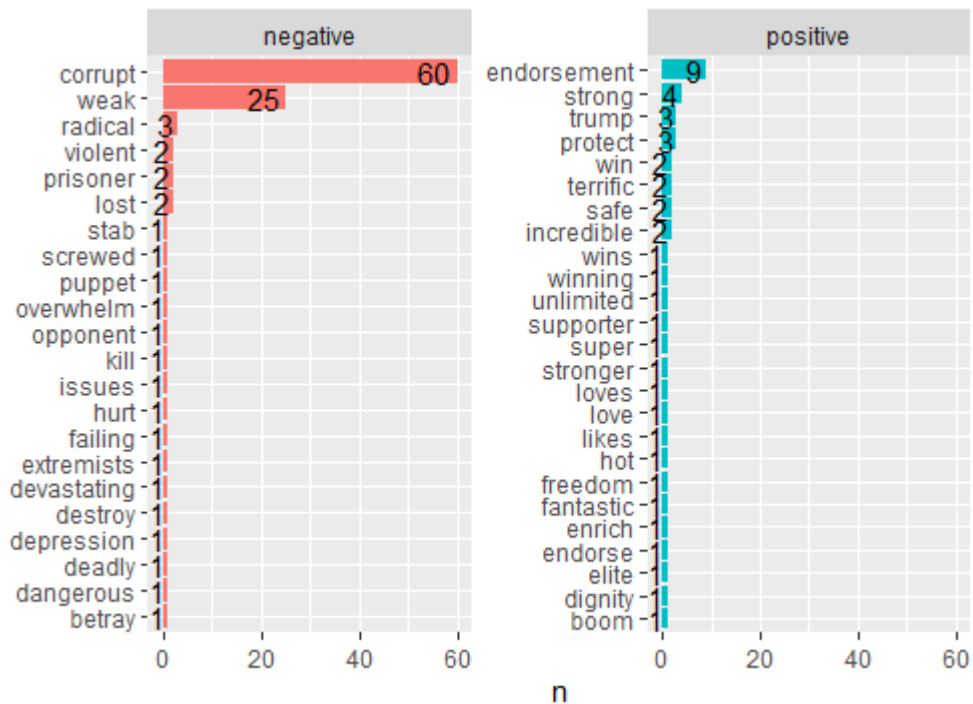


Figura 54. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@realDonaldTrump”. Léxico “bing”. Fuente: Elaboración propia

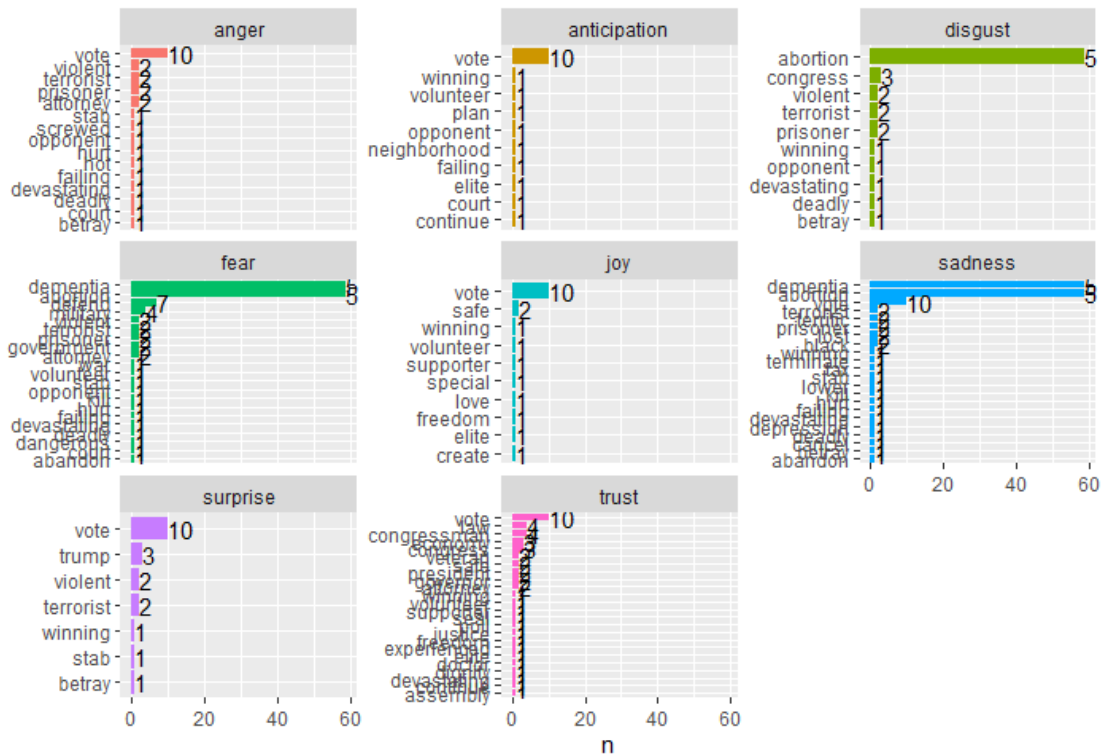


Figura 55. Palabras clasificadas por tipo de sentimiento. Usuarios con más menciones de “@realDonaldTrump”. Léxico “nrc”. Fuente: Elaboración propia

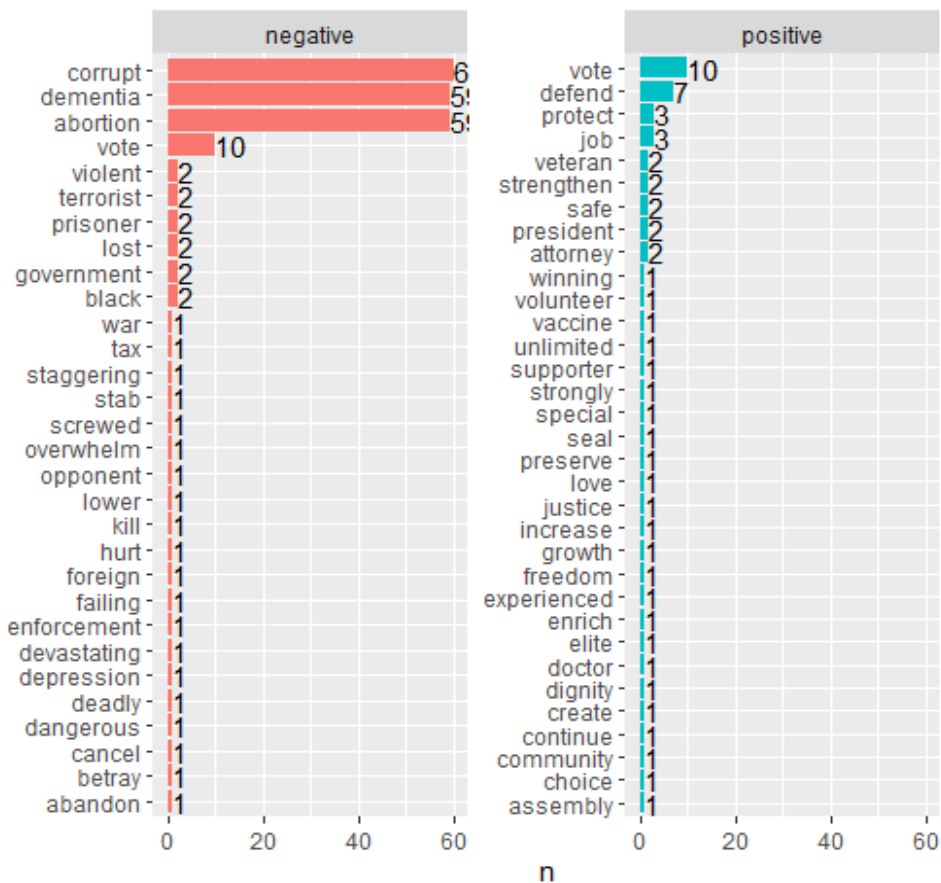


Figura 56. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@realDonaldTrump”. Léxico “nrc”. Fuente: Elaboración propia

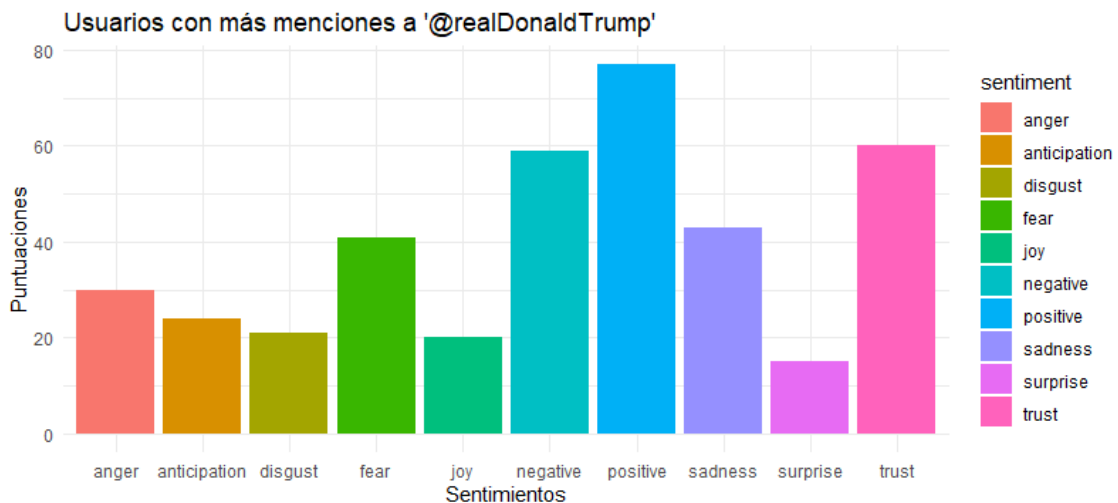


Figura 57. Puntuación sentimientos encontrados usuarios con más menciones a “@realDonaldTrump”. Fuente: Elaboración propia

Los resultados obtenidos son coherentes con los que se obtuvieron al analizar la palabra clave “@realDonaldTrump”, si bien se advierte que el discurso de los usuarios que más la han mencionado se vuelve todavía de corte más conservador o republicano. Hallamos palabras que advierten de corrupción (“corrupt”), aluden a China o al aborto (“prochina”, “proabortion”) o bien intentan prevenir de futuros incrementos de tasas (“taxes”). Encontramos más puntuación de sentimientos positivos que negativos en el diagrama de barras de sentimientos, pero ambos

léxicos, “bing” y “nrc”, dan más puntuación global negativa que positiva. Veamos también qué dicen algunos hashtags a modo de ejemplo:

***“A Biden win would mean the end of Fracking in Pennsylvania, Texas, and everywhere else. Millions of jobs would be lost, and Energy prices would soar.” @OANN And we would no longer be Energy Independent!!!”***

***“Joe Biden is a corrupt politician who has done nothing but betray you for 47 long years - selling out to lobbyists, China, big pharma & foreign countries. He will stab you in the back. He only cares about political power & he will use that power to hurt you & enrich his cronies! <https://t.co/8B9kZ5sdEJ>”***

***“President Obama meant a lot to Black people, but he didn’t do a thing for Black people”. Vernon Jones, Georgia State Representative, Trump supporter. @ffweekend”***

***“MAKE AMERICA GREAT AGAIN!”***

***“Biden and the Radical Left are asking for the largest tax hike in the history of the U.S. Remember that when you VOTE!”***

El contenido de estos mensajes es claro: las consecuencias negativas de un supuesto voto al candidato demócrata Joe Biden. Estas consecuencias se concretan en pérdidas de puestos de trabajo, guiño a las grandes compañías farmacéuticas chinas y aumento de impuestos. Se trata, por tanto, de un calco del mensaje habitual de Donald Trump en Twitter durante la campaña electoral. Los que más mencionan su nombre de perfil, por consiguiente, son bastante afines a su partido. Nótese que uno de los mensajes es el eslogan de campaña del propio Donald Trump.

Por último, se analiza el discurso de los usuarios con más menciones a “@JoeBiden” (se trata de los mensajes de los usuarios “@PAmegazo” y “@Michael32957419”).

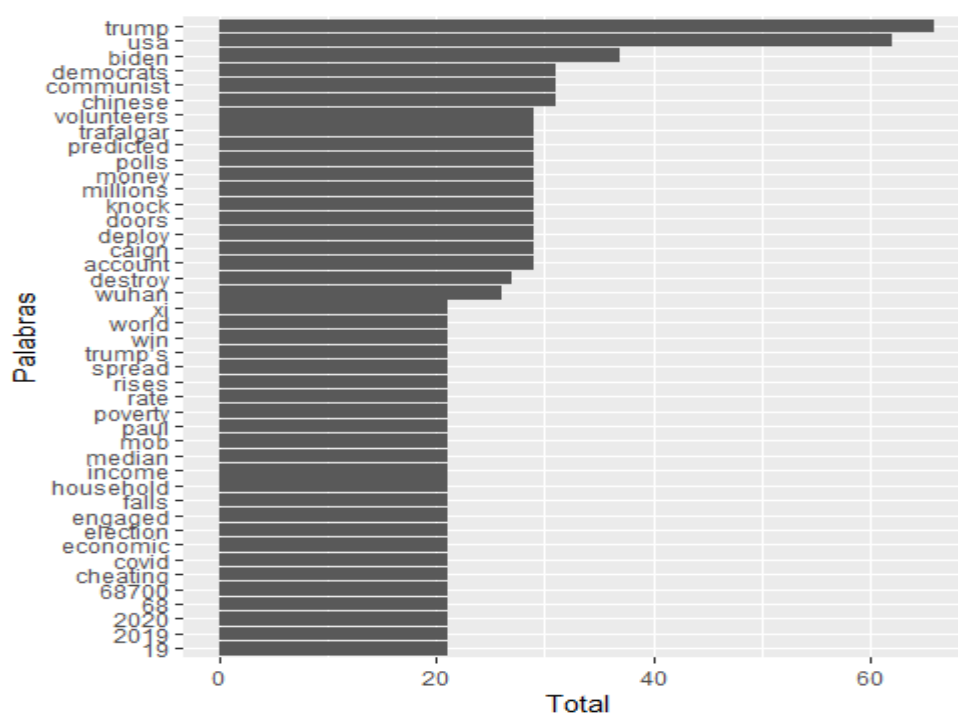


Figura 58. Palabras más frecuentes usuarios con más menciones de “@JoeBiden”.  
Fuente: Elaboración propia



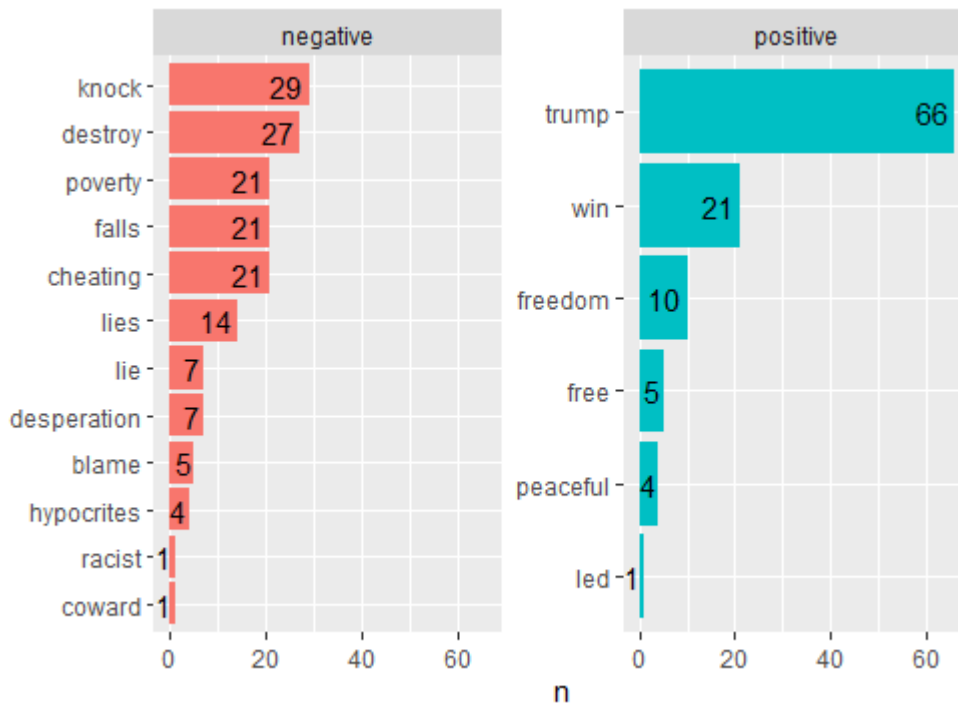


Figura 59. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@JoeBiden”. Léxico “bing”. Fuente: Elaboración propia

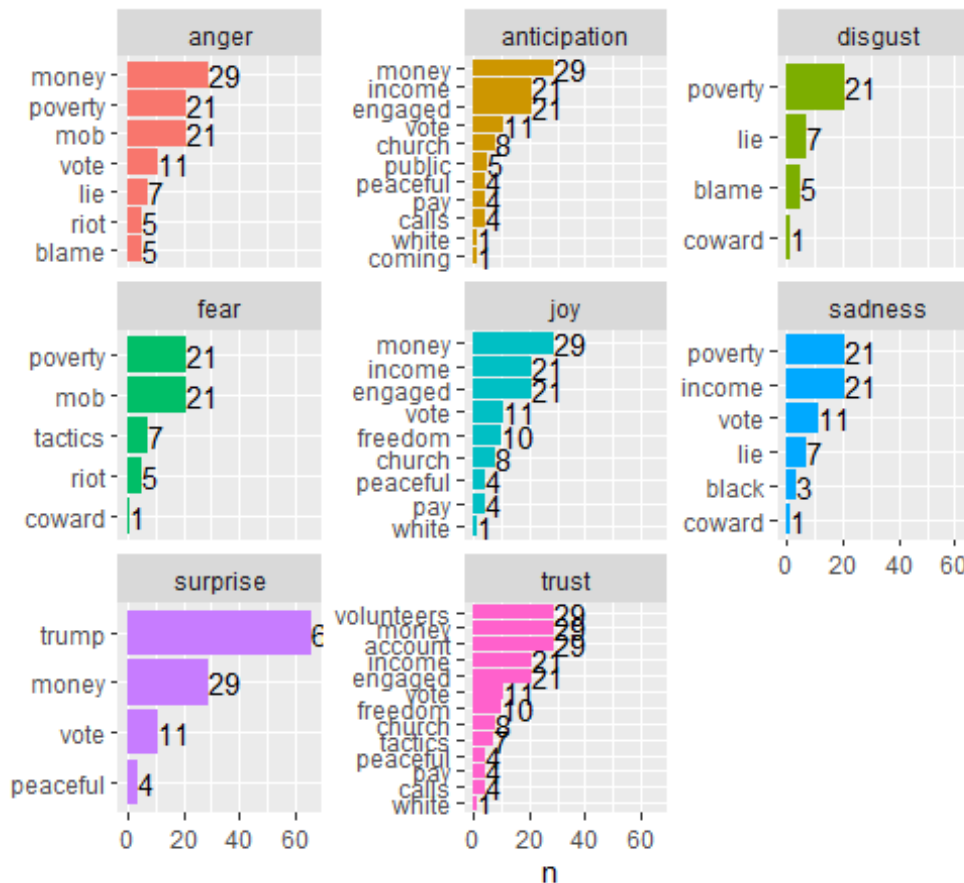


Figura 60. Palabras clasificadas por tipo de sentimiento. Usuarios con más menciones de “@JoeBiden”. Léxico “nrc”. Fuente: Elaboración propia

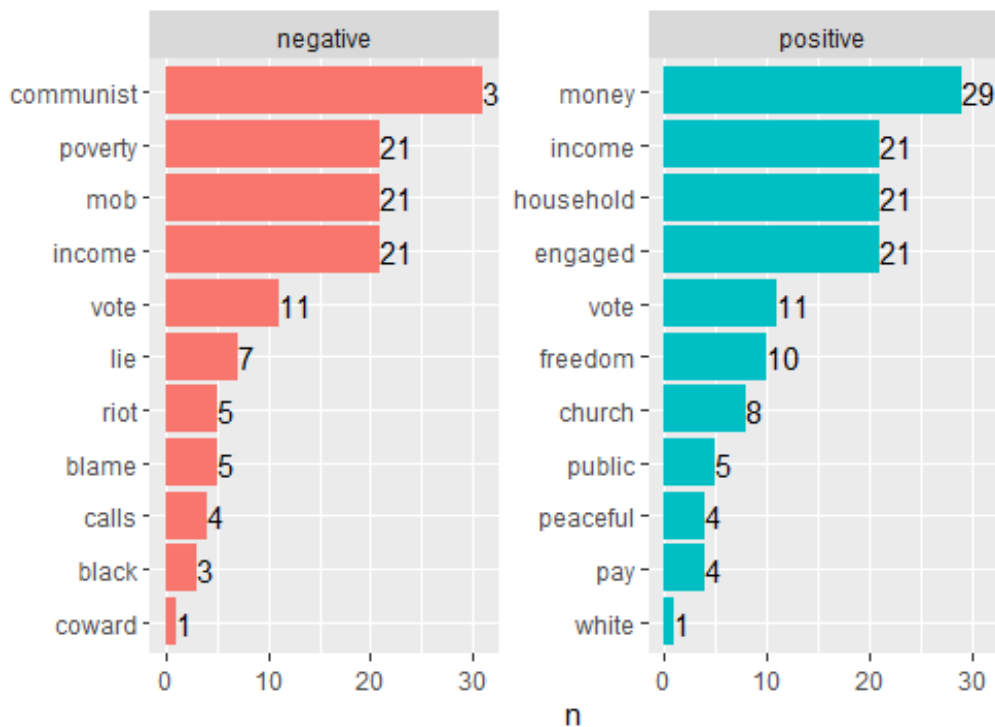


Figura 61. Palabras clasificadas por sentimiento binario. Usuarios con más menciones de “@JoeBiden”. Léxico “nrc”. Fuente: Elaboración propia

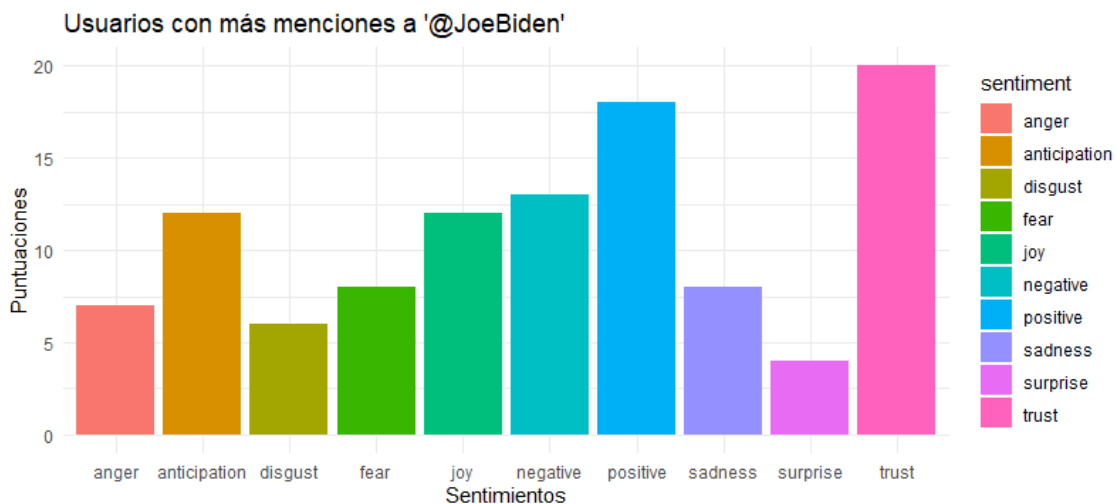


Figura 62. Puntuación sentimientos encontrados usuarios con más menciones a “@JoeBiden”. Fuente: Elaboración propia

Ambos léxicos dan más puntuación positiva que negativa a las palabras encontradas. Los resultados son coherentes con los obtenidos a partir de “@JoeBiden” y se observan palabras nuevas en el discurso, como “communist” o “wuhan”. A continuación, se muestran algunos tuits:

***"@hydehunter2017 @JoeBiden Which church does Trump attend? Is it Christlike to teargas & run over peaceful protesters to go & hold the Bible upside down in front of a church? Is it christlike to sleep with pornstars & pay them off for political power? Jesus calls it, "Hypocrites".***

**"@MiBoomer @JoeBiden Desperation & Scaremongering is Trump's tactics, but the problem is Biden isn't Hillary because many people know Joe & will not buy Trump's lies! Statistics don't lie! Only ghosts will believe Trump's lies!"**

**"@JoeBiden Median U.S. household income rises 6.8% to \$68,700 in 2019, poverty rate falls for fifth year. Paul D. USA TODAY. But democrats J. B. engaged Chinese Communist of Xi mob to spread Wuhan COVID -19, destroy economic in USA and cheating the the whole world in 2020 to win election."**

**"@JoeBiden We shall vote Trump again and let him stop engagement between Chinese communist party and democrats in past 42 years. Trump created more jobs for Americans and restore economics in the USA by adding more sanctions to CCP, N Korea & Russia who against human right & freedom."**

El contenido de estos mensajes es más ambiguo que el de los que más han mencionado el perfil de Trump. Detectamos ataques a ambos candidatos electorales. El primer tuit es un ataque visceral a Donald Trump, que dice llamarse cristiano pero que ataca con gas lacrimógeno a los manifestantes y se acuesta con actrices porno. El segundo tuit sigue en una línea parecida y acusa a Trump de embustero al que las estadísticas acabarán poniendo en evidencia. No obstante, los dos últimos tuits son un ataque a la causa demócrata. Se les acusa de involucrarse con la mafia comunista china para difundir el covid y hundir la economía americana, así como de hacer trampas para ganar las elecciones. El último tuit es una declaración de intenciones: se votará a Trump para que siga creando empleo y aplique sanciones a China y a Rusia, medidas necesarias para garantizar derechos humanos y libertades. A la vista de los resultados, no sabría decir a ciencia cierta de qué lado se posicionan los que más han mencionado el perfil de Joe Biden, ya que encuentro opiniones contrapuestas en los tuits que se han seleccionado para analizar sus contenidos. Lo que sí que puedo constatar es que los que más han usado la mención a Joe Biden lo han interpelado tanto positiva como negativamente.

Los tuits que incluyen el Hashtag neutro se distinguen en cuanto a temática y tono de los que mencionan a los candidatos. Si bien en estos últimos el tono habitual es de crítica y de advertencia, en el primer caso, a menos que se trate de un insulto, se aborda la temática covid, posibles vaticinios de resultados y se plantea cómo se debe gestionar el tiempo para un procesamiento correcto de los votos. No es tan habitual hallar mensajes de crítica, reproche o ataque.

Para terminar esta sección, el siguiente *word cloud* establece una comparativa entre las palabras más usadas por los que más han mencionado a "@realDonaldTrump" y los que más han mencionado a "@JoeBiden". El color rojo corresponde al bando republicano y el azul al demócrata. Nótese la presencia de palabras nuevas tanto en el discurso republicano ("prochina", "proabortion", "taxes"), como en el demócrata ("communist", "wuhan") respecto a la nube de temática similar que se mostró en la Figura 32. Esta nube de palabras nos muestra una vista rápida de los temas o aspectos más repetidos en cada uno de los discursos de los que han protagonizado más menciones.





**Donald J. Trump** ✓  
@realDonaldTrump



Joe Biden is the candidate of rioters, looters, arsonists, gun-grabbers, flag-burners, Marxists, lobbyists, and special interests. I am the candidate of farmers, factory workers, police officers, and hard-working, law-abiding patriots of every race, religion and creed! #MAGA

[Traducir Tweet](#)

11:35 p. m. · 1 nov. 2020 · Twitter for iPhone

Figura 64. Ataque personal a Joe Biden y elogio a su propia candidatura.  
Fuente: <https://twitter.com/realDonaldTrump/status/1323030938952060928>



**Donald J. Trump** ✓  
@realDonaldTrump



If you vote for Biden, your kids will not be in school, there will be no graduations, no weddings, no Thanksgiving, no Christmas, and no Fourth of July!

[Traducir Tweet](#)



De **Team Trump (Text VOTE to 88022)** ✓

Figura 65. Advertencias a los votantes de Joe Biden.  
Fuente: <https://twitter.com/realDonaldTrump/status/1321594088676761601>



**Donald J. Trump** ✓  
@realDonaldTrump



Thank you Goodyear, Arizona! Get out and VOTE —  
Together, we are going to MAKE AMERICA GREAT  
AGAIN!

Traducir Tweet



0:05 1,6 M reproducciones

De **Dan Scavino** 🇺🇸 🇺🇸 ✓

12:19 a. m. · 29 oct. 2020 · Twitter for iPhone

Figura 66. Instando a la población de Arizona a votar y proclamando su eslogan.  
Fuente: <https://twitter.com/realDonaldTrump/status/1321592371570569217>



**Donald J. Trump** ✓  
@realDonaldTrump



Joe Biden is a corrupt politician and he's  
COMPROMISED. China is desperate for Biden to win  
because if Biden Wins, CHINA WINS – and China will  
OWN AMERICA. This corruption is EXACTLY why I  
decided to run for President in the first place...

Traducir Tweet

4:06 a. m. · 24 oct. 2020 · Twitter for iPhone

Figura 67. Acusando a Biden de político corrupto que va a dejar el país en manos de China.  
Fuente: <https://twitter.com/realDonaldTrump/status/1319822456480100354>

En definitiva, Trump insta a los norteamericanos a votar, ataca de forma feroz a su adversario Biden y se reivindica como única opción válida ante los peligros que amenazan la continuidad de unos valores conservadores que considera la esencia y la grandeza de la nación americana. Apuesta claramente por la polarización y el

enfrentamiento, en el que apenas se atisban narrativas constructivas. Su discurso, además, es claramente nacionalista y proteccionista. Esta estrategia está en perfecta sintonía con los que más le mencionan en Twitter, que incluso radicalizan su discurso cuando se trata de atacar a su adversario Biden. El discurso de estos últimos, por consiguiente, es un calco del suyo propio.

Por su parte, el candidato demócrata Joe Biden presenta un perfil más bajo o actividad menos vertiginosa en redes sociales que su contrincante Donald Trump. Su usuario es “@JoeBiden” y en su imagen de perfil muestra un rostro sonriente, a diferencia de la de Trump, que hace gala de un gesto más serio y adusto. Su biografía es más extensa que la de su adversario, desvela más sobre su personalidad y vida privada, lo que denota un cierto interés en darse a conocer. “Senador, vicepresidente, candidato a la presidencia en 2020, marido de @DrBiden, orgulloso padre y abuelo. Amante del helado, la aviación y @Amtrak”, podemos leer en su perfil. Si bien desde que abrió su cuenta en 2007 el número de tuits y su número de seguidores es más discreto que el de Trump, así como su porcentaje de retuits, en esta campaña electoral, no obstante, ha publicado más tuits que su adversario (véase el gráfico con la distribución de tuits capturados mostrado anteriormente). Cada vez que comparte una publicación, suele ser de @TeamJoe (la cuenta oficial de su campaña), de la candidata a la vicepresidencia @KamalaHarris o de su esposa, @DrBiden. Una diferencia notable con Trump es que es más proclive a realizar menciones. En ellas, se nombra a él mismo y a la fundación Biden (@bidenfoundation). También recurre a las etiquetas en un porcentaje mayor que Trump, siendo las más comunes “#demdebate” (debate demócrata) y “#demconvention” (convención demócrata). Además, sus publicaciones en Twitter están repletas de palabras alentadoras y reconfortantes. Mientras que entre las 50 que más repite el actual presidente solo cuatro son positivas (genial, gracias, orden y mejor), entre las de Biden se cuentan hasta 11, y la mitad de ellas se repiten más de diez veces. El candidato demócrata tuitea con mucha frecuencia términos como juntos, ayuda, futuro, construir, cambio y proteger. Tuitea más frecuentemente la palabra América y derivados que el propio Donald Trump, que ha centrado buena parte de su estrategia en la confrontación. Biden se muestra como un candidato optimista, capaz de devolver la concordia al pueblo estadounidense. Podría decirse que la estrategia de Biden, a partir del seguimiento de sus tuits en esta campaña, ha sido la de intentar mantener un perfil bajo, buscando no cometer errores y teniendo como eje central de su discurso echar a Trump. El candidato demócrata ha querido centrar su discurso en lo malo de su oponente y en la visión social de su programa. Otra baza importante de su estrategia es la introducción de Kamala Harris como candidata a la vicepresidencia, quien ofrece más activos que él mismo, tanto por frescura como por sectores sociales, con los que puede llegar a conectar. Veamos algunos ejemplos:



**Joe Biden** ✓  
@JoeBiden



President Obama and I left President Trump a playbook on how to deal with pandemics. He flat-out ignored it. And we're suffering every day because of it.

[Traducir Tweet](#)

8:25 p. m. · 31 oct. 2020 · TweetDeck

Figura 68. Crítica a Trump por su mala gestión sobre la pandemia.  
Fuente: <https://twitter.com/JoeBiden/status/1322620606098137088>



**Joe Biden** ✓  
@JoeBiden



Presidents lead by example.

Mask up.

[Traducir Tweet](#)



12:20 a. m. · 1 nov. 2020 · Twitter Web App

**Figura 69. Compromiso y seriedad frente a la pandemia.**  
Fuente: <https://twitter.com/JoeBiden/status/1322679745662881798>



**Joe Biden** ✓ @JoeBiden · 25 oct.



Folks, it's go-time. There are just 10 days left until Election Day, and we need your help.

Chip in \$10 to help bring this across the finish line:



Chip in to elect Joe Biden and Kamala Harris

Join us! Contribute today.

[secure.actblue.com](https://secure.actblue.com)

**Figura 70. Solicitando un donativo a sus votantes, junto a Kamala Harris.**  
Fuente: <https://twitter.com/JoeBiden/status/1320137745696649216>





Figura 71. Ejemplo de mensaje positivo. Promesa de tiempos mejores que debe cumplirse.  
Fuente: <https://twitter.com/JoeBiden/status/1320122899655634944>

No puedo afirmar con la misma rotundidad que en el caso de Trump que los que más han mencionado a Biden estén alineados con su estrategia electoral, ya que en los discursos analizados, si bien hay mensajes con una clara tendencia de echar a Trump, también encontré críticas y ataques a la causa demócrata. No obstante, en el análisis de sentimientos de sus menciones hallé palabras en la misma línea alentadora y reconfortante de Biden, como por ejemplo “freedom” , “peaceful” o “win”.

Una vez conseguido el propósito de esta sección y desveladas cada una de las estrategias, conviene citar también que ambos candidatos tienen en común que en sus discursos han utilizado de forma recurrente la palabra “votar” (algo de esperar tratándose de una campaña electoral), así como el apellido de su contrincante (Trump nombra muchas veces a Biden y Biden a Trump). No obstante, hay temas que no comparten en absoluto. Mientras Biden hace alusión a la pandemia con términos como “mascarilla” o “Covid-19”, el coronavirus no forma parte del eje central del discurso en los tuits de Trump. En su lugar, Trump nombra más veces conceptos relacionados con la economía, como “trabajos” o “impuestos”.

## 4. Obtención de resultados y conclusiones

Una vez finalizado el diseño e implementación del trabajo, en esta sección se enumeran y valoran los resultados obtenidos y se extraen las conclusiones de estos. Los resultados obtenidos representan una aportación original, la cual actualiza los conocimientos que hubiese hasta el momento sobre usos de Twitter durante campañas electorales que tocan a su fin. En particular, se ha estudiado el uso de Twitter durante los últimos días de las elecciones presidenciales de EEUU en 2020.

### 4.1 Obtención de resultados

Mediante la captura y procesamiento de tuits a partir del entorno de desarrollo RStudio, he podido cumplir de forma satisfactoria el primer objetivo principal de este proyecto, la obtención, limpieza y radiografía de los mensajes generados y difundidos en Twitter durante los últimos días de la campaña electoral correspondiente a las presidenciales de 2020 en EEUU. No obstante, si bien el número de tuits para la presente investigación es aceptable, podría haber sido mucho más copioso y haber

contemplado más hashtags neutros representativos que ayudasen a enriquecer el trabajo aún más si cabe. Si bien inicialmente me había propuesto realizar el seguimiento de la campaña electoral durante la quincena previa al día de las elecciones (3 de noviembre), limitaciones tecnológicas debidas a Twitter y a la función del lenguaje R *search\_tweets* acotaron la investigación entre los días 30 de octubre y 2 de noviembre de 2020. Otro propósito fallido fue sacar toda la información del timeline y las características de los usuarios @realDonaldTrump y @JoeBiden mediante la función *userTimeline()* de R. Esto hubiese facilitado el seguimiento de su actividad en Twitter a lo largo del tiempo en periodos mucho más espaciados y mejor definidos que los que se muestran en este trabajo en cuanto a distribución de tuits por horas, un poco mermados al restringirse sólo a un día de la investigación (30 de octubre). En este sentido, otra cuestión crucial fue realizar toda la captura de tuits de una única vez y durante el mismo día. Haber realizado capturas durante varios días para unir después todos los datos hubiese enriquecido los resultados y hubiese facilitado un análisis de evolución temporal más completo. Además, cada una de estas capturas podría haber incorporado algún que otro hashtag neutro adicional al que se ha usado en este trabajo.

Por otra parte, el resultado final cumple también con el segundo objetivo principal de este trabajo, esto es, se ha realizado un análisis de las publicaciones correspondientes tanto a los que mencionan a los principales candidatos electorales como a los que usan el hashtag neutro que alude a la campaña electoral y se ha detectado lo dicho por parte de los que mencionan cada conjunto de datos. Se ha realizado un gran ejercicio de análisis y obtención de resultados, capaz de satisfacer la totalidad de los objetivos más específicos o secundarios (detección de palabras más frecuentes, obtención de algún *Word Cloud* y análisis de sentimientos, etc.). No obstante, el camino recorrido ha sido arduo y difícil y no ha estado exento de contratiempos metodológicos. Si bien este trabajo fue concebido para realizar el seguimiento electoral de una campaña electoral en Twitter, dando cierta relevancia a la estrategia electoral seguida por cada uno de los candidatos, ha evolucionado hacia el uso de Twitter durante el final de una campaña electoral. La razón principal de este giro metodológico es que el trabajo fue tomando forma en torno a lo que comentan los que interpelan a los principales candidatos y los que utilizan el hashtag neutro de las elecciones, dando mucho peso a cuestiones como el uso que se hace de los tuits y las posibles diferencias en el tono utilizado para cada uno de los conjuntos de datos. Ello ha potenciado la presencia en este trabajo de palabras más frecuentes y del análisis de sentimientos a partir de diccionarios a tal efecto. La estrategia seguida en Twitter por cada uno de los principales candidatos electorales, finalmente, pasa a un segundo plano en el contexto de este trabajo de ciencia de datos y se compara con lo expuesto por los usuarios que más menciones han realizado a cada uno de ellos con el fin de hallar si estos últimos corresponden a los líderes que mencionan o bien están ideológicamente en las antípodas. En este sentido, los resultados han sido más dispares entre los usuarios que más han mencionado el perfil de Joe Biden que entre los que han mencionado el perfil de Donald Trump. Para estos últimos, se observó que incluso se radicalizaba el tono del discurso del propio candidato, volviéndose si cabe de corte más conservador.

## 4.2 Conclusiones

Habiendo señalado en el subapartado anterior que el resultado final cumple con los principales objetivos planteados, en esta sección, a la vista de los resultados obtenidos, confirmo mi hipótesis inicial de investigación conforme **existen diferencias significativas en el contenido de las menciones para cada conjunto de datos, especialmente entre los perfiles relevantes que más han mencionado a cada uno**

**de los candidatos.** Las razones por las que considero que esta hipótesis queda validada se han expuesto a lo largo de este trabajo, pero se resumen a continuación:

- Las menciones al hashtag neutro “Elections2020”, a diferencia de las menciones a los candidatos electorales, representan el único conjunto de datos que se relaciona con el tema covid19, aportando así algo de información sobre el contexto en que se desarrolla la campaña electoral.
- Mientras que el tono dominante de las palabras TOP para el conjunto de datos “Elections2020” es de anticipación y positivismo, en el caso de “@realDonaldTrump” es de tristeza y negatividad y en el caso de “@JoeBiden” es de confianza y positividad. En este sentido, conviene señalar también que en el análisis de sentimientos de los usuarios con más menciones a “@realDonaldTrump” tanto el léxico “bing” como el “nrc” otorga más puntuación global negativa que positiva.
- El contenido de las menciones al conjunto de datos “@realDonaldTrump” es mucho más claro que el correspondiente a las menciones al conjunto de datos “@JoeBiden”. Si bien el primero no hace más que advertir de las posibles consecuencias negativas de un posible voto a Joe Biden, en el segundo hallamos tanto elogios como ataques a ambos candidatos electorales.
- Si bien el tono habitual de las menciones a los candidatos es de crítica y advertencia, en el caso de las elecciones hallamos contenidos más diversos y es más difícil encontrar mensajes de crítica, reproche o ataque.

Por todo lo expuesto anteriormente, atendiendo a los resultados obtenidos, se concluye que esta investigación cumple con sus principales objetivos iniciales y que se verifica también su hipótesis de partida.

### **4.3 Aspectos susceptibles de mejora**

Esta investigación presenta algunas limitaciones que merecen ser discutidas. La primera es en torno al marco temporal en el que se ha desarrollado el estudio, ya que un espectro cronológico superior hubiese permitido una comparativa de uso de Twitter antes, durante e incluso después de la campaña electoral. Ello, lógicamente, hubiese permitido plantearse nuevas preguntas en el marco de un seguimiento de campaña electoral en Twitter mucho más completo, como por ejemplo si el tono o el mensaje observado entre los que más interpelan a los candidatos varía durante los días de campaña.

Por otra parte, la ausencia de software adecuado necesario para abarcar un mayor corpus de estudio o que ofreciese información estadística que permitiera identificar ciertos resultados ha impedido, por ejemplo, obtener ratios que ayudasen a entender qué mensajes funcionan mejor o determinar las comunidades creadas junto a los perfiles relevantes, así como las posibles relaciones entre las comunidades creadas a nivel morfológico.

Por último, si bien en este trabajo se ha abordado de forma notable el análisis de sentimientos para determinar el tono y la orientación de las menciones, una herramienta de monitorización que ofreciera un análisis de sentimiento para conocer reacciones en la red o contextualizarlas mediante la recopilación de opiniones hubiera permitido, además, conocer el grado de aceptación o controversia generado por las menciones a los candidatos, permitiendo determinar así la actitud de la ciudadanía norteamericana ante los temas abordados en la red en vísperas de elecciones.

#### 4.4 Líneas de trabajo futuras

En esta sección comentaré brevemente las cuestiones relacionadas con el trabajo que podrían ser abordadas en el futuro por otros investigadores.

En este trabajo se han clasificado palabras en función de sentimientos y polaridad para descifrar qué dicen los que hacen menciones a ciertos conjuntos de datos. Si bien esto ya es un gran logro, no obstante, desde el primer momento se ha reconocido la necesidad de poner estos resultados en entredicho en función del contexto, ya que el análisis de palabras individuales no es capaz de diferenciar el doble sentido. Líneas de investigación futuras deberían contemplar la información no textual que acompaña a los mensajes y publicaciones en el análisis de la polaridad. Elementos como vídeos, imágenes, animaciones, *likes* y *retweets* facilitarán aún más la clasificación de sentimientos y permitirán conocer mucho más de lo que se percibe en el texto escrito. Confío que en un futuro no muy lejano se resuelva un tema clásico pendiente del análisis de sentimientos como es la detección del sarcasmo, la ironía, la ambigüedad o la dependencia del contexto. No obstante, es muy posible que también aparezcan nuevos problemas en lo referente al análisis de sentimientos, como por ejemplo el spam de opiniones, mensajes o publicaciones que pretendan manipular el sentimiento global de las personas hacia una determinada cuestión. Tener un control sobre este problema será necesario para garantizar un análisis de sentimientos de cierta calidad. Por otra parte, estoy convencido de que el nivel de detalle en que en este trabajo se han medido los sentimientos también cambiará en el futuro. Dejando atrás la escala binaria de lo positivo y negativo, los nuevos sistemas de análisis podrán saber cómo se sienten las personas y darán el salto desde los sentimientos hasta las emociones. En el futuro, el análisis de sentimientos nos va a abrir las puertas a un mundo tan inquietante como fascinante a partes iguales. En un contexto de elecciones electorales, puede que estudios mucho más completos que la presente investigación sean por fin capaces de predecir el resultado de las mismas a partir de mensajes publicados en redes sociales como Twitter. Atrás quedarán los tradicionales modelos de Machine Learning dependientes de bases de datos previamente etiquetadas para poder ser entrenadas y utilizadas, dando paso a sistemas basados en técnicas de inteligencia artificial capaces de acometer empresas como la predicción de resultados en unas elecciones. Científicos de la Universidad de Granada son pioneros en el uso de este tipo de técnicas y han publicado su trabajo en la revista internacional *IEE Access*. Puede que el futuro inquietante y fascinante a partes iguales antes mencionado esté mucho más cerca de lo que imaginamos.

## 5. Bibliografía / Webgrafía

**Abejón, P. ; Sastre, A. ; Linares, V.** (2012). *FACEBOOK Y TWITTER EN CAMPAÑAS ELECTORALES EN ESPAÑA*. Anuario Electrónico de Estudios en Comunicación Social "Disertaciones" E-ISSN: 1856-9536 disertaciones@urosario.edu.co. Universidad del Rosario. Colombia.

**Alonso-Muñoz, L. ; Casero-Ripollés, A.** (2018). *COMUNICACIÓN DE LOS LÍDERES POPULISTAS EUROPEOS EN TWITTER: CONSTRUCCIÓN DE LA AGENDA Y EFECTO "MÁS ES MENOS"*. El profesional de la información, 2018, noviembre-diciembre, v. 27, n. 6. eISSN: 1699-2407.

**Azorín-Richarte, D. ; Orduna-Malea, E. ; Ontalba-Ruipérez, J.A.** (2016). *Redes de conectividad entre empresas tecnológicas a través de un análisis métrico longitudinal de menciones de usuario en Twitter*. *Revista Española de Documentación Científica*, 39(3): e140.

**Bustos, J. ; Ruiz, F.J.** (2019). *Sobreabundancia comunicativa de los líderes políticos en Twitter. El caso español*. ÁMBITOS. REVISTA INTERNACIONAL DE COMUNICACIÓN. N.º. 45. (2019) | © Universidad de Sevilla. ISSN: 1139-1979 | E-ISSN: 1988-5733. N.º DOI: HTTP://DX.DOI.ORG/10.12795/AMBITOS

**Caldevilla, D.** (2009). *Democracia 2.0: La política se introduce en las redes sociales*. Universidad Complutense de Madrid. *Pensar la Publicidad* 2009, vol. III, nº 2, 31-48. ISSN: 1887-8598.

**Cambria, E. ; Schuller, B. ; Liu, B. ; Wang, H. ; Havasi, C.** (2013). *Knowledge-Based Approaches to Concept-Level Sentiment Analysis*. IEEE INTELLIGENT SYSTEMS.

CEA, M. A. (2001): *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Madrid: Síntesis.

**Congosto, M.L. ; Aragón, P.** (2012). *TWITTER, DEL SONDEO A LA SONDA: NUEVOS CANALES DE OPINIÓN, NUEVOS MÉTODOS DE ANÁLISIS. MÁS PODER LOCAL. ESPECIAL. ESTUDIOS DE CASO*.

**Conover, M.D. et al.** (2011). *Political Polarization on Twitter*. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media

**Cumbicus, O.M.** (2017). *CATEGORIZACIÓN AUTOMÁTICA DE TWEETS SOBRE EL TEMA POLÍTICO ELECTORAL APLICANDO ALGORITMOS DE CLASIFICACIÓN SUPERVISADA*. Facultad de Informática, UPV/EHU. [Tesis de Máster].

**Da Cunha, I.** (2015). *El trabajo de fin de grado y de máster: Redacción, defensa y publicación*. Barcelona: Universitat Oberta de Catalunya.

**ESTADO DEL ARTE DE BIG DATA** (2017). *Oportunidades Industria 4.0 en Galicia*. Xunta de Galicia.

**Fernández, M.** (2018). *Twitter, el pájaro que llevó a Donald Trump a la Casa Blanca*. Barcelona. Universitat Oberta de Catalunya [Trabajo de fin de máster].

**García, C.** (2013). *Estado del arte de los estudios, tesis, artículos o ensayos de redes sociales y Twitter en Latinoamérica, Estados Unidos y Europa*. Guayaquil-Ecuador. Universidad Casa Grande. [Trabajo final para la obtención del Título Licenciado en Diseño Gráfico y comunicación visual].

**García-Ortega, C. ; Zugasti-Azagra, R.** (2018). *GESTIÓN DE LA CAMPAÑA DE LAS ELECCIONES GENERALES DE 2016 EN LAS CUENTAS DE TWITTER DE LOS CANDIDATOS: ENTRE LA AUTORREFERENCIALIDAD Y LA HIBRIDACIÓN MEDIÁTICA*. Universidad de Zaragoza. El profesional de la información, 2018, noviembre-diciembre, v. 27, n. 6. eISSN: 1699-2407

**Gayo-Avello, D. ; T. Metaxas, P.; Mustafaraj, E.** (2011). *Limits of Electoral Predictions Using Twitter*. p. 490-493 . Association for the Advancement of Artificial Intelligence.

**Gayo-Avello, D.** (2012). *No, You Cannot Predict Elections With Twitter*. University of Oviedo, Spain. 1089-7801/12/\$31.00 © 2012 IEEE. Published by the IEEE Computer Society..

**González, C.** (2015). *CLASIFICACIÓN AUTOMÁTICA DE TEXTO PARA EL SEGUIMIENTO DE CAMPAÑAS ELECTORALES EN REDES SOCIALES*. Universidad Carlos III de Madrid. [Trabajo fin de grado].

**Guerrero, M.V.** (2017). *LOS TUITSTARS: LÍDERES DE OPINIÓN EN TWITTER*. Universidad del Salvador. [Tesis monográfica].

**Juares, W. ; Pérez. B. ; Reclusa, A. ; Ferreira, S.** (2019). *Twitter y #Corrupción: un balance sobre problemas y técnicas para pensar la política en hashtags..* Universidad Nacional de Mar de Plata. ISBN 978-987-544-895-7.

**Leetanu, K.H.** (2012). *DATA MINING METHODS FOR THE CONTENT ANALYST. An Introduction to the Computational Analysis of Content*. Taylor & Francis Group. Third Avenue, New York.

**López, A.** (2015). *Redes sociales y campañas electorales: Twitter como fuente informativa en las elecciones catalanas del #25N*. Comunicació: Revista de Recerca i d'Anàlisi [Societat Catalana de Comunicació] <http://revistes.iec.cat/index.php/TC>. Vol. 32 (2) (novembre 2015), p. 115-137  
ISSN (ed. impresa): 2014-0304 / ISSN (ed. electrònica): 2014-0444 / DOI: 10.2436/20.3008.01.138.

**López-Meri, A. ; Marcos-García, S. ; Casero-Ripollés, A.** (2017). *¿QUÉ HACEN LOS POLÍTICOS EN TWITTER? FUNCIONES Y ESTRATEGIAS COMUNICATIVAS EN LA CAMPAÑA ELECTORAL DE 2016*. El profesional de la información, 2017, septiembre-octubre, v. 26, n. 5. eISSN: 1699-2407.

**Machado, N. ; Capdevila, A.** (2017). *Los issues de PSOE y Podemos en Twitter durante la campaña electoral de mayo de 2015 en España*. Comunicación y Hombre - Número 13 · Año 2017.

**Márquez, L.** *¿Son los debates electorales la nueva Eurovisión? Análisis del seguimiento de dos eventos masivos a través de Twitter*. Revista DÍGITOS • 3 • 2017 • ISSN: 2444-0132.

**Martínez, J.** (2019). *Estrategia electoral en Twitter de los principales líderes políticos durante las Elecciones Generales en España 2019*. Universitat Oberta de Catalunya (UOC). [Trabajo de fin de máster].

**Puyosa, I.** (2017). *Bots políticos en Twitter en la campaña presidencial #Ecuador2017*. Universidad Central de Venezuela. Contratexto n.º 27, enero-junio 2017, ISSN 1025-9945, pp. 39-60.

**Rodríguez, R. ; Ureña, D.** (2011). *Diez razones para el uso de Twitter como herramienta en la comunicación política y electoral*. Universidad Pontificia de Salamanca. Comunicación y pluralismo, ISSN 1885-8201, N.º. 10, 2011, págs. 89-116.

**Sánchez, A.** (2018). *Sistema de inteligencia de negocio para pacientes y clínicos dentro del marco de las enfermedades llamadas trastornos cognitivos*. Barcelona: Universitat Oberta de Catalunya. [Trabajo de fin de máster].

**Tumasjan, A. ; Sprenger, T.O. ; Sandner, P.G. ; Welpe, I. M.** (2010). *Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment*. Technische Universität München. Leopoldstraße 139, 80804 Munich, Germany.

**Yeste, E. ; Franch, P.** (2018). *TRUMP VS LOS MEDIOS. TRATAMIENTO DE LA PRENSA DESDE LA CUENTA DE TWITTER DEL PRESIDENTE DE EUA. El profesional de la información*, 2018, septiembre-octubre, v. 27, n. 5. eISSN: 1699-2407.

[https://es.wikipedia.org/wiki/Elecciones\\_presidenciales\\_de\\_Estados\\_Unidos\\_de\\_2020](https://es.wikipedia.org/wiki/Elecciones_presidenciales_de_Estados_Unidos_de_2020) (17/09/2020).

[https://es.wikipedia.org/wiki/Web\\_2.0](https://es.wikipedia.org/wiki/Web_2.0) (21/09/2020).

<https://es.wikipedia.org/wiki/RStudio> (22/09/2020).

[https://es.wikipedia.org/wiki/R\\_\(lenguaje\\_de\\_programaci%C3%B3n\)](https://es.wikipedia.org/wiki/R_(lenguaje_de_programaci%C3%B3n)) (22/09/2020).

<https://es.wikipedia.org/wiki/ProjectLibre> (23/09/2020).

<https://www.expansion.com/economia/politica/2019/11/01/5dbac420e5fdea52698b4577.html> (07/10/2020)

<https://www.expansion.com/economia/politica/elecciones-generales/2019/04/10/5cac4e5222601d1c6f8b45cf.html> (07/10/2020)

<http://agendapublica.elpais.com/twitter-como-espacio-de-transformacion-y-debate-politico/> (13/10/2020)

<https://www.elmundo.es/television/2018/06/05/5b1589de22601d84308b45ec.html> (13/10/2020)

<https://arxiv.org/abs/1801.01665> (13/10/2020)

<https://ieeexplore.ieee.org/abstract/document/6113109/authors#authors> (15/10/2020)

<https://academic.oup.com/joc/article-abstract/64/2/239/4085969> (15/10/2020)

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/1944-2866.POI354> (15/10/2020)

<https://212.128.240.21/handle/10115/15773/> (15/10/2020)

<https://medium.com/the-artificial-impostor/analyzing-tweets-with-r-92ff2ef990c6> (28/10/2020)

<https://uvastatlab.github.io/2019/05/03/an-introduction-to-analyzing-twitter-data-with-r/> (29/10/2020)

[https://www.raynamharris.com/blog/sacnas\\_tweet/](https://www.raynamharris.com/blog/sacnas_tweet/) (29/10/2020)

<https://rpubs.com/HAVB/tweet> (03/11/2020)

[https://rstudio-pubs-static.s3.amazonaws.com/595002\\_2d1617098c8c44b494bc2ec97018a82b.html](https://rstudio-pubs-static.s3.amazonaws.com/595002_2d1617098c8c44b494bc2ec97018a82b.html) (05/11/2020)

<https://levelup.gitconnected.com/sentiment-analysis-with-twitter-hashtags-in-r-af02655f2113> (05/11/2020)

<https://sites.google.com/site/miningtwitter/questions/talking-about/wordclouds/comparison-cloud> (07/11/2020)

<https://sites.google.com/site/miningtwitter/questions/sentiment/sentiment> (07/11/2020)

[https://rpubs.com/Joaquin\\_AR/334526](https://rpubs.com/Joaquin_AR/334526) (09/11/2020)

[https://www.granadahoy.com/granada/analisis-Twitter-predecir-resultados-electorales\\_0\\_1516648324.html](https://www.granadahoy.com/granada/analisis-Twitter-predecir-resultados-electorales_0_1516648324.html) (25/12/2020)

<https://es.wikipedia.org/wiki/Bot> (26/12/2020)

## **6. Anexo. Código R utilizado**

En esta sección, se proporciona al lector interesado acceso al enlace <https://drive.google.com/drive/u/0/folders/1qdEIZSvmmr0YL7fvsxnw4xw2HBRIMfjL>, que contiene el documento Wordpad “full\_code” con el código R utilizado para el desarrollo del presente trabajo.