

Análisis de Sentimiento en Twitter de las principales Aerolíneas Estadounidenses

Miguel Pérez Caro, miguel.perez.caro@hotmail.com

Abstracto: Las redes sociales proporcionan una cantidad masiva de información diariamente a la que se puede acceder con relativa facilidad para su uso en investigaciones. El objetivo de este estudio es aplicar técnicas de análisis de sentimiento para conocer la opinión de los pasajeros sobre las principales aerolíneas estadounidenses y buscar posibles deficiencias en sus servicios a través del dato que se puede obtener por medio de la red social Twitter. El estudio muestra una significativa interacción por parte de los pasajeros, a partir de la cuál es posible extraer una idea de la opinión que tienen los pasajeros hacia la aerolínea, así como la existencia de ciertos problemas que aparecen con mucha frecuencia en el servicio proporcionado. Dicha interacción puede aportar información adicional que puede ser usada para mejorar la gestión y el servicio de las aerolíneas.

Palabras Clave: Análisis de sentimiento; aerolíneas; servicio al cliente; twitter

1. Introducción

Las aerolíneas tienen un papel importante en un mundo tan globalizado como el actual, en el que un transporte rápido y a grandes distancias se considera una necesidad básica. Debido a las ingentes cantidades de ingresos que se generan en este sector, hay una gran competencia entre diferentes aerolíneas del mercado [1], y el ser capaz de dar un mejor servicio como aerolínea al pasajero se considera vital para mejorar la reputación de esta, existiendo desde principios de siglo estudios sobre los puntos clave para mejorar la calidad del servicio ofrecido por las compañías aéreas [2], de forma que conocer la opinión de los pasajeros sobre los aspectos a mejorar de la aerolínea puede ser un factor diferencial con la competencia.

Durante los últimos años, el uso de las redes sociales se ha ido incrementando, y una de las más populares es la red social Twitter, cuyo crecimiento se puede ver reflejado en [3], alcanzando más de 300 millones de usuarios activos a nivel mundial. Este gran número de usuarios activos convierte a Twitter en un gran medio de comunicación, ya que mucha gente se informa a través de dicha red social, y en un importante medio de interacción, de forma que una parte de la población publica situaciones de su vida personal a través de dicha red social, lo que conlleva que Twitter sea una herramienta excelente para extraer la opinión de sus usuarios sobre diversas temáticas [4]. Por este motivo puede que muchos usuarios de Twitter publiquen sus experiencias a la hora de coger un vuelo, ya sean positivas o negativas, sus opiniones sobre los servicios de la aerolínea, cualquier problema que haya podido surgir durante el embarque, etc. Todas estas publicaciones son información que puede ser usada por la aerolínea, para mejorar su situación en el mercado, o para conocer la opinión de sus usuarios acerca de sus servicios, sustituyendo así formas antiguas de comunicación con los pasajeros para conocer sus opiniones como podrían ser encuestas en el avión, o llamadas telefónicas. Evidentemente, lo comentado para las aerolíneas es extensible a cualquier otro sector que pueda encontrar información relevante en Twitter acerca de sus operaciones [5].

El objetivo de este proyecto es el de mezclar los dos campos comentados anteriormente, es decir, las aerolíneas y Twitter, de forma que se pretende analizar la

interacción que se produce por parte de los usuarios de determinadas compañías aéreas hacia estas a través de la red social. Para este proyecto se ha decidido acotar el marco del estudio al mercado estadounidense, centrando el análisis únicamente en las principales aerolíneas de Estados Unidos, que se consideran que son las 4 aerolíneas que tuvieron una mayor cuota de mercado nacional entre abril de 2019 y marzo de 2020 según los datos de la web statista [6].

Se espera que dicho análisis pueda mostrar como hay usuarios que expresan sus experiencias en dicha red social, y que las compañías aéreas, a través de algoritmos de procesamiento de lenguaje natural, pueden ser capaces de detectar el sentimiento que se genera en la red social hacia la compañía. También se prevé que se puedan encontrar los servicios que están funcionando, y los que están fallando según la opinión de los pasajeros, para poder actuar acorde a la situación y así poder mejorar la experiencia de los usuarios. Por último, se espera que sea posible detectar ciertos patrones, que pueden estar relacionados con fallos de la aerolínea como cancelaciones o retrasos de vuelos, o con servicios positivos, de forma que se pueda conocer a que servicio corresponden la mayoría de las quejas por parte de los usuarios. Con toda esta información, también se pretende generar una comparación entre las 4 aerolíneas para conocer el sentimiento que se genera hacia ellas en Twitter y qué servicios destacan, tanto negativa como positivamente, en cada una de ellas.

Tras esta introducción, el resto del proyecto se articulará tal y como se especifica a continuación:

- Sección 2 que incluirá un análisis de trabajos previos directamente relacionados.
- Sección 3 en la que se describirá la metodología que se ha llevado a cabo para la obtención del dato y su posterior análisis.
- Sección 4 donde se analizarán los resultados obtenidos y se compararán con los de artículos similares si se considera pertinente.
- Sección 5 en la que redactarán las conclusiones obtenidas del trabajo realizado, así como posibles líneas de trabajo futuro.

2. Estado del arte

El proyecto se centra en el procesamiento de lenguaje natural de tweets relacionados con determinadas aerolíneas, y aunque quizá el objetivo de este trabajo sea ligeramente diferente a otros, este campo es muy amplio y existen muchos estudios que han trabajado de forma similar sobre esta temática.

Se considera relevante mencionar estudios previos sobre el análisis del dato que proviene de Twitter, los tweets, con técnicas diferentes a las de un tratamiento de procesado de lenguaje natural típico, y con enfoques diferentes para predecir el sentimiento de dicho tweet, como sucede en el artículo [7], donde los autores emplean un enfoque novedoso que busca encontrar la orientación semántica de las combinaciones de verbo y adverbio para valorar el sentimiento del tweet posteriormente mediante una ecuación lineal. En el estudio [8], los autores pretenden usar dato de Twitter para construir un modelo capaz de predecir el sentimiento de un tweet, y para clasificar el sentimiento de los tweets hacen uso de las técnicas TextBlob [9] y SentiWordNet [10], más similares a las que se pretenden usar para este proyecto.

También existen muchos artículos que han intentado abordar el uso de técnicas de Machine Learning para predecir si el sentimiento del tweet es negativo o positivo. Todos parten de un set de datos obtenido a partir de Twitter y cuyo sentimiento ya ha sido previamente clasificado, por lo que carecen de un análisis de sentimiento de los textos, y se centran en la parte predictiva. Entre estos estudios se encuentra [11], en el que los autores comparan estudios anteriores que usaron diferentes modelos con un nuevo

enfoque mediante el modelo AdaBoost [12], concluyendo que el análisis de este tipo de dato puede ser diferencial en el mercado. En [13], los autores vuelven a intentar predecir el sentimiento del texto mediante diferentes modelos concluyendo que un ensamble de los modelos probados es el mejor enfoque, y se adquiere una precisión suficiente para que las compañías aéreas puedan usar este dato en sus estudios de mercado. En [14] se vuelve a hacer un estudio predictivo, probando diferentes técnicas de preprocesamiento y con la novedad de que el software WEKA [15] es usado para implementar los métodos realizados. Los autores de [16] llevan a cabo un exhaustivo análisis de previos estudios y comprueban una multitud de modelos y de técnicas de Machine Learning para buscar una mayor capacidad predictiva. En otro estudio [17], los autores hacen uso del mismo set de datos, pero se centran en la importancia del preprocesamiento aplicando diferentes técnicas de Word2Vec [18], aunque finalmente los resultados no son los esperados comparado con estudios similares, como los comentados anteriormente. En definitiva, se demuestra que se puede obtener una idea de la opinión de los pasajeros, y que las aerolíneas pueden hacer uso de modelos predictivos para conocer el sentimiento de las opiniones de sus usuarios.

Los autores de [19] realizan una búsqueda del dato por Twitter haciendo un análisis de sentimiento a través de TextBlob, evaluando el sentimiento de cada tweet y de las aerolíneas en su conjunto, y buscando también combinaciones de 2 o 3 palabras (bigrams o trigrams), con el objetivo de buscar patrones que puedan ser motivo de quejas o, al contrario, de experiencias positivas, concluyendo que el dato proveniente de Twitter puede ser usado para analizar posible fallos que estén sucediendo en la operación de la aerolínea.

Los autores de [20] también llevan a cabo un estudio de análisis de sentimiento a través de Twitter, pero en este caso se centran en un aeropuerto en vez de en aerolíneas. Buscan medir el ASQ (Airport Service Quality) a través de información obtenida en Twitter, sustituyendo así otros estudios previos cuyo dato se obtenía a partir de encuestas. Para el procesamiento, hicieron una búsqueda de palabras clave relacionadas con parámetros que componen el ASQ, y para el análisis de sentimiento se usaron las herramientas Theysay [21], desarrollada por lingüistas computacionales de la Universidad de Oxford, y la herramienta Twinword [22]. Los autores concluyen que, aunque un análisis más detallado incluyendo otras herramientas o un preprocesamiento más exhaustivo se considera necesario, con el trabajo realizado se pueden identificar los servicios que se califican como positivos en el aeropuerto en cuestión, y los que tienen opiniones más negativas, lo que convierte al dato que proviene de esta red social en una información muy valiosa para mejorar la calidad de los servicios de un aeropuerto.

Por último, el autor de [23] hace un estudio similar al que se propone en este proyecto, pero sobre aerolíneas del mercado europeo. Para ello, recoge tweets de un periodo de un mes sobre dichas aerolíneas, y para el análisis de sentimiento, en este caso, se hace uso de los léxicos Bing [24] y NRC Emotion Lexicon [25] de forma que se clasifican las palabras por diferentes categorías asociadas a los sentimientos y de esta forma, a través del sentimiento de cada palabra, se obtiene el sentimiento de los documentos, que en este caso son los tweets.

3. Materiales y métodos

En esta sección se pretende describir como se ha obtenido el dato, los métodos seguidos para el análisis de sentimiento de los documentos y las técnicas utilizadas para la búsqueda de patrones que puedan existir para cada compañía aérea.

A la hora de estructurar el trabajo realizado, se ha seguido la metodología CRISP DM [26], exceptuando la última sección que sería el despliegue al no considerarse necesario para este proyecto.

3.1. Datos

Como se ha comentado previamente, el dato se ha obtenido a través de la red social Twitter. Para ello, se ha hecho uso de la API de Twitter [27], que permite obtener tweets con una existencia no superior a 7 días y que impone unas restricciones de un máximo de 180 tweets cada 15 minutos. Para poder operar con dichas restricciones, se ha hecho uso de la librería de Python tweepy [28], que mediante una serie de parámetros gestiona dichas restricciones, facilitando la obtención de los tweets. Se han usado como palabras claves para conseguir los tweets deseados, los usuarios de las cuentas de Twitter de las 4 aerolíneas y los nombres de las compañías, generando un set de datos por cada palabra clave, y por cada vez que se recogían los datos, lo cuál se ha llevado a cabo desde el 3 de Agosto hasta el 7 de Octubre. Como el objetivo es el análisis del texto, se han recogido únicamente los campos correspondientes al id del tweet, la fecha, el texto, y los retweets y favoritos por si pudiesen ser de utilidad. Una vez se obtiene el set de datos, comienza la parte del preprocesado de los mismos. En primer lugar, se ha de tener en cuenta la existencia de duplicados del mismo tweet, que se debe a la metodología de extracción realizada, existiendo 3 posibles escenarios:

- Para la recogida de tweets de una aerolínea, se usan dos formatos, por el usuario, y por el nombre de la aerolínea. Es posible que un mismo tweet mencione la aerolínea por su usuario y por su nombre, lo que haría que existiese un duplicado.
- Existe la posibilidad de que un mismo tweet mencione 2, 3 o las 4 aerolíneas, por lo que sería recogido en cada set de datos correspondiente a cada aerolínea.
- Finalmente, también puede suceder, debido a los tramos de tiempo de recogida, que un mismo tweet aparezca en dos sets de datos que se hayan recogido en diferentes fechas.

Se eliminan los duplicados por el id del tweet, teniendo en cuenta que un mismo tweet puede estar referenciando a más de una aerolínea. Una vez eliminados los tweets duplicados por id, se hace lo propio por el texto, ya que durante el análisis se encontró una gran cantidad de textos duplicados debido, probablemente, a menciones y retweets, por lo que se sigue una metodología de eliminación de duplicados igual a la anterior. Una vez se ha conseguido un set de datos sin duplicados, se comienza con el preprocesamiento del texto el cual incluirá:

- Tratamiento de emojis y emoticonos, habituales en el lenguaje utilizado en las redes sociales y que sí que muestran un sentimiento. Para ello se hace uso de la librería de Python emot [29].
- Eliminar los nombres de usuario ya que no aportan información relevante al análisis de sentimiento del texto.
- Eliminar el símbolo de almohadilla de los hashtags, ya que el resto del contenido sí que puede ser útil para el análisis posterior.
- Eliminar los números de vuelo, y los términos referidos a la posición en la navegación de los aviones, ya que tampoco aportan sentimiento.

- Eliminar signos de puntuación, stopwords, enlaces web, y cualquier vocablo que se encuentre de forma frecuente y no sea útil para el análisis pudiendo incluso empeorarlo.
- Tras llevar a cabo esta limpieza, puede haber nuevamente textos duplicados debido a que el tweet inicial se diferenciaba en un enlace web al que hacían referencia o en un nombre de usuario, por lo que se eliminan dichos tweets duplicados por el texto con la misma metodología seguida anteriormente.

Una vez se han realizado todos los pasos anteriores, se considera que el texto está limpio y listo para comenzar el análisis de sentimiento. La cantidad de tweets total es de 244522, con la siguiente distribución por compañía aérea teniendo en cuenta que algún tweet puede hacer referencia a más de una aerolínea:

Tabla 1. Tweets por compañía aérea

Aerolínea	Número de Tweets
American Airlines	80687
Delta Airlines	77584
Southwest Airlines	54379
United Airlines	49973

3.2. Análisis de Sentimiento

Para este apartado, se van a plantear dos técnicas diferentes para comparar posteriormente sus resultados:

- La primera de ellas es la librería de python TextBlob, que ya ha sido mencionada anteriormente en alguno de los estudios comentados, y permite realizar diversas tareas de procesamiento de lenguaje natural, incluyendo el análisis de sentimientos. De esta forma, permite obtener el valor de polaridad de una palabra, una frase o un documento. Dicho valor oscila en el rango de -1 a 1, de forma que todo valor menor que 0 es considerado como un texto negativo, y todo valor mayor que 0 es considerado como un texto positivo, mientras que el valor 0 se considera neutro, es decir, carece de sentimiento. Para obtener el valor de polaridad, TextBlob utiliza un léxico predefinido a partir del cuál aporta un valor de polaridad, subjetividad e intensidad a la palabra, que se entiende como si dicha palabra modificase la siguiente. Con estos valores, se obtiene la polaridad de cada término y por un promediado, el del texto que se pretende analizar.
- La segunda metodología que se va a emplear para el análisis de sentimiento es el uso de la librería de Python VADER [30], enfocada principalmente al análisis de textos procedentes de las redes sociales, como es el caso que se aborda en este proyecto. Esta técnica, también tiene un léxico propio en el que contiene palabras con un valor según el sentimiento que generan, pero también es capaz de entender todo tipo de negaciones, contracciones verbales, o incluso el lenguaje coloquial que se suele usar en las redes sociales. VADER obtiene el sentimiento del texto a través de una métrica que calcula la suma de las valoraciones de todo el léxico normalizado al rango -1 a 1, y se considera sentimiento negativo todo valor inferior a -0.05, y positivo todo valor superior a 0.05, de forma que los resultados entre estos dos valores serán clasificados como neutros.

Para llevar a cabo el análisis del sentimiento, se ha hecho uso de dos sets de datos diferentes. El primero se ha obtenido tras la limpieza comentada en el apartado anterior, mientras que el segundo es idéntico al primero, pero se excluye la limpieza de los signos

de puntuación, de los emoticonos y de los emojis, ya que ambas técnicas son capaces de preprocesar dichos elementos, y se pretende comprobar si el resultado final varía. Los resultados obtenidos, así como las comparaciones entre las 4 cuatro aerolíneas se encuentran en la sección 4.

Finalmente, se ha de mencionar que el motivo por el que estas técnicas se han escogido para realizar el análisis de sentimiento es, en el caso de TextBlob, que es la más usada en los artículos estudiados para obtener mayor información sobre el análisis de sentimiento, por lo que se consideró que era una buena opción probar su uso, y en el caso de VADER, el principal motivo es que es una herramienta enfocada al análisis de textos procedentes de redes sociales, por lo que también es una opción muy válida para este proyecto y, además, el uso de ambas técnicas podría llevar a una comparación que podría indicar cuál es la más adecuada, aportando un valor añadido a los resultados obtenidos en el estudio.

3.3. Búsqueda de Patrones

Este último apartado se puede dividir en dos partes. La primera de ellas consiste en la conversión de los documentos en formato texto a vectores que puedan ser posteriormente procesados por modelos de Machine Learning u otros algoritmos. Se van a usar diferentes técnicas de conversión, en algunos casos debido a que son necesarias para la aplicación de algún algoritmo específico, y en otros porque se pretende hacer uso de todas ellas para buscar alguna diferencia en los resultados finales e intentar obtener alguna conclusión sobre qué forma de conversión puede ser más adecuada para este específico caso de uso. Las técnicas que se usarán para esta conversión, y cuyos resultados se compararán posteriormente, son las siguientes:

- La primera técnica que se va a utilizar es Bag of Words (BoW), que es un modelo conocido como bolsa de palabras que interpreta un documento como un conjunto de palabras, ignorando el orden de estas, por lo que cada texto se convertirá a un vector que consistirá en el número de veces que aparece cada palabra. Se hace uso de este modelo porque es necesario para la agrupación por topics que se realizará posteriormente.
- En segundo lugar, se va a aplicar una transformación a través de modelos que ya han sido previamente entrenados, que son una colección de embeddings que han sido ya creados y solo hay que descargar para poder aplicarlos. Esta técnica es de gran utilidad porque permite tener acceso a un modelo entrenado con una cantidad masiva de datos a la que de otra forma no se podría tener acceso. La desventaja es que el corpus contenido puede no captar ciertas peculiaridades del conjunto de datos con el que se está trabajando, pudiendo ser muy diferente el corpus con el que se han entrenado estos modelos al obtenido por Twitter. Para obtener uno de los múltiples modelos que se pueden encontrar, se va a hacer uso de la librería Spacy [31], ya que está diseñada para un rendimiento rápido, y es perfecta para análisis sencillos y de poca duración, y que además trae los modelos consigo, a diferencia de otras opciones donde se requieren más operaciones para poder obtenerlos.
- La siguiente conversión a vector se va a realizar a través de un modelo Doc2vec, que es la extensión de Word2vec, ya referenciada anteriormente, a embeddings de documentos. Word2vec es un algoritmo que incrusta palabras en un espacio vectorial de menor dimensión haciendo uso de una red neuronal poco profunda. El resultado es un conjunto de vectores de palabras donde los vectores cercanos en el espacio vectorial tienen significados similares basados en el contexto, y los vectores de palabras distantes entre sí tienen significados diferentes. El problema es que el modelo Word2vec solo calcula vectores por palabras, y de ahí que se haya generado el modelo Doc2vec que generalmente supera el resultado de hacer un simple

promedio de los vectores Word2vec. Como principal característica de esta técnica, diferenciándola de la anterior, aquí solo se hace uso del corpus con el que se está trabajando.

- La última conversión a vector se va a realizar a través de BERT [32], que es un modelo de procesamiento de lenguaje natural que aplica técnicas que en su diseño fueron novedosas y mejoraban los algoritmos con los que se estaba trabajando. BERT aplica un entrenamiento bidireccional de transformadores, que contrasta con las técnicas previas de leer el texto de derecha a izquierda, de izquierda a derecha o en ambas. Esta técnica ayuda a captar mucho mejor el contexto de las frases. Para ello, el modelo ha sido entrenado con un corpus sin clasificar de grandes dimensiones que incluye, entre otros, toda la Wikipedia y multitud de libros. BERT puede ser aplicado a muchas tareas de procesamiento de lenguaje natural ya que permite adaptar este modelo pre-entrenado, haciendo uso de todo lo aprendido.

Una vez comentados los métodos de conversión que se van a utilizar, se procede con la segunda sección de este apartado, que son los procedimientos usados para la búsqueda de clústeres y/o patrones:

- El primer método es la agrupación por topics, que consiste en el uso de algoritmos de topic modeling que buscan grupos de topics dentro de un conjunto de documentos. Hay varios modelos dentro de esta técnica que podrían ser usados, y se ha escogido el modelo Latent Dirichlet Allocation (LDA), en base a los resultados observados en el estudio [33], que analiza diferentes algoritmos de topic modeling para textos cortos obtenidos de redes sociales, al igual que se hace en este proyecto, concluyendo que LDA obtiene grandes resultados, y es de los más usados. LDA es un modelo generativo que presupone que los documentos están compuestos por una distribución de topics, y cada topic compuesto por una distribución de palabras.

Entre las cuestiones técnicas, cabe destacar que este modelo se basa en los modelos de bolsas de palabras, generalmente sin aplicar normalizaciones sobre los mismos, por lo que el orden de las palabras no se tiene en cuenta. Un aspecto importante de los modelos LDA es la dificultad de encontrar el número óptimo de topics, ya que este es uno de los parámetros del modelo. Para ello, se busca una forma empírica para averiguar el número óptimo, que es el coherence topic, una métrica que mide para cada topic el grado de similitud semántica entre las palabras de puntuación alta en dicho topic.

- El segundo método que se va a aplicar es la reducción de dimensionalidad no lineal, a través de la cual se puede obtener una aproximación a un espacio de baja dimensionalidad donde los datos tengan una estructura global y local similar al espacio original, pudiendo por tanto visualizar el corpus correspondiente respetando su estructura interna y proporcionando información muy valiosa en la interpretación. Hay diferentes métodos para realizar dicha reducción, pero se va a hacer uso del algoritmo UMAP [34], escogido sobre otros métodos como t-Stochastic Neighbor Embedding debido a la inclusión de varios parámetros que permiten modelar la reducción de dimensionalidad, como el parámetro `min_dist`, que controla como de estricto es el algoritmo a la hora de agrupar puntos, de forma que valores bajos de este parámetro resultarán en embeddings más agrupados, lo cual puede resultar útil para la obtención de clústeres.

UMAP emplea la técnica conocida como aprendizaje de manifolds, basada en la idea que la alta dimensionalidad de los datos se debe a una distribución artificial de estos datos en alta dimensionalidad, y existe un espacio de menor dimensionalidad que puede contener estos mismos datos de una forma muy similar. Encontrar esta función que reduce la dimensionalidad, en este caso, consiste en encontrar la distribución de los datos en alta dimensionalidad, de forma que podamos encontrar

proyecciones de esa distribución a espacios de baja dimensionalidad sin perder una gran cantidad de información. Esta técnica se aplicará a los corpus obtenidos a través del modelo pre-entrenado de Spacy, mediante el modelo Doc2Vec y por el algoritmo BERT.

- El ultimo método que se va a utilizar en este apartado es la búsqueda de clústeres a través del algoritmo de aprendizaje no supervisado más utilizado en Machine Learning, K-Means, siendo el procesamiento de lenguaje natural una de sus múltiples aplicaciones. Esta técnica se aplicará a los corpus obtenidos a través del modelo pre-entrenado de Spacy, mediante el modelo Doc2Vec y por el algoritmo BERT. Para obtener el número de clústeres óptimos en cada corpus se van a comprobar dos métodos. El primero es conocido como Dispersión, y utiliza la distancia al cuadrado de los puntos al centroide más próximo frente al número de clústeres de forma que en el momento en el que cambia la velocidad de reducción de la dispersión se entiende que se ha llegado a la cantidad óptima de clústeres. El segundo método es Silhouette, que consiste en calcular un coeficiente que se consigue a partir de la distancia intraclúster media de los elementos de un clúster y la distancia media a los elementos del clúster más cercanos.

4. Resultados y Discusión

En este apartado se van a analizar y comparar los resultados obtenidos en el estudio, que se ha centrado en dos secciones muy diferenciadas, que son, el análisis de sentimiento y la búsqueda de patrones, por lo que se van a comentar por separado.

4.1. Análisis de Sentimiento

En primer lugar, es necesario puntualizar que, como se comentó en la sección 3, se ha hecho uso de dos datasets con diferentes preprocesamientos de limpieza, aunque muy similares. Como los resultados se asemejan en casi todos los análisis, solo se va a exponer lo obtenido con el set de datos al que se le realizó la limpieza completa, puntualizando las diferencias que haya existido entre ambos análisis en aquellos aspectos que así lo requieran.

Para comenzar a analizar los resultados, se van a mostrar los valores obtenidos para las dos técnicas que se han usado. La información que se muestra es la siguiente:

- Total de tweets positivos y porcentaje sobre el total de tweets de cada aerolínea.
- Total de tweets neutros y porcentaje sobre el total de tweets de cada aerolínea.
- Total de tweets negativos y porcentaje sobre el total de tweets de cada aerolínea.
- Sentimiento medio de la aerolínea, es decir, la media de los valores numéricos que evalúan el sentimiento de un tweet en un rango de -1 a 1.

Tabla 2. VADER – Análisis de sentimiento

Aerolínea	Tweets Positivos	Tweets Neutros	Tweets Negativos	Sentimiento Medio
American Airlines	28548 (35.38%)	27161 (33.66%)	24978 (30.96%)	0.029
Delta Airlines	31858 (41.06%)	26027 (33.55%)	19699 (25.39%)	0.093
Southwest Airlines	22389 (41.17%)	17959 (33.03%)	14031 (25.80%)	0.097
United Airlines	21997 (44.02%)	16432 (32.88%)	11544 (23.10%)	0.122

Tabla 3. TextBlob – Análisis de sentimiento

Aerolínea	Tweets Positivos	Tweets Neutros	Tweets Negativos	Sentimiento Medio
American Airlines	27149 (33.65%)	36995 (45.85%)	16543 (20.50%)	0.043
Delta Airlines	27595 (35.57%)	35397 (45.62%)	14592 (18.81%)	0.065
Southwest Airlines	20459 (37.62%)	24350 (44.78%)	9570 (17.60%)	0.077
United Airlines	19072 (38.16%)	22630 (45.28%)	8271 (16.55%)	0.083

De los datos obtenidos se puede observar como American y Delta son las compañías que tienen una mayor interacción por Twitter, seguidas por Southwest, y, por último, United. Dado que se ha hecho el análisis con dos herramientas diferentes, se van a comentar los resultados por separado.

En cuanto a los datos obtenidos por VADER, para todas las aerolíneas los valores positivos superan a los negativos, lo cual es una buena señal, aunque la cantidad de tweets negativos es relativamente alta. Los porcentajes permiten conocer como se distribuyen los tweets para cada aerolínea. Se observa como el porcentaje de tweets neutros es muy similar en todos los casos, existiendo diferencias en los porcentajes positivos y negativos, donde se aprecia como American es la aerolínea con un mayor porcentaje negativo, por lo que se entiende que es la que recibe más críticas o quejas a través de Twitter, mientras que United, que es la que tiene una interacción menor, es la más positiva. En cuanto a Southwest y Delta tienen una interacción muy similar. Por último, el sentimiento medio reafirma las conclusiones ya mencionadas. American obtiene el sentimiento medio más bajo de todas las aerolíneas, mientras que United es la que tiene una media más alta, por lo que es la aerolínea que recibe una mayor interacción positiva. Por otra parte, sobre Southwest y Delta, que obtuvieron valores totales muy similares, vuelven a obtener un sentimiento medio similar, pero ligeramente superior en el caso de Southwest.

Para TextBlob, los resultados son diferentes a los obtenidos con VADER. En este caso, la clasificación que predomina para todas las aerolíneas es la de tweets neutros, obteniendo unos valores muy elevados, lo cual presenta una clara diferencia entre ambas técnicas. También destaca que la cantidad de tweets negativos es baja, por lo que no encuentra fácilmente tweets donde los clientes puedan expresar una queja o disgusto sobre el servicio ofrecido. En lo relativo a la distribución del sentimiento por aerolínea sí que se pueden obtener conclusiones similares. El porcentaje de tweets clasificados como neutros para cada aerolínea es prácticamente el mismo. Nuevamente American es la que recibe una peor crítica, aunque con una menor diferencia que en el caso anterior. United vuelve a ser la que recibe una interacción más positiva, mientras que Delta y Southwest vuelven a obtener valores similares, pero ahora sí que se puede concluir que Southwest recibe una interacción más positiva que Delta. También cabe destacar que los resultados en este caso son más parejos que en el anterior. Finalmente, el sentimiento medio, al igual que pasaba en el caso anterior, permite obtener ideas similares. American vuelve a ser la compañía aérea que muestra un peor resultado, mientras que United destaca sobre el resto, y Southwest obtiene un sentimiento medio claramente superior al de Delta.

Para continuar con el análisis de sentimiento, se han buscado los bigrams, más repetidos en los tweets que mencionan cada aerolínea para intentar obtener alguna conclusión adicional. En todos los casos, “customer service” es de los bigrams que más se repiten, así como “flight attendant” o “flight attendants”, lo cual hace indicar que puede existir una clara necesidad de analizar dichos tweets ya que pueden incluir información relevante sobre la opinión de los usuarios de Twitter acerca de los servicios ofrecidos por la compañía aérea. También destacan en gran medida para todas las aerolíneas la aparición de emojis o emoticonos en el set de datos al que se le ha realizado un preprocesamiento completo, ya que los emojis o emoticonos son convertidos a un conjunto de palabras que expresan la imagen del mismo, mientras que en el set de datos

que incluye los emojis y los emoticonos, esto no sucede, ya que no se detectan los mismos como bigrams, pero del primer set de datos, se puede concluir el frecuente uso de dichos elementos.

Siguiendo con el análisis de los conjuntos de palabras, se observan algunos términos negativos que se consideran muy importantes en el sector que se pretende analizar en este proyecto, como son, “flight delayed”, “cancelled flight”, “flight cancelled” o “never fly”, y que muestran un claro problema en los servicios ofrecido por la compañía, siendo una de las razones por las que existe un porcentaje relevante de tweets negativos. Estos términos aparecen con una frecuencia elevada en el caso de American, mientras que, en el resto de las compañías, no son tan frecuentes. Otros términos que son observados con una alta frecuencia, y que aportan siempre un sentimiento, ya sea negativo o positivo, son los emojis o emoticonos, los cuáles tienen una gran importancia en los resultados obtenidos porque los bigrams más frecuentes en el caso del set de datos que contiene los emojis sin preprocesar son en su mayoría de sentimiento neutro.

Aparte de lo comentado, y con un sentimiento neutro, se encuentran otros términos que eran relevantes en el sector en el momento de la obtención de los tweets como pueden ser los relacionados con las vacunas como “covid vaccine”, “get vaccinated”, “delta variant” o “wear mask”.

Una vez expuestos todos los resultados obtenidos de los análisis realizados para conocer el sentimiento de los tweets, se aprecia que existe una diferencia clara entre los resultados obtenidos por VADER y por TextBlob. En un 42.82% de los tweets, los resultados han sido diferentes. La gran diferencia se produce en el caso en el que VADER detecta un sentimiento en el tweet mientras que TextBlob no lo hace, lo cuál representa un 53.66% del total de tweets en cuyo resultado no es el mismo para ambas técnicas. Esto se debe a que el análisis de TextBlob es menos sensible que el de VADER, mostrando este último una mayor capacidad para analizar los emojis y emoticonos. También es relevante destacar que VADER detecta como algo negativo la cancelación o el retraso de un vuelo, mientras que TextBlob no, lo cuál hace que, para el sector que ocupa en este proyecto, VADER sea un mejor analizador de sentimiento que TextBlob para futuras aplicaciones dentro de los textos relacionados con el análisis de opinión de pasajeros sobre los servicios ofrecidos por una aerolínea.

En cuanto a las aerolíneas, se ha observado como American es la aerolínea que recibe una interacción más negativa, siendo también la que muestra una mayor frecuencia de tweets con términos relacionados a la cancelación o retraso de un vuelo, problema que también aparece para las otras 3 aerolíneas, pero en menor medida. También se observa como United obtiene una interacción más positiva que el resto, por lo que parece que recibe un menor número de críticas por parte de los pasajeros, mientras que Southwest obtiene una interacción ligeramente más favorable que Delta.

Por último, queda comparar los resultados obtenidos con alguno de los estudios similares comentados en la sección 2:

- En [19] se recoge dato de Twitter y se analiza sobre muchas aerolíneas entre las que se encuentran Delta, Southwest y United, para las que se calcula el sentimiento medio semanal durante un mes. El rango de valores del sentimiento medio es similar al obtenido en este estudio, pero en este caso Southwest obtiene un sentimiento medio en torno a 0.135, Delta cercano a 0.065, mientras que United muestra muchas variaciones con una media que se aproxima a 0.05, por lo que contrasta con los resultados obtenidos donde United era la aerolínea que mostraba una interacción más positiva.
- En [23], el autor analiza el sentimiento sobre aerolíneas que operan mayoritariamente en el mercado europeo. Para analizar el sentimiento utiliza otras técnicas de léxicos,

pero es destacable que en su mayoría clasifica los tweets como neutros, y los positivos superan a los negativos, obteniendo resultados similares a los obtenidos en este estudio con TextBlob.

- Por último, en la sección 2 se han analizado muchos estudios que partían de un set de datos obtenido a través de Twitter ya clasificado manualmente que se encuentra en Kaggle. Dicho set de datos contiene tweets de aerolíneas estadounidenses, entre las que se encuentran las 4 en las que se ha centrado este estudio. En la Tabla 4 se pueden encontrar los resultados que comparan el análisis de sentimiento manual y el realizado con las mismas técnicas aplicadas en este estudio. Se muestra como manualmente se obtiene una mayor interacción negativa, y los resultados son drásticamente diferentes, en cuanto a los porcentajes de tweets negativos y positivos. También es destacable que las diferencias entre las aerolíneas son considerablemente similares, es decir, con ambas clasificaciones por sentimiento, se aprecia qué aerolíneas obtienen una respuesta más negativa, y cuáles más positiva, por lo que el sentimiento medio obtenido en el estudio, aunque manualmente pudiera haber sido muy diferente, viendo este ejemplo, es bastante posible que American siguiera siendo la aerolínea con una interacción más negativa, mientras que la más favorable sería United.

Tabla 4. Análisis de sentimiento Kaggle Dataset

Aerolínea	Positivos	Neutros	Negativos	Positivos VADER	Neutros VADER	Negativos VADER
American Airlines	12.18%	16.78%	71.04%	41.79%	24.10%	34.11%
Delta Airlines	24.48%	32.54%	42.98%	45.68%	31.23%	23.09%
Southwest Airlines	23.55%	27.44%	49.01%	50.12%	25.25%	24.63%
United Airlines	12.87%	18.24%	68.89%	43.59%	22.97%	33.44%

4.2. Búsqueda de Patrones

Se ha intentado buscar patrones o clústeres dentro del conjunto de datos a partir de 3 técnicas diferentes para poder comparar los resultados entre ellas y también poder obtener una mayor información. Se analizan a continuación los resultados para cada una de las técnicas utilizadas:

- El primer método que se ha utilizado para la búsqueda de patrones es el modelo LDA con el objetivo de encontrar diferentes topics dentro del set de datos que pudieran aportar una idea de lo que se encuentra en dicho conjunto de datos. Se ha generado un modelo para 12 topics, número óptimo obtenido a través de la métrica coherence score mencionada anteriormente, y para su análisis se ha visualizado a través de la herramienta pyLDAvis [35]. Dentro de los topics encontrados en la visualización, destacan 2 que parecen aglutinar muchos de los relacionados con el sector y el servicio. El primero de ellos incluye como término principal “flight” y otros como “delayed” o “cancelled”, importantes para una aerolínea, mientras que el segundo al que se hace referencia incluye palabras como “customer”, “service” o “worst”, y en ambos casos el sentimiento medio negativo de los tweets pertenecientes a este topic es muy bajo, lo cual muestra que existe una parte del conjunto de datos recogidos que referencian quejas por parte de los usuarios.

Por otro lado, también se encuentra un topic que incluye palabras de agradecimiento o términos que demuestran positividad, siendo el topic que tiene un sentimiento medio superior al resto con una clara diferencia, por lo que parece indicar que existe un cierto porcentaje de tweets con mensajes positivos hacia la aerolínea. Otros dos topics de cuyos términos se pueden extraer conclusiones son, un primero relacionado con un tema derivado del clima que incluye términos como “weather” o “issue”, lo

cuál tiene sentido en el sector de la aviación comercial, y un segundo que también está altamente relacionado con el sector en el momento actual y que trata asuntos acerca del virus Covid19 incluyendo términos como “vaccine”, “vaccinated” o “employees”, por lo que puede incluir temas como la política de vacunación de la aerolínea, las vacunas a los empleados etc. Por último, se puede encontrar otro topic con términos como “luggage”, “bag”, “bags” o “lost”, lo cuál parece indicar que hace referencia a las quejas sobre maletas o bolsas de equipaje extraviadas, y que nuevamente evidencia un problema en los servicios de la aerolínea.

Del resto de topics es difícil sacar conclusiones. El modelo LDA permite conocer más a fondo el conjunto de datos observando que existen ciertos patrones, como tweets que hacen referencia a las cancelaciones o retrasos de vuelo y que muestran un sentimiento negativo, tweets de agradecimiento, el tema de las vacunas y el virus Covid19 que era esperable que apareciese, temas climatológicos o problemas con el equipaje.

- El segundo método utilizado para este apartado ha sido la reducción de dimensionalidad a través del algoritmo UMAP. Dicho algoritmo se ha aplicado a los corpus obtenidos por medio de aplicar el modelo Doc2vec, el modelo pre-entrenado de Spacy y el algoritmo BERT al conjunto de datos, y también se ha calculado para una combinación de los parámetros min_dist y n_neighbors con diferentes valores, de forma que se obtiene una mayor variedad de visualizaciones para cada gráfico con el objetivo de poder facilitar la búsqueda de clústeres dentro de las imágenes.

Por lo general, la gran mayoría de gráficos obtenidos muestran una masa compacta de puntos donde apenas se aprecian diferentes grupos que puedan sugerir la existencia de clústeres, aunque hay pequeños casos en los que sí se observan diferencias. En el caso de Doc2vec, la visualización se compone de un gran conjunto de puntos en el que apenas hay pequeñas porciones de los mismos que se alejan ligeramente del conjunto principal, pero se pueden apreciar cuatro posibles clústeres pequeños que pueden aportar algo de información. El primero parece claramente positivo, incluyendo muchos tweets de agradecimiento, mientras que el segundo es negativo, centrándose en las cancelaciones. El tercero está relacionado con el Covid19, las vacunas o las mascarillas y, por último, el cuarto parece tratar sobre el equipaje al referenciar términos relacionados al mismo. En el caso de Spacy, nuevamente se aprecia un conjunto de puntos muy denso, en el que se puede visualizar una zona con una mayor cantidad de tweets positivos, y otra con una mayor densidad de tweets negativos, pero muy difícil de observar, aunque en esta visualización, a diferencia de la anterior, sí que hay dos conjuntos de puntos que se encuentran claramente alejados del principal siendo ambos de sentimiento neutro. El primero está relacionado con tweets que mencionan coordenadas de los aviones, aunque de una forma diferente a la habitual, ya que esta fue incluida en la limpieza, y el segundo parece incluir tweets que anuncian vuelos baratos debido al contenido de los mismos, pudiendo pertenecer a algún tipo de cuenta que se dedica a promocionar ofertas. Por último, para el corpus generado a través del algoritmo BERT, la visualización obtenida no presenta ninguna novedad a lo ya comentado para los datos obtenidos a través de las otras técnicas apreciándose nuevamente clústeres relacionados con las coordenadas de vuelos, con vuelos baratos, y zonas con mayor densidad tanto de tweets negativos como positivos.

Tras el análisis realizado, se comprueba que existe una clara dificultad para encontrar clústeres bien definidos, aunque sí que hay ciertos patrones que se encuentran, que ya habían sido mencionados tras el análisis del resultado de aplicar el modelo LDA. Entre estos patrones hay que destacar las zonas con una mayor densidad de tweets positivos ya que, aunque muchos de ellos así lo son, hay un gran porcentaje que sufren

una clara perturbación de los resultados debido a la ironía usada por los usuarios a través de los emojis o emoticonos, puesto que se han encontrado varios casos en los que el tweet referencia un problema que el pasajero ha experimentado con la aerolínea finalizando el mismo con un emoji de agradecimiento, claramente de forma irónica, pero la herramienta utilizada para el análisis de sentimiento, VADER, no es capaz de detectar dicha ironía y el peso positivo del emoji es superior al negativo ocasionado por la experiencia del pasajero resultando en un tweet que parece de agradecimiento con sentimiento positivo pero es todo lo contrario. También se aprecia un claro grupo de tweets negativos con términos relacionados con la cancelación, cambio o retraso de un vuelo, siendo estos términos un patrón se encuentra de manera habitual por lo que parece que es un tema recurrente en la interacción de los usuarios con las aerolíneas. Por último, existen ciertos grupos menos relevantes que hacen referencia en menor medida a temas como son los relacionados con el virus Covid19 y las vacunas, tweets que informan de las posiciones geográficas de los aviones que se encuentran en ruta o tweets que parecen promocionar vuelos a bajo precio.

- La última técnica usada para la búsqueda de clústeres es el algoritmo K-Means. En este caso, para encontrar el número óptimo de clústeres se han usado las técnicas de Dispersión y de Silhouette. Para el corpus obtenido con Doc2vec ninguno de los dos métodos ayuda a elegir el número óptimo de clústeres. Para el corpus obtenido con el modelo pre-entrenado de Spacy, parece que el óptimo podría ser 3 clústeres, mientras que, para el obtenido a través de BERT, la técnica Silhouette parece indicar un óptimo de 4 clústeres mientras que para el método Dispersión deberían ser 3, aunque ninguno parece ser muy claro. Ninguna de las técnicas permite obtener conclusiones claras acerca del número óptimo de clústeres. Se van a probar diferentes combinaciones para analizar posteriormente, de forma que el corpus generado a través de Doc2vec se buscarán 2 y 3 clústeres, igual que para Spacy mientras que para BERT serán 3 y 4.

Dados estos parámetros parece que solo hay dos modelos de los que se pueden extraer conclusiones y que permiten hacer diferencias entre alguno de sus clústeres. Para analizar los resultados se han obtenido los conjuntos de 2 palabras más comunes por cada clúster y su sentimiento medio. El primero de los modelos es el de 3 clústeres aplicado al vector obtenido a través del modelo pre-entrenado de Spacy, en el cual sí que se aprecian diferencias. El primer clúster parece que tiene como tema principal tweets relacionados con las vacunas y el Covid19, el segundo clúster en cambio incluye términos más con los vuelos, cancelaciones, retraso, etc. Por último, en el tercer clúster los términos más comunes hacen referencia a los emojis/emoticonos, por lo que parece incluir los tweets con estos elementos. En cuanto al sentimiento, tiene sentido que el segundo clúster tenga el sentimiento más bajo, ya que incluye términos negativos en la experiencia de los pasajeros, mientras que el último muestra un sentimiento más positivo, por lo que se intuye que el uso de emojis/emoticonos es en su mayoría positivo. El otro modelo del que se puede sacar alguna conclusión es el aplicado al vector obtenido a través de BERT para 4 clústeres, donde el segundo clúster incluye términos relacionados con el vuelo, mientras que en el resto los términos no aportan ningún valor añadido y no se puede extraer ninguna conclusión.

La búsqueda de clústeres a través de un modelo K-Means solo ha sido ligeramente fructuosa en el caso del vector obtenido a través del modelo pre-entrenado de Spacy para el caso de 3 clústeres, pero únicamente se aprecian de forma limitada temas ya comentados, por lo que es de poca utilidad.

En general, se evidencia una clara dificultad para encontrar clústeres bien definidos en el corpus, aunque sí se ha demostrado que se pueden encontrar cierto patrones o temas que son de alta relevancia para el sector.

Comparando los resultados con estudios similares, en ninguno de ellos se busca la obtención de clústeres, que en el caso de este estudio no ha finalizado con un resultado favorable, pero sí se obtienen ciertos patrones o temas frecuentes.

En [19], se muestran bigrams similares a los obtenidos como “customer service”, “flight delayed” o “baggage service”, y también otros diferentes, pero igualmente relacionados con el sector como “confirmation number” o “reservation recipient”. Por último, a partir del set de datos clasificado manualmente al que se hacía referencia en el apartado anterior, a parte de clasificar el sentimiento del tweet, también se clasificó de forma manual el motivo del sentimiento negativo en el caso de que existiera, destacando motivos como problemas con el servicio al cliente, retraso de un vuelo, vuelo cancelado, pérdida o daño al equipaje, experiencia negativa en el vuelo, quejas sobre las azafatas o problemas a la hora de reservar un vuelo. Muchos de estos problemas se han detectado en el corpus como patrones, por lo que se evidencia que son problemas frecuentes, y que los pasajeros hacen uso de Twitter para mostrar su disconformidad o enfado acerca de las malas experiencias sufridas.

5. Conclusiones

El estudio refleja la viabilidad del uso de la red social Twitter como una fuente de datos para conocer la opinión de los pasajeros sobre sus servicios y poder medir el grado de satisfacción con la misma. Todo ello se refleja en la información que se ha podido encontrar en los tweets recogidos. Un aspecto importante del estudio era el análisis de sentimiento sobre el dato recogido para cada aerolínea. Se ha demostrado que la efectividad de las técnicas de análisis utilizadas dista mucho de ser perfecta al compararse con un análisis de sentimiento manual. Bien es cierto que VADER es capaz de detectar en mejor medida que TextBlob las emociones representadas en ciertos términos, pero ambas carecen de la habilidad de detectar la ironía usada por los pasajeros a la hora de publicar los tweets, por lo que, o bien, encontrar una herramienta que cumpla mejor dicha función, o conseguir mediante alguna regla intermedia que VADER sea capaz de detectar la ironía, se considera necesario. De todas formas, los resultados obtenidos por aerolínea a la hora de comparar las 4 sí que son válidos, y se observa un mayor porcentaje de críticas o quejas hacia la aerolínea American Airlines, que parece ofrecer un peor servicio hacia los pasajeros, o que sus vuelos sufren más contratiempos, por lo que dicha información podría ser usada para mejorar los servicios de la misma.

Se pretendía poder encontrar aquellos servicios en los que las aerolíneas destacaban y en los que, por el contrario, ofrecían un sentimiento bastante mejorable. En cuanto a lo positivo, apenas se detecta un tipo de patrón que es el del agradecimiento, aunque este está claramente perturbado por la interpretación de la ironía por parte del analizador de sentimiento. En cambio, sí que se pueden detectar ciertos patrones negativos en los tweets relacionados, en su mayoría, con los retrasos o cancelaciones de los vuelos, lo cual parece un problema recurrente, y que obviamente enoja a los pasajeros, haciendo que su confianza hacia la aerolínea disminuya, por lo que es un factor clave para poder mejorar sus servicios. Otros temas recurrentes son el servicio al cliente, la pérdida o daño del equipaje, de forma que una rápida respuesta por parte de los servicios de comunicación hacia el pasajero que publica dicha queja podría apaciguar el enfado, y no incrementar así la insatisfacción del cliente por el error cometido, ya que estos son fallos en los que una rápida respuesta por parte de la aerolínea puede incluso revertir la situación y que el pasajero no tenga una mala opinión hacia los servicios prestados. También se detectan otros patrones que son menos relevantes como pueden ser temas relacionados con el Covid19 o problemas meteorológicos que, aunque no son aspectos relacionados directamente con los servicios de la aerolínea, en muchos casos pueden ser dudas de los pasajeros que podrían ser resueltas generando una experiencia positiva. Para la obtención

de estos patrones se han usado diferentes técnicas de conversión, y se han extraído conclusiones similares a partir de todas ellas por lo que no parece que ninguna ofrezca ventaja sobre las demás mientras que la búsqueda de clústeres ha sido infructuosa, mostrando la dificultad de encontrar clústeres en este tipo de corpus con textos tan cortos, aunque los modelos de topics son una muy buena opción para averiguar los temas principales existentes en el corpus y permiten conocer la opinión que se genera a través de Twitter.

Se concluye que los pasajeros expresan sus opiniones y sus problemas a través de la red social Twitter, pudiendo ser usada esta información tanto para conocer qué servicios fallan de forma más frecuente, como los comentados a lo largo del estudio, como para dar una respuesta rápida al pasajero para aliviar la experiencia negativa, por lo que es una fuente de información muy válida para la gestión de aerolíneas, y se ha de buscar la forma más eficiente de utilizarla.

Finalmente, como líneas de trabajo a futuro que podrían mejorar el análisis realizado en este estudio destaca, como se ha comentado, la utilización de una herramienta que sea capaz de captar la ironía en los textos y así analizar con una mayor efectividad el sentimiento del texto, y poder comparar la posición de la aerolínea con respecto a su competencia en el marco de la interacción de los pasajeros a través de Twitter de una manera más fiable. Otra característica que se podría obtener, a pesar de la dificultad de encontrar clústeres, es una clasificación por topic, ya que sí que hay ciertos temas recurrentes, por lo que se podría hacer un filtrado para obtener dichos tweets que tratan sobre esos temas, bien por algún tipo de regla lingüística o mediante aprendizaje automático supervisado a través de una clasificación previa a un corpus obtenido con el que se pueda trabajar. Un uso de este tipo de clasificaciones permitiría generar grupos por problemas típicos de forma que los tweets fuesen asignados inmediatamente al grupo correspondiente, lo cual ayudaría a los servicios de comunicación a responder de una forma más rápida y eficiente, incrementando la satisfacción de los pasajeros y, por tanto, su confianza hacia la aerolínea en futuros vuelos.

Referencias

- Mazzeo, M.J. Competition and Service Quality in the U.S. Airline Industry. *Review of Industrial Organization* 22, 275–296 (2003). <https://doi.org/10.1023/A:1025565122721>
- Dogan Gursoy, Ming-Hsiang Chen, Hyun Jeong Kim, The US airlines relative positioning based on attributes of service quality, 2005, <https://doi.org/10.1016/j.tourman.2003.08.019>.
- Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019. Disponible online: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>
- Iqbal, Saqib & Zulqurnain, Ali & Wani, Yaqoob & Hussain, Khalid. (2015). The Survey of Sentiment and Opinion Mining for Behavior Analysis of Social Media. *International Journal of Computer Science & Engineering Survey (IJCSSES)*. 6. 21–27. 10.5121/ijcses.2015.6502.
- Jamilah, & Handayani, P. W. (2017). Analysis on effects of brand community on brand loyalty in the social media: A case study of an online transportation (UBER). In 2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016 (pp. 239–244). [7872746] (2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016). Institute of Electrical and Electronics Engineers Inc.. <https://doi.org/10.1109/ICACSIS.2016.7872746>
- Distribución de la cuota de mercado nacional de las aerolíneas estadounidenses entre abril de 2019 y marzo de 2020. Disponible online: <https://es.statista.com/estadisticas/598766/las-aerolineas-lideres-en-los-ee-uu-cuota-de-mercado-nacional/>
- Kumar, Akshi & Sebastian, Teeja. (2012). Sentiment Analysis on Twitter. *International Journal of Computer Science Issues*. 9. 372–378.
- Hasan, Ali & Moin, Sana & Karim, Ahmad & Band, Shahab. (2018). Machine Learning-Based Sentimental Analysis for Twitter Accounts. *Mathematical and Computational Applications*. 23. 11. 10.3390/mca23010011.
- TextBlob: Simplified Text Processing. Disponible online: <https://textblob.readthedocs.io/en/dev/>
- SentiWordNet. Disponible online: <https://github.com/aesuli/SentiWordNet>
- Eswaran, Prabhakar & Santhosh, M & Krishnan, A & Kumar, T. (2019). Sentiment Analysis of US Airline Twitter Data using New Adaboost Approach. *International Journal of Engineering and Technical Research*. 7. 1–3.

12. AdaBoost Classifier. Scikit Learn documentation. Disponible online: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html> 769
770
13. Wan, Yun & Gao, Qigang. (2015). An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis. 1318-1325. 10.1109/ICDMW.2015.7. 771
772
14. Hakh, Heba & Aljarah, Ibrahim & Al-Shboul, Bashar. (2017). Online Social Media-based Sentiment Analysis for US Airline companies. 773
774
15. Hall, Mark & Frank, Eibe & Holmes, Geoffrey & Pfahringer, Bernhard & Reutemann, Peter & Witten, Ian. (2008). The WEKA data mining software: An update. SIGKDD Explor. Newsl.. 11. 10-18. 775
776
16. Rustam, Furqan & Ashraf, Imran & Id, + & Mehmood, Arif & Ullah, Dr. Saleem & Choi, Gyu Sang & Khan, Yar. (2019). Tweets Classification on the Base of Sentiments for US Airline Companies. Entropy. 21. 10.3390/e21111078. 777
778
17. Ahsene Djaballah, Kamel & Kamel, Boukhalifa & Boussaid, Omar. (2019). Sentiment Analysis of Twitter Messages using Word2vec by Weighted Average. 223-228. 10.1109/SNAMS.2019.8931827. 779
780
18. Word2Vec. Disponible online: tensorflow.org/tutorials/text/word2vec 781
19. Khan, Rida & Urolagin, Siddhaling. (2018). Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction using Twitter Data. International Journal of Advanced Computer Science and Applications. 9. 10.14569/IJACSA.2018.090652. 782
783
20. Martin-Domingo, Luis & Martín, Juan Carlos & Mandsberg, Glen. (2019). Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). Journal of Air Transport Management. 78. 10.1016/j.jairtraman.2019.01.004. 784
785
21. TheySay. Disponible online: <https://innovation.ox.ac.uk/success-stories/theysay/> 786
22. Twinword. Sentiment Analysis. Disponible online: <https://www.twinword.com/api/sentiment-analysis.php> 787
23. Amigo Portillo, Adrián (2020). Análisis de sentimiento en Twitter. Reputación de las principales aerolíneas europeas tras la crisis sanitaria del covid-19 [Trabajo fin de Máster, Universidad Internacional de Andalucía] 788
789
24. Bing sentiment lexicon. Disponible online: https://search.r-project.org/CRAN/refmans/textdata/html/lexicon_bing.html 790
25. NRC Word-Emotion Association Lexicon. Disponible online: <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm> 791
792
26. What is CRISP DM? Disponible online: <https://www.datascience-pm.com/crisp-dm-2/> 793
27. Twitter API. Disponible online: <https://developer.twitter.com/en/docs/twitter-api> 794
28. Tweepy API. Disponible online: <https://docs.tweepy.org/en/stable/api.html> 795
29. Description of the emot:3.1 library. Disponible online: <https://github.com/NeelShah18/emot> 796
30. VADER-Sentiment Analysis. Disponible online: <https://github.com/cjhutto/vaderSentiment> 797
31. Spacy. Disponible online: <https://spacy.io/> 798
32. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL. 799
800
33. Albalawi, Rania & Yeap, Tet & Benyoucef, Morad. (2020). Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. Frontiers in Artificial Intelligence. 3. 10.3389/frai.2020.00042. 801
802
34. McInnes, Leland & Healy, John & Saul, Nathaniel & Grossberger, Lukas. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software. 3. 861. 10.21105/joss.00861. 803
804
35. pyLDAvis Documentation. Disponible online: <https://pyldavis.readthedocs.io/en/latest/modules/API.html> 805
806