

# Minería de datos en Twitter: análisis del sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19

Enrique Edgardo Condor-Tinoco

enricoti@unajma.edu.pe / Universidad Nacional José María Arguedas

Emerson Loa-Navarro

emer.loan@gmail.com / Universidad Nacional José María Arguedas

Josué Alexi Huarcaya-Ccoicca

jalexihuarc@gmail.com / Universidad Nacional José María Arguedas

Carlos Yinmel Castro-Buleje

ccastro@unajma.edu.pe / Universidad Nacional José María Arguedas

Recepción: 1/9/2020 Aceptación: 1/10/2020

**RESUMEN.** Con la aparición del COVID-19, se vino una crisis socioeconómica en cada país que toca esta pandemia; los países hispanohablantes son los más afectados por su estructura económica, es así que autores como Orgaz advierten que el COVID-19 ocasionará globalmente que, entre abril y junio del 2020, el 6,7 % de las horas de empleos desaparezca, lo que equivale a perder 195 millones de puestos de trabajo a tiempo completo. Con la finalidad de conocer lo que piensa la población, en la presente investigación se realizó un modelo que permitiría analizar el sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19 en la red social Twitter, haciendo uso de la metodología SEMMA y apoyados con la herramienta Orange Canvas. En los resultados se obtuvo que, de los seis mil tuits analizados, en los comentarios acerca del tema del desempleo, este tema está asociado con los términos *pobreza*, *pandemia*, *país* y *trabajo*; también se identificó que los usuarios tienen un sentimiento negativo y neutro en cuanto al tema de desempleo, considerando como un factor principal a la pandemia.

**PALABRAS CLAVE:** minería de datos / desempleo / análisis de sentimiento / COVID-19

## Data Mining on Twitter: Sentiment Analysis on Unemployment in the Spanish-Speaking Population in Times of COVID-19

**ABSTRACT.** The onset of the COVID-19 pandemic has caused a socioeconomic crisis in each country hit by the disease, being the Spanish-speaking countries the most affected due to their economic structure. Authors such as Orgaz warn that COVID-19 is expected to wipe out 6.7 % of working hours across the world only between April and June 2020, which is the equivalent of 195 million full-time workers losing their jobs. In order to know people's opinion, this research developed a model to analyze the emotional effects of unemployment in the Spanish-speaking population in times of COVID-19 on Twitter social network using the SEMMA methodology and the Orange Canvas tool. The results showed that, out of the 6,000 tweets analyzed, the unemployment is associated with the terms poverty, pandemic, country and work. It was also determined that Twitter users have a negative and neutral feeling regarding unemployment, and consider the pandemic as its main cause.

**KEYWORDS:** data mining / unemployment / sentiment analysis / COVID-19

## 1. INTRODUCCIÓN

El mundo se encuentra en una pandemia causada por el COVID-19; es una crisis de salud global. Desde su aparición a finales del año pasado, el virus ha llegado a casi todos los continentes. El COVID-19 es la enfermedad infecciosa causada por el coronavirus; este nuevo virus, como la enfermedad que provoca, eran desconocidos antes de que apareciera el brote en Wuhan (China) en diciembre del 2019 (Organización Mundial de la Salud, OMS, 2020). La OMS lo ha declarado como una pandemia desde el 11 de marzo del 2020.

La pandemia es más que una crisis de salud, es también una crisis socioeconómica; tiene el potencial de crear impactos: sociales, económicos y políticos, en cada país que la padece, dejando profundas y duraderas cicatrices. Sobre lo económico, la BBC News Mundo, en su portal web, publicó una nota periodística (abril del 2020) en donde se indica que “el efecto de la pandemia global de coronavirus sobre el empleo va a ser ‘devastador’ y superará con creces lo sucedido durante la crisis financiera mundial del 2008-2009, según un informe de la Organización Internacional del Trabajo” (Orgaz, 2020); en esa misma línea Orgaz advierte que el COVID-19 ocasionará globalmente que, entre abril y junio del 2020, el 6,7 % de las horas de empleos desaparezca, lo que equivale a perder 195 millones de puestos de trabajo a tiempo completo. En la página web del diario *Gestión* se señala que en Brasil se perdieron 7,8 millones de puestos de trabajo, Chile tiene su tasa más alta de desempleo en 10 años, en Bolivia el desempleo subió en casi dos puntos: de 5,7 % a 7,3 % y solo en abril, en Colombia, la tasa de desempleo urbano subió a un nivel más alto desde el año 2001 (“El empleo, otra víctima de la pandemia en América Latina”, 2020). En el Perú el desempleo aumentó a 16,3 % en el trimestre de abril a junio, un récord ocasionado por la cuarentena, que semiparalizó la economía de ese país. Este resultado trimestral constituye un aumento de 10 puntos porcentuales respecto al año 2019; la PEA (población económicamente activa) disminuyó en un 49,8 % y la tasa de desempleo fue de 6,3 a 16,3 %, en el segundo trimestre del 2020, también con relación al año 2019 (“Desempleo récord de 16,3 % en Perú durante confinamiento por pandemia”, 2020). “En el segundo trimestre del año, el producto bruto interno (PBI) se redujo en 30,2 % como consecuencia de las medidas adoptadas por el gobierno, destinadas a afrontar el COVID-19” (Instituto Nacional de Estadística e Informática, 2020).

La red social Twitter fue fundada por Jack Dorsey; es un servicio de microblogueo, una herramienta que limita sus mensajes a 280 caracteres. Twitter es una plataforma en línea para publicar y leer mensajes de texto de hasta 140 caracteres (Orihuela, 2011). En Twitter tienes a seguidores (*followers*) y los seguidos (*followed*); si sigues a alguien, eso quiere decir que verás sus tuits (mensajes que escribe en Twitter) (Rubín, 2020).

Los comentarios sobre el desempleo en Twitter en esta temporada de la pandemia se han incrementado y esto se debe justamente a la crisis de la economía en diferentes países que son afectados por esta enfermedad; y las opiniones son diversas. Con los datos de Twitter sobre el desempleo y el *software* Orange Canvas 3.2 se ha realizado minería de datos sobre el análisis de sentimiento. La minería de datos (MD) es “la metodología que permite explotar los datos

con el objetivo de generar modelos que posibiliten describir, encontrar patrones, establecer agrupaciones, clasificar, segmentar o asociar productos, clientes o cualquier otra entidad con el objeto de obtener conocimiento” (Rámila, 2019). El análisis de sentimientos es una técnica para detectar diversas opiniones favorables y desfavorables hacia temas específicos (Nasukawa y Yi, 2003), y Liu (2010) define el análisis de sentimientos como el estudio que analiza las opiniones de las personas, sentimientos, apreciaciones, evaluaciones y emociones sobre entidades como productos, servicios, individuos, eventos, tópicos y sus atributos.

El objetivo de la presente investigación es desarrollar un modelo para analizar el sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19 en la red social Twitter, mediante la metodología SEMMA que fue desarrollada por el Instituto SAS. SEMMA significa *sample, explore, modify, model y assess*, considera cinco etapas para su proceso, las cuales son estudiadas en la metodología de la presente investigación.

Se ha descrito el problema y se ha planteado el objetivo de la investigación en la introducción; además, se han descrito las herramientas y la metodología que se usa en el desarrollo de la investigación. Se explica, paso a paso, un modelo de minería de datos con respecto al análisis del sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19, haciendo uso de la metodología SEMMA; después, se muestran los resultados obtenidos, y al final se dan las conclusiones de la investigación.

## 2. METODOLOGÍA

SEMMA fue desarrollada por el instituto SAS que es el acrónimo de *sample, explore, modify, model y assess*. Su traducción sería muestra, explora, modifica, modela y evalúa; metodología para desarrollar el proceso de minería de datos.

El desarrollo de la investigación está basado en los pasos de la metodología SEMMA, siguiendo la secuencia descrita en la figura 1; la herramienta usada para el procesamiento de los datos es Orange Canvas. Dichos datos fueron extraídos de la red social Twitter; el tema es sobre el análisis del sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19.

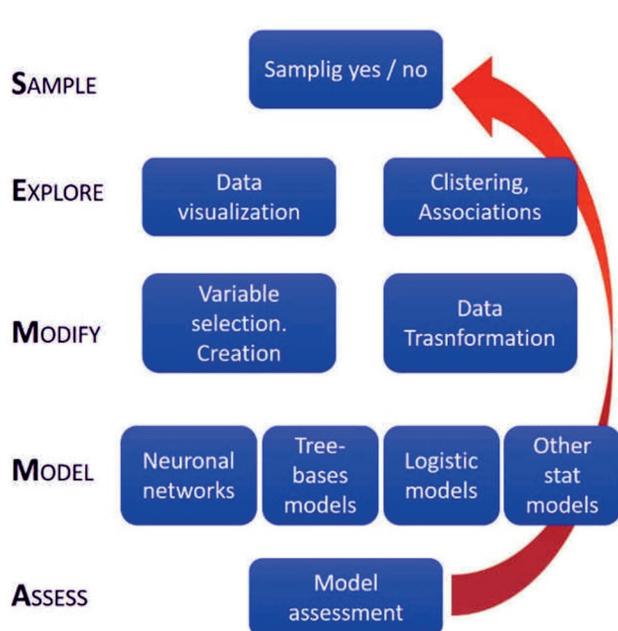


Figura 1. Metodología SEMMA

Fuente: Romero (2019)

## 2.1 Recursos utilizados

Software Orange Canvas, versión 3.2, con *widgets* descritos en la siguiente tabla:

Tabla 1  
Widgets de Orange Canvas usados en la investigación

Widget	Función
 Twitter	Envía datos al servidor de Twitter mediante una consulta realizada, después recupera una colección de tuits.
 Corpus Viewer	Permite la visualización de archivos de texto o instancias de corpus.
 Preprocess Text	Recibe una colección de documentos (en este caso tuits), luego los divide en <i>tokens</i> , los filtra y ejecuta la derivación y lematización.
 Word Cloud	Visualiza <i>tokens</i> en el corpus; su tamaño indica la frecuencia de la palabra en el corpus.
 Sentiment Analysis	Predice el sentimiento de los documentos asignando una puntuación.

(continúa)

(continuación)

 Data Table	Recibe unos conjuntos de datos y muestra el resultado como una hoja de cálculo de Excel.
 Select Columns	Permite seleccionar una o varias columnas, componer manualmente el dominio de datos.
 Linear Projection	Muestra proyecciones lineales de los datos etiquetados en la clase.
 Distributions	Muestra las distribuciones de los valores de un atributo.

Elaboración propia

## 2.2 Metodología SEMMA

### 2.2.1 Muestreo de los datos

En el *software* Orange Canvas, se procedió a recolectar y extraer los datos con el *widget* Twitter; nos conectamos a la API de dicha red social, previamente solicitadas las credenciales de acceso. Se hizo la consulta de la palabra *desempleo*, como resultado se obtuvo un total de seis mil tuits procedentes de los mensajes e interacciones de los usuarios. Estos mensajes e interacciones se recopilieron desde los siete días anteriores respecto a la fecha consultada (27 de agosto del 2020), ya que el *widget* Twitter obtiene los mensajes solo con una anterioridad de siete días. Se visualizan los comentarios extraídos mediante el *widget* Corpus Viewer (figura 2).

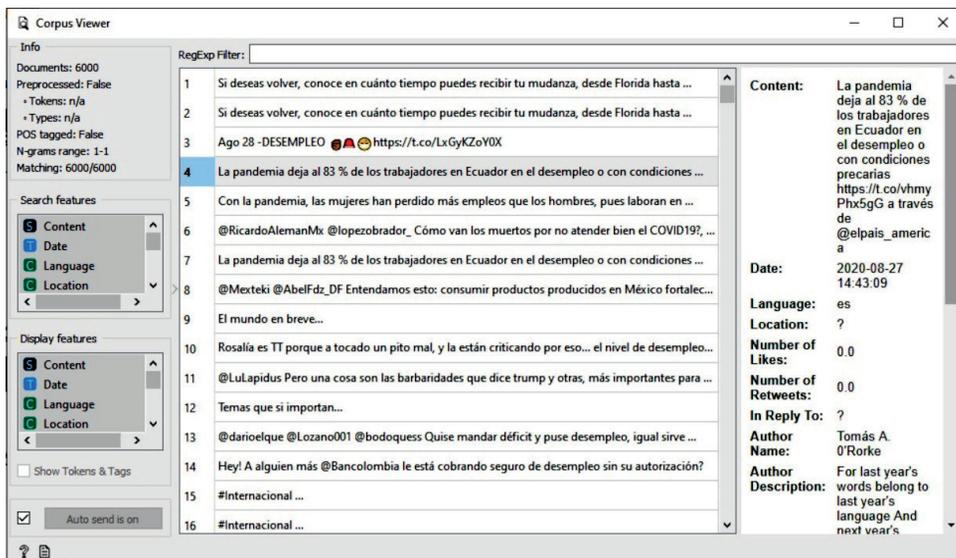


Figura 2. Visualización de tuits mediante Corpus Viewer

Elaboración propia en Orange Canvas

### 2.2.2 Exploración de los datos

Se ha realizado una exploración de los datos mediante el *widget* Word Cloud, para ver las frecuencias de las palabras o cantidad de veces que aparece una palabra en los tuits; se usó el *widget* Sentiment Analysis, dicha herramienta nos proporciona el módulo de VADER para poder analizar los sentimientos: evaluar y puntuar como positivo, negativo, neutro y compuesto a cada tuit, donde las puntuaciones antes mencionadas se basan en la validación empírica de múltiples jueces humanos independientes. VADER incorpora un léxico de sentimiento estándar que está especialmente en sintonía con contextos similares a los de microblogs (Hutto y Gilbert, 2014). Se usó el *widget* Distribution (gráfico de distribución) y Linear Projection (proyección lineal) para determinar las puntuaciones para cada comentario: positivo, negativo, neutro y compuesto.

### 2.2.3 Modificación de datos

Cuando se extrajeron los datos, se tenían muchos que no eran materia del propósito de la investigación, por lo que se procedió a realizar el preprocesamiento del texto de los comentarios, utilizando el *widget* Preprocess Text, dividiendo en unidades más pequeñas (*tokens*), como también omitiendo las puntuaciones, filtrando, ejecutando derivación y lematización; este paso se realizó de manera secuencial. Primeramente, se transformó el texto en minúsculas, enseguida se eliminaron todos los signos diacríticos y acentos, como también las URL y las etiquetas HTML; se procedió también a la eliminación de las palabras vacías como *y*, *o*, *en*. Las palabras por eliminar fueron cargadas desde un archivo propio en formato .txt el cual contenía una lista de palabras vacías.

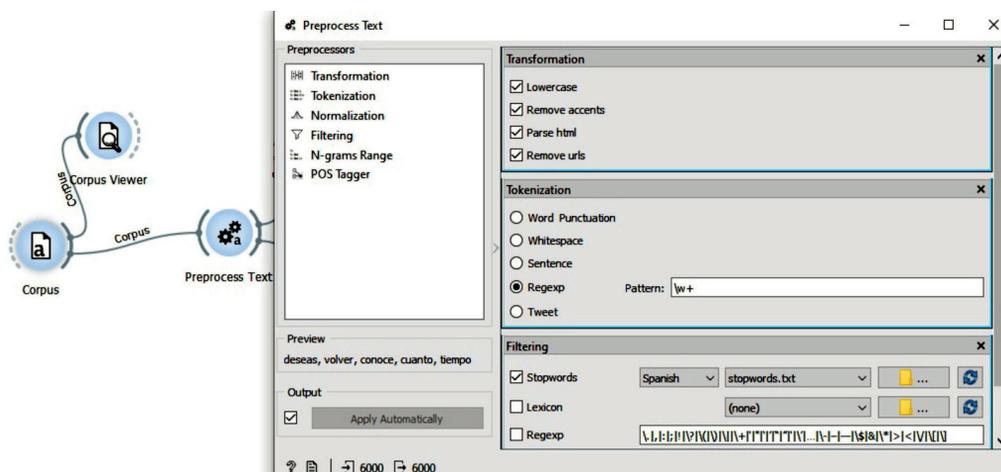


Figura 3. Reprocesamiento de texto (los tuits)

Elaboración propia en Orange Canvas

Mostramos los *tokens* de toda la colección de tuits, obteniendo datos más “limpios” y legibles, mediante el *widget* Word Cloud.

### 2.2.4 Modelado de datos

El modelo del análisis del sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19 culminado es como se muestra en la figura 4.

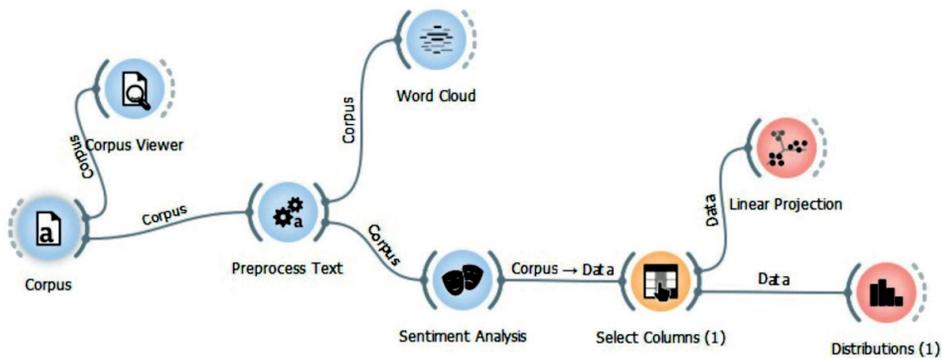


Figura 4. Diseño del modelo del análisis del sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19

Elaboración propia en Orange Canvas

### 2.1.5 Evaluación de datos

Una vez obtenidos datos más legibles, se aplicó el *widget* Sentiment Analysis; mediante este componente se analizaron los sentimientos, el cual evalúa y puntúa como positivo, negativo, neutro y compuesto. Para el análisis, el *widget* Sentiment Analysis nos provee dos módulos de sentimiento: el primero de ellos es el método de Liu Hu; el segundo método que se va a utilizar para esta investigación es el método VADER (*Valence Aware Dictionary and Entiment Reasoner*). Este es una herramienta de análisis de sentimientos basada en reglas y léxico que está específicamente en sintonía con los sentimientos expresados en las redes sociales. VADER usa una combinación de un léxico de sentimiento, es una lista de características léxicas que generalmente se etiquetan de acuerdo con su orientación semántica como positiva o negativa. El léxico de sentimientos de VADER es sensible tanto a la polaridad como a la intensidad de los sentimientos expresados (Hutto y Gilbert, 2014). Respecto a la puntuación compuesta se calcula sumando las puntuaciones de valencia de cada palabra en el léxico, se ajusta de acuerdo con las reglas y luego se normaliza para estar entre  $-1$  (más extremo negativo) y  $+1$  (más extremo positivo). Los valores umbrales típicos son:

- sentimiento positivo: compuesto  $\geq 0,05$
- sentimiento neutral: compuesto  $> -0,05$  y compuesto  $< 0,05$
- sentimiento negativo: compuesto  $\leq -0,05$

En cambio, los puntajes positivo, negativo y neutro representan a las proporciones de texto que caen en cada categoría (por el que todos deben sumar 1 o cerca de él con la operación flotante).

Este método VADER está dentro de Natural Language Toolkit (NLTK), es una plataforma de código abierto y gratis, para construir programas en lenguaje Python que trabajen con datos de lenguaje humano. Contiene librerías de procesamiento de texto para clasificar, *tokenización*, *stemming*, *tagging*, parseo y razonamiento semántico, además de librerías robustas para Natural Language Processing (NLP) (González Figueroa, 2018).

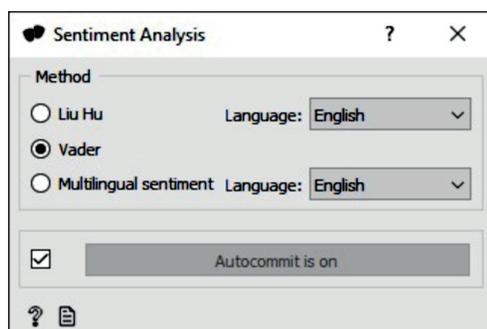


Figura 5. Selección método Vader NLTK en el *widget* Sentiment Analysis

Elaboración propia en Orange3 Text Mining

Se usó el *widget* Select columns para filtrar solo las puntuaciones (figura 6): positiva, negativa, neutra y compuesta; de esta manera, visualizar los resultados mediante los *widgets* Word Cloud (nube de palabras), Linear Projection (Proyección lineal) y Distribution (gráfico de distribución).

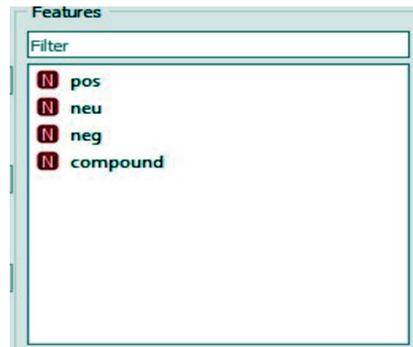


Figura 6. Filtrando solo las puntuaciones mediante el *widget* Select Columns  
Elaboración propia en Orange Visual Programming

### 3. RESULTADOS

El *widget* Word Cloud realiza el conteo de las palabras de acuerdo a la frecuencia con que estas aparecen, con ello se obtiene en total 21 131 palabras contadas, donde la palabra *desempleo* se menciona 5998 veces, lo que representa el 28 % del total, seguido por la palabra *pobreza* con un total de 704 repeticiones, lo que representa el 3,33 % del total, y la palabra *pandemia*, con una frecuencia de 637 menciones, lo que representa el 3,03 %. Son las tres primeras palabras que aparecen, según la frecuencia, según se muestra en la lista en la figura 7.

De esta manera, se concluye que en la red social Twitter los comentarios realizados por los usuarios hispanohablantes acerca del tema del desempleo están estrechamente relacionados con la pandemia y la pobreza.

Weight	Word
5998	desempleo
704	pobreza
637	pandemia
521	pais
468	economia
408	trabajo
403	solo
399	gobierno
398	empleo
368	gente
353	reales
333	seguro

Figura 7. Lista de palabras según su frecuencia, *widget* World Cloud  
Elaboración propia en Orange3 Text Mining

Los siguientes resultados se muestran mediante el *widget* Distribution (gráfico de distribución) y Linear Projection (proyección lineal); para ello, hemos seleccionado las variables a mostrar mediante el *widget* Select Columns; estas variables son las puntuaciones para cada comentario: positivo, negativo, neutro y compuesto.

### 3.1 Puntuación positiva

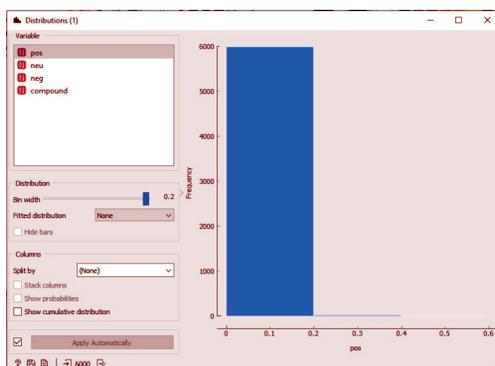


Figura 8. Distribución, puntuación positiva  
Elaboración propia en Orange Canvas

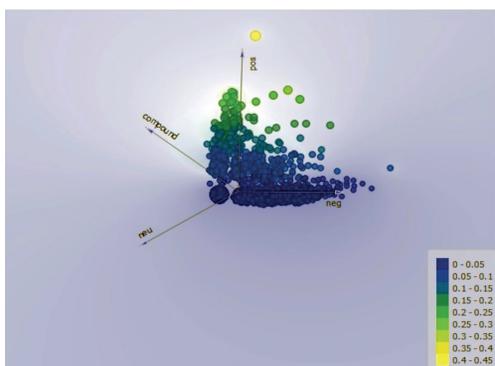


Figura 9. Proyección lineal, puntuación positiva  
Elaboración propia en Orange Canvas

En la figura 8, tenemos el gráfico de distribución, donde las barras representan la frecuencia de los comentarios con base en el nivel de puntuación. Estas puntuaciones representan el nivel de intensidad del sentimiento (positivo, negativo o neutro) que contiene el conjunto de tuits. Un puntaje cercano a 1 significa que hay mayor cantidad de tuits con intensidad de sentimientos (positivos, negativos o neutros) en todos los comentarios tomados. Las barras de color

azul de la figura 8 nos muestran que 5983 tuits tienen una puntuación de entre 0 y 0,2, lo cual las sitúa en los niveles de positividad, lo que representa el 99,72 % del total de 6000 tuits que se tomaron para el estudio. A ello se siguen 16 tuits con una puntuación de entre 0,2 y 0,4, lo que representa el 0,27 %. Y se finaliza con un tuit que tiene la puntuación mayor o igual a 0,4, lo que representa el 0,02 % del total. De la misma forma, se observan las métricas en la figura 9, donde el color azul alude al conjunto de datos con las puntuaciones antes mencionadas y en que la cantidad de las burbujas representa la predominancia de los sentimientos positivos en todo el conjunto de datos. Se concluye que se tiene una intensidad baja de positividad, considerando que el 99,72 % de tuits se encuentra entre 0 y 0,2.

### 3.2 Puntuación negativa

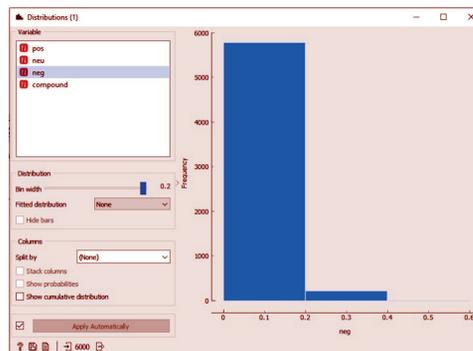


Figura 10. Distribución, puntuación negativa  
Elaboración propia en Orange Canvas

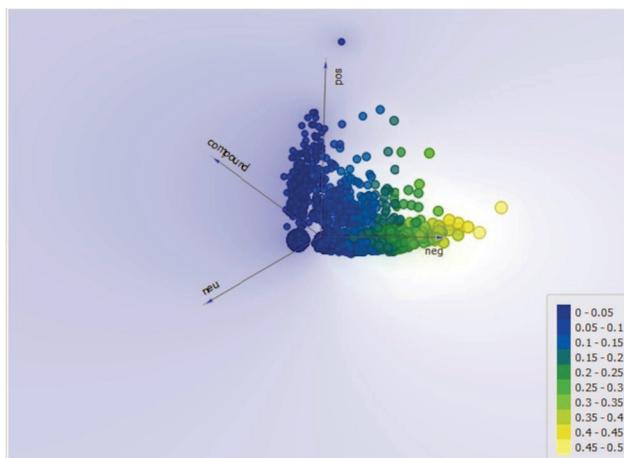


Figura 11. Proyección lineal, puntuación negativa  
Elaboración propia en Orange Canvas

Acerca de la puntuación negativa, en la figura 10 se muestra un total de 5779 tuits con una puntuación de entre 0 y 0,2, lo cual representa el 96,32 % del total del conjunto de datos, seguida por 214 tuits con puntuaciones de entre 0,2 y 0,4, lo que representa el 3,57 % del total. La figura finaliza con siete tuits con puntuaciones de entre 0,4 y 0,6 con un porcentaje de 0,12 %. La predominancia de la intensidad del sentimiento negativo se observa en la figura 11, la cual, a comparación de la puntuación positiva, tiene mayor predominancia. Se concluye que se tiene una intensidad baja de negatividad, considerando que el 96,32 % de tuits se encuentra entre 0 y 0,2.

### 3.3 Puntuación neutra

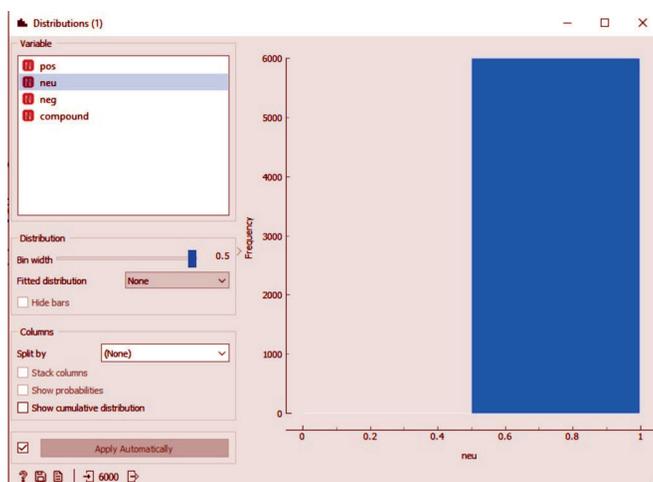


Figura 12. Distribución, puntuación neutral

Elaboración propia en Orange Canvas

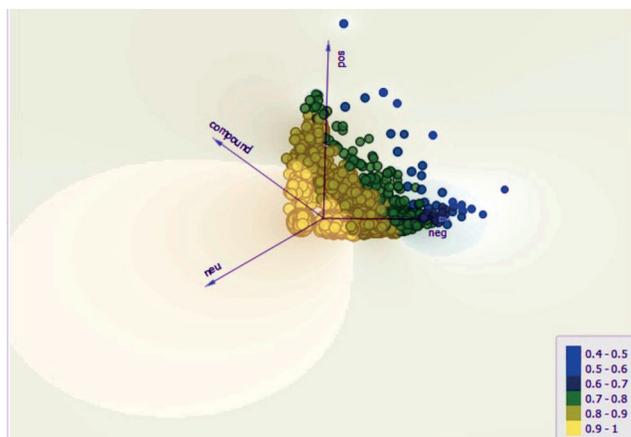


Figura 13. Proyección lineal, puntuación neutral

Elaboración propia en Orange Canvas

En la puntuación neutral (figura 12) los tuits obtuvieron un nivel de puntuación de entre 0 y 0,5, lo cual representa el 0,02 % del total de tuits. En cuanto al tuit restante, representa el 99,98 % con una puntuación de entre 0,5 y 1. En la figura 13 se observa mayor predominancia de la intensidad de los sentimientos neutrales en comparación con los negativos, donde el color de las burbujas representa las puntuaciones mencionadas anteriormente. Se concluye que se tiene una intensidad alta de neutralidad, al considerar que el 99,98 % de tuits se encuentra entre 0,5 y 1.

### 3.4 Puntuación compuesta

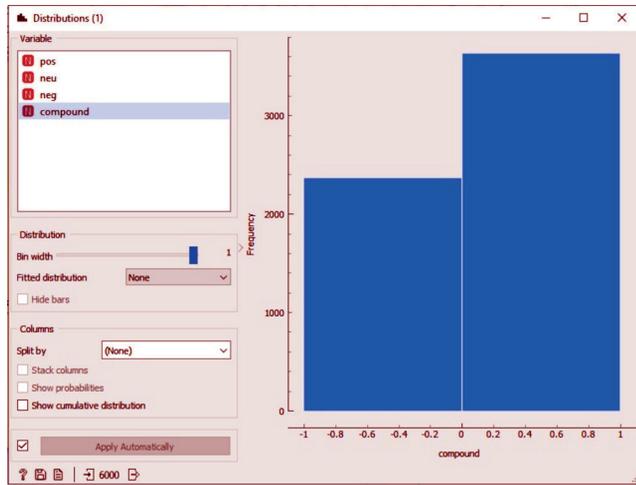


Figura 14. Distribución, puntuación compuesta  
Elaboración propia en Orange Canvas

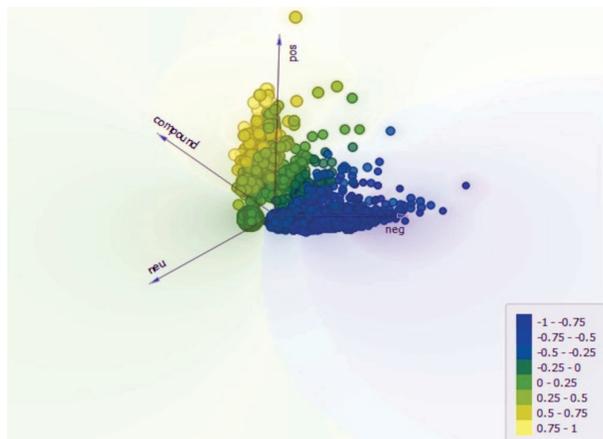


Figura 15. Proyección lineal, puntuación compuesta  
Elaboración propia en Orange Canvas

La puntuación compuesta muestra la puntuación total de un comentario, cuanto más cercano a 1 sea el nivel puntuación es más positivo; de lo contrario, se acerca a  $-1$  y tiende a ser más negativo. Si observamos la figura 14, 2367 tuits tienen una puntuación de entre  $-1$  y 0, lo cual representa el 39,45 % del total de tuits; en cambio, las puntuaciones de entre 0 y 1 se obtuvieron en 3633 tuits que representan el 60,55 % del total. En la proyección lineal (figura 15) también se observa la predominancia del sentimiento compuesto.

En la figura 16 se muestran las puntuaciones compuestas; en este caso se ajustó el gráfico de distribuciones a los valores umbrales típicos para, de esta manera, determinar la puntuación total de los comentarios de manera unidimensional con base en los valores umbrales.

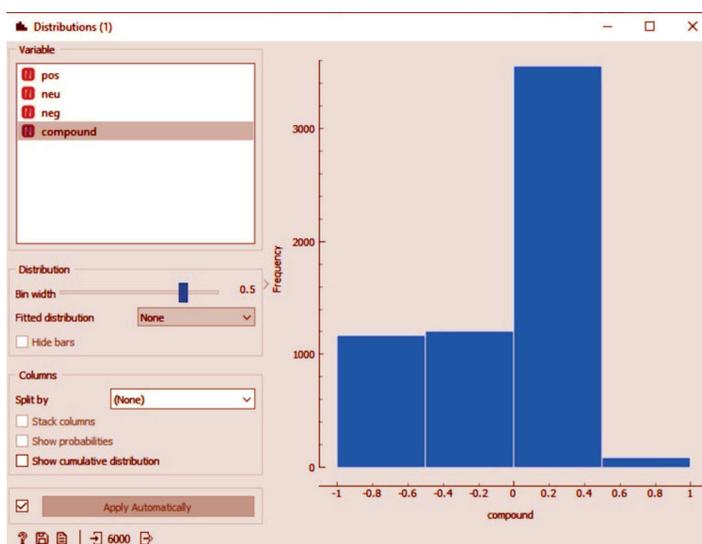


Figura 16. Distribución, valores umbrales  
Elaboración propia en Orange Canvas

El primer umbral es la puntuación entre  $-1$  y  $-0,05$ ; este rango representa la determinación de que un comentario es negativo. Es así que, según la figura 16, la primera barra representa el primer umbral con 1165 tuits como negativos, lo que equivale al 19,42 % del total.

Los segundos valores umbrales están en el rango de  $-0,05$  y  $0,05$ , determinando en este rango los comentarios neutros. La segunda y la tercera barra aluden a estas puntuaciones donde dichas barras representan 4753 tuits como neutros, lo que equivale al 79,21 % del total.

Como tercer umbral tenemos los valores entre  $0,05$  y  $1$ , que es el rango de los comentarios positivos. La cuarta barra de la figura 15 representa a 82 tuits como positivos con un porcentaje de 1,37 % del total.

#### 4. CONCLUSIONES

- a. Para el análisis del sentimiento del desempleo en la población hispanohablante en tiempos del COVID-19 en la red social Twitter, se procesaron seis mil tuits; en los comentarios realizados por los usuarios hispanohablantes se encontró que las cinco palabras más utilizadas son: *desempleo* (5998 veces), *pobreza* (704 veces), *pandemia* (637 veces), *país* (521 veces), *economía* (468 veces) y *trabajo* (408 veces), lo que indica que la mayoría de los tuits estaban compuestos por estas palabras. Los datos se obtuvieron con la palabra *desempleo* y esta a su vez vinculó a términos como *pobreza*, *pandemia*, *país* y *economía*.
- b. En los tuits de los usuarios hispanohablantes tomados los días 25, 26 y 27 de agosto del 2020, el resultado de la puntuación compuesta muestra un sentimiento neutral (79,21 %), seguido de un sentimiento negativo (19,42 %) y solo un 1,37 % tiene sentimiento positivo sobre este tema.

#### 5. RECOMENDACIONES

- a. Para un análisis más profundo que incluya una mayor cantidad de datos, es necesario recopilar dicha información en períodos de tiempo, después aplicar el modelo desarrollado en Orange Canvas. De este modo se obtendrán resultados más ajustados a la realidad.
- b. En cuanto a la limpieza de los datos, se debe armar una lista de las palabras y símbolos que no representan dato para el análisis con la finalidad de evitar consumir tiempo en el procesamiento de los datos.

#### REFERENCIAS

- Artacho Torres, P. (2014). *Minería de opiniones basada en patrones semánticos, en Twitter* [Tesis de grado]. Universidad de Málaga.
- El empleo, otra víctima de la pandemia en América Latina. (7 de julio del 2020). *Gestión*. <https://gestion.pe/mundo/internacional/el-empleo-otra-victima-de-la-pandemia-en-america-latina-noticia/?ref=gesr>
- González Figueroa, M. (2018). *Búsqueda, descubrimiento y análisis de sentimientos en Twitter de la institución* [Tesis de maestría, Universidad Técnica Federico Santa María]. [https://www.mti.cl/wp-content/uploads/2019/01/Tesina\\_2018\\_Gonz%C3%A1lez-Marcela.pdf](https://www.mti.cl/wp-content/uploads/2019/01/Tesina_2018_Gonz%C3%A1lez-Marcela.pdf)

- Hutto, C., y Gilbert, E. (Junio del 2014). *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*. International Conference on Weblogs and Social Media (ICWSM-14). <https://www.scinapse.io/papers/2099813784>
- Desempleo récord de 16,3 % en Perú durante confinamiento por pandemia. (17 de julio del 2020). *Infobae*. <https://www.infobae.com/america/agencias/2020/07/17/desempleo-record-de-163-en-peru-durante-confinamiento-por-pandemia/>
- Instituto Nacional de Estadística e Informática. (20 de agosto del 2020). *Producto bruto interno se redujo 30,2 % durante el II trimestre del 2020* [Nota de prensa]. <https://www.inei.gob.pe/prensa/noticias/producto-bruto-interno-se-redujo-302-durante-el-ii-trimestre-del-202-12350/>
- Liu, B. (2010). Sentiment Analysis and Subjectivity. En N. Indurkha y F. J. Damerau (Eds.), *Handbook of Natural Language Processing* (2.ª ed., pp. 627-666). Chapman & Hall/CRC.
- Mitra, S., y Acharya, T. (2005). *Data Mining. Multimedia, Soft Computing and Bioinformatics*. Wiley-Interscience.
- Nasukawa, T., y Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language. En *K-CAP '03: Proceedings of the 2nd International Conference on Knowledge* (pp. 70-77). Association for Computing Machinery. <https://doi.org/10.1145/945645.945658>
- Organización Mundial de la Salud. (2020). *Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19)*. <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>
- Orgaz, C. (8 de abril del 2020). *Coronavirus: “Se perderán 195 millones de empleos en solo 3 meses” por la pandemia, el alarmante informe de la OIT (y cómo afectará a América Latina)*. BBC Mundo. <https://www.bbc.com/mundo/noticias-america-latina-52220090>
- Orihuela, J. L. (2011). *Mundo Twitter: una guía para comprender y dominar la plataforma que cambió la red*. Alienta.
- Rámila, H. (29 de julio del 2019). Minería de datos: cómo aplicarlo al análisis predictivo. *Semrush Blog*. [https://es.semrush.com/blog/mineria-de-datos/?kw=ycmp=LM\\_SRCH\\_DSA\\_Blog\\_Analytics\\_ES&label=dsa\\_blogyNetwork=gyDevice=cyutm\\_content=438280918963&kwid=dsa-835963616070&ycmpid=8044598079&yclid=EAIAIQobChMI0favs8LA6wIVxIVaBR117AezEAAYAiAAEgK2EPD\\_BwE](https://es.semrush.com/blog/mineria-de-datos/?kw=ycmp=LM_SRCH_DSA_Blog_Analytics_ES&label=dsa_blogyNetwork=gyDevice=cyutm_content=438280918963&kwid=dsa-835963616070&ycmpid=8044598079&yclid=EAIAIQobChMI0favs8LA6wIVxIVaBR117AezEAAYAiAAEgK2EPD_BwE)
- Romero, J. (11 de junio del 2019). *Metodologías de minería de datos*. <https://jorgeromero.net/metodologias-de-mineria-de-datos/>
- Rubín, R. (22 de mayo del 2020). Qué es Twitter, cómo funciona y qué te puede aportar esta red social. *Ciudadano 2.0*. <https://www.ciudadano2cero.com/twitter-que-es-como-funciona/>