TRABAJO DE GRADO APLICADO

PRUEBA DE CONCEPTO – PSICOLINGUÍSTICA – MODELOS DE SEGMENTACIÓN PARA IDENTIFICAR TENDENCIAS O PATRONES INDICADORES DE DEPRESIÓN EN REDES SOCIALES

CAOBA – PONTIFICIA UNIVERSIDAD JAVERIANA

Andrés Eduardo Mendoza Molina

Jorge Eduardo Enciso Agudelo

Asesores

Juan Pablo Pájaro Hernández

Juan Pablo Mora López

Mayo 2022

Maestría en Analítica para la Inteligencia de Negocios

Facultad de Ingeniería

Pontificia Universidad Javeriana

Bogotá D.C

CONTENIDO

CONTENIDO DE TABLAS

CONTENIDO DE GRÁFICOS

1. ENTENDIMIENTO DE NEGOCIO

1.1. Contexto del Negocio

Una buena salud mental es uno de los principales motivos por el cual las personas pueden interactuar con otras personas de una forma que se pueda crear una relación y como se toman las decisiones. Asimismo, en los últimos años esta ha sido uno de los principales motivos por los cuales las personas se enferman física y mentalmente, los factores que afectan a este estado mental son muchos como lo son la política, el medio ambiente, la religión, la educación, etc. Por lo que no se puede atribuir solo ha un factor en específico. De acuerdo con la organización mundial de la salud (WHO) [7], es muy común que los pacientes que tienen mala salud mental sufran de trastornos depresivos, lo que provoca que síntomas como la fatiga, insomnio, pérdida de apetito, entre otros, se manifiesten en la vida diaria de las personas afectando su rendimiento, lo cual puede llegar a acabar en ideas suicidas o intentos de suicidio. En el 2018 la WHO estimó que alrededor de 300 millones de personas habían sido diagnosticadas con depresión en la última década. Este indicador ha incrementado a lo largo del tiempo y se estima que desde el 2005 al 2015 ha tenido un crecimiento aproximado del 18%. [15]

Hoy en día, más del 90% de la sociedad está conectada en internet, por lo cual el sector de la salud ha decidió observar con detalle este ecosistema donde las personas interactúan de forma virtual y pueden expresar sus ideas libremente, la gente que sufre de depresión usualmente usa estas redes sociales para compartir sus experiencias, buscar ayuda o con consejos y saber cómo manejar su vida social[6].Por lo cual, las historias que comparten o sus estados en estas tienden a un comportamiento similar, ya que resaltan palabras claves similares y la forma en la que hablan es diferente a otros usuarios que también usan la misma plataforma. Por esta razón, la tecnología relacionada a procesamiento de texto usando como técnicas de procesamiento de lenguaje natural o aprendizaje de máquina permite a los investigadores encontrar estos patrones de comportamiento que caracterizan a esta gente, ayudándolos a prevenir que los trastornos mentales no se desarrollen.

Twitter es una de las redes sociales más famosas en la última década debido a su facilidad de publicar los pensamientos de los usuarios que la usan en tan poco tiempo. El número de tweets por día puede estar alrededor de 58 millones sin contar los comentarios asociados a estos[16]. Al ser una de las redes sociales con mayor número de usuarios activos, el factor de anonimidad les da oportunidad a las personas a mostrar sus verdaderos pensamientos, por lo que también es una de las redes sociales con mayor toxicidad y contenido relacionado al acoso coloquialmente entendido como Bullying. Debido a la cantidad de información de texto que se maneja en esta plataforma, los investigadores neurolingüísticas pueden encontrar muchos comportamientos que la gente publica por su forma de hablar y cómo interactúan con otros. Por esta razón, es primordial que la ciencia de datos brinde herramientas y conclusiones relevantes para ayudar al sector de la salud para que así puedan tomar mejores decisiones respecto a las iniciativas para identificar posibles usuarios que puedan tener síntomas de depresión o ansiedad.

1.2. Revisión Bibliográfica (Estado del Arte)

Al realizar una investigación bibliográfica, se encontraron alrededor de 40 documentos relacionados con el tema a tratar y que pueden apalancar la prueba de concepto que se está desarrollando, los artículos investigados son definidos como títulos en su mayoría publicaciones e investigaciones que hablan de cómo realizar un proceso de detección de elementos relacionados con depresión en personas jóvenes que hacen uso de redes sociales como medio de expresión de su situación psicológica [3].

Ahora, para comprender mejor cómo la investigación bibliográfica ayuda a entender y modelar el problema principal, el cual consiste en identificar patrones y comportamientos manifestados por personas con diagnóstico confirmado o con tendencias depresivas, se deben listar brevemente las formas, métodos y modelos más empleados para capturar información relevante a esas conductas; si bien esta literatura provee en su mayoría elementos descriptivos, hay documentos que nos muestran el uso de técnicas avanzadas de analítica como el uso de Redes Neuronales profundas [14].

Algoritmos	Cantidad
NLP (TDIF)	8
Arboles de decisión	5
SVM	5
NLP (Análisis de Sentimientos)	4
NLP (LIWC)	4
Bayes Ingenuo	2
NLP (Análisis de Polaridad)	2
NLP (Lista de Palabras)	2
NLP (Detección de Emociones)	2
DNN (Redes Neuronales Profundas)	1
Red Bayesiana	1
TAN (Redes Neuronales Transducidas)	1
BPNN (Redes Neuronales de Propagación Inversa)	1
Clústeres (KNN o Similar)	1
Total	39

Tabla 1 Resumen de algoritmos listados en la investigación

Como conclusión del ejercicio de validación bibliográfica, se determinó que en términos prácticos, el uso de técnicas basadas en NLP es una de las metodologías más usadas para desarrollar algoritmos que nos permita identificar patrones o estresores relacionados con la depresión en personas tendientes a esta enfermedad basados en palabras clave [7], por otro lado, en promedio, se habla de dos (2) algoritmos por documento investigativo consultado de veinte (20) intrínsecamente relacionados con el objeto de este trabajo analítico, siendo un valor cercano al 50% del total de los métodos empleados.

1.3. Objetivos

1.3.1. Objetivo de Negocio

Detectar tendencias de perfiles o comentarios que posiblemente tengan rasgos o características de síntomas de depresión o algún trastorno parecido a este como lo es la ansiedad o los pensamientos suicidas, con el fin de que la Pontificia Universidad Javeriana reduzca el tiempo de reacción y se pueda tomar decisiones antes de que se desarrollen los síntomas.

1.3.1.1. Objetivos específicos del negocio

1. Describir tendencias que los usuarios tienen cuando escriben en las redes sociales sobre temas relacionados a la depresión.
2. Consolidar el informe de resultados de las redes sociales, con el fin de que sea puedan tomar decisiones a nivel estratégico en la Pontificia Universidad Javeriana.

1.3.2. Objetivos de *minería de datos*

Realizar un modelo de procesamiento de texto para poder identificar patrones de comportamiento y palabras claves, las cuales permitan caracterizar a los usuarios que tienden a sufrir síntomas de ansiedad o depresión.

1.3.2.1. Objetivos de minería de datos específicos

Identificar y caracterizar palabras claves y grupos de palabras que permitan identificar síntomas de estrés o de depresión validados por expertos en el tema.

1.4. Supuestos y restricciones del proyecto

a) Debido a que existen múltiples niveles de depresión o ansiedad, se tomara como supuesto que todos solo se categorizara un tipo de depresión, por lo cual no habrá discriminación a ningún tipo de variante que esta pueda tener.
b) Las bases de datos suministradas por el equipo de CAOBA estarán basadas únicamente en Twitter y se contemplarán solo tweets que estén en el idioma español.
c) No se podrá asumir que los usuarios de Twitter que consolidaron la base de datos sufren de depresión, solo se podrá asumir que pueden estar sintiendo síntomas relacionados con depresión o ansiedad.

1.5. Criterios de éxito

Identificar patrones del lenguaje y patrones de los usuarios que permitan identificarlos frente a un grupo control y que los expertos del negocio, que este caso serán psicólogos. Puedan confirmar que las características encontradas son acordes a los síntomas de depresión presentados en la teoría.

1.6. Hardware y Software

Procesador: Intel(R) Core™ i5 – 6300U CPU@ 2.30GHz.

Memoria RAM: 8 GB

Sistema Operativo: Windows 10

Software: Python 3.8, Pandas 1.4.1, Sklearn 0.16.1, Spacy 1.8.2, Gensim 3.8.3

1.7. Flujograma del proyecto

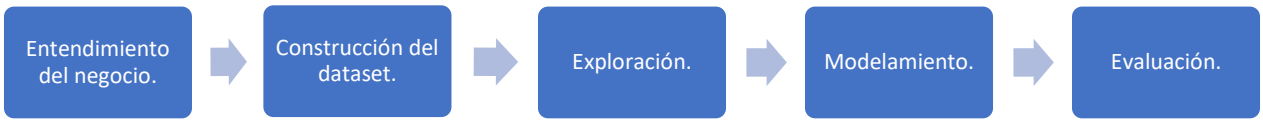


Gráfico 1 Flujograma del proyecto

Entendimiento del negocio: Comprender el estado del arte de la conexión entre la red social de Twitter y los síntomas relacionados con depresión y ansiedad identificando posibles hipótesis a probar en la base de datos creada.

Construcción de la base de datos: Creación de una base de datos de tweets relacionada con palabras importantes encontradas en los lexicones psicológicos y conclusiones de otras investigaciones.

Exploración: Se hará una exploración de datos con los tweets usando técnicas de NLP para encontrar patrones lingüísticos que puedan tener los usuarios que tienden hablar relacionados a temas relacionados con depresión o ansiedad.

Modelamiento: Se usarán técnicas de aprendizaje no supervisado y semi supervisado para poder encontrar grupos de palabras y tendencias de palabras claves que permitan identificar comportamientos específicos de personas que sufren trastornos.

Evaluación: Se comprobarán los resultados obtenidos de la exploración de datos y el modelamiento con un psicólogo/psiquiatra o doctor que permitan validar los resultados encontrados después del modelaje.

1.7.1. Flujograma Construcción fuente de datos.

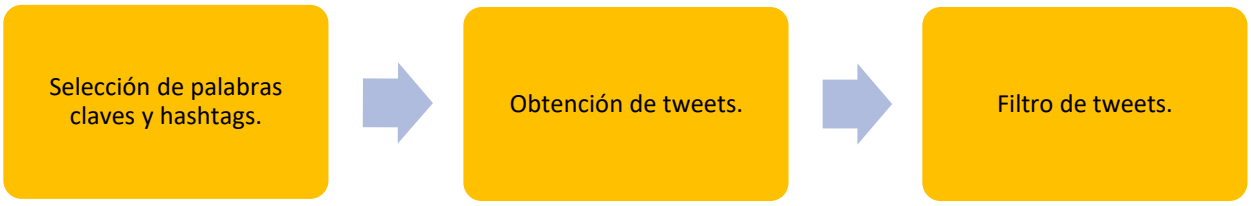


Gráfico 2 Flujograma construcción fuente de datos

Selección de palabras claves y hashtags: Para la construcción de la base de datos se hará una lista de las palabras claves usando diccionarios de términos psicólogos que tengan palabras que usualmente tengan relación con trastornos de depresión o ansiedad. Además de buscar también los hashtags que el estado del arte ha encontrado que describen las personas que tienen síntomas.

Obtención de tweets: Se solicitan los tweets referentes a un tema específico usando una herramienta de extracción de tweets de la alianza CAOBA la cual ha sido la fuente de datos para varios proyectos relacionados a las redes sociales en los últimos años.

Filtro de tweets: Para disminuir el ruido y los falsos positivos que pueden traer los tweets se hará un filtrado de los usuarios encontrados dependiendo la actividad en la plataforma y el uso de las palabras claves de acuerdo con un contexto especifico.

1.7.2. Flujograma exploración

En esta parte el estado del arte se proponen varios experimentos para hacer experimentos y probar cual es lo que mejor de resultados ya que depende mucho de la base de datos en tratada. Los pasos que estén de color naranja serán los pasos en donde se hará experimentos donde se abrirá en 2 caminos, haciéndolo y no haciéndolo

Gráfico 3 Flujograma exploración

Fichado (Tokenization): En esta parte se debe separar las frases en palabras. Usualmente este el primero proceso que se hace en un proceso de procesamiento de lenguaje natural para tratar las frases de un texto como si fueran valores específicos.

Etiquetado de POS: Esta parte del proceso etiqueta las palabras según el componente que le corresponde en un texto según el componente que significa en la frase. Este proceso por lo general va después de eliminar las palabras vacías, sin embargo, autores como Coppersmith [8] o Rozano [6] proponen hacer el experimento con las variables de salida y poner la eliminación de palabras vacías después del etiquetado.

Eliminar palabras vacías: La eliminación de palabras vacías o *stopwords* consiste en quitar todas las palabras que no le aporten valor semántico o valor agregado al texto. Comúnmente, las palabras que se eliminan son preposiciones, aunque dependiendo el tipo de tarea a ejecutar se pueden agregar o quitar palabras a esta lista. En este caso se usarán las palabras que ya vienen predefinidas en la librería de Spacy 1.8.2 en el apartado de español, el que constituye en alrededor de 90% preposiciones.

Análisis de dependencia: El análisis de dependencia se refiere a entender como está compuesta la oración a nivel semántico. Con este análisis se puede identificar cual es el objeto predicado en los tweets y si este tiene relación con los síntomas de ansiedad o de estrés o si solo es una palabra más del análisis. Este componente lo propone el autor Salas-Zárate [1] , ya que usualmente los tweets que tienen como objeto principal palabras relacionadas con medicamentos o algún tema polémico como lo es la religión o política tienden a parecerse a tener síntomas de depresión.

Lematización de las palabras: Luego de haber hecho un análisis semántico para encontrar cuales son las palabras importantes en los tweets, se recomienda hacer una lematización de las palabras para dejar las raíces y quitar los derivados de las palabras. Esto se hace en función de reducir la cantidad de palabras únicas que pueden aparecer cuando se hace la vectorización de campos.

Identificación de la polaridad de los textos: Uno de los pasos que usan se utiliza en la literatura, pues es la identificación de polaridad. Esto se refiere a detectar el sentimiento de una frase o un texto. Normalmente esto en si es una tarea analítica sin embargo autores como Xi [16]o Cambria [12] lo usan como un insumo para poder definir si las personas tienen síntomas de depresión o ansiedad.

Identificación de entidades: La identificación de entidades es por si sola una tarea analítica, sin embargo, hay autores como Salas-zarate[1] o Tai [19] que usan diccionarios especializados como lo es SENTICON o el LIWC para poder identificar emociones y conceptos psicológicos que pueden ayudar a entender con mayor profundidad el tema del texto.

Vectorización: Por último, el paso final antes de pasar a la modelación de datos es la parte de vectorización, en esta parte los autores como Rajput [5] utilizan técnicas ortodoxas como lo es el TF IDF o el Word Embedding basándose en algoritmos basados en corpus de Twitter.

1.7.3. Flujograma de Modelamiento

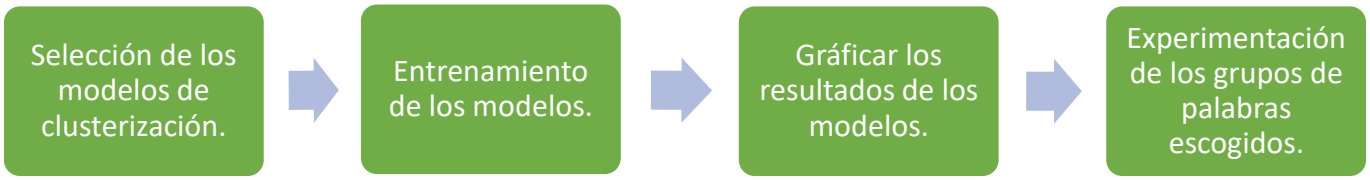


Gráfico 4 Flujograma modelamiento

Selección de modelos de Segmentación (Clústeres): Luego de haber desarrollado el preprocesamiento de los algoritmos y de la vectorización de las oraciones, lo siguiente seria escoger los modelos de clúster para poder segmentar a los tweets. En este caso se usarán dos tipos de algoritmos. Uno es el LDA (Latent Dirichlet Allocation) el cual es una red neuronal no supervisada y el K-means para poder agrupar estas palabras en grupos.

Entrenamiento de los modelos: Para el entrenamiento de los modelos se usará toda la base de datos ya que no hay necesidad de hacer una partición debido a que no se está tratando de hacer un algoritmo supervisado.

Graficar resultados de los modelos: Uno de los pasos más importantes en el desarrollo del proyecto y con el cual se medirá el desempeño de este son las visualizaciones de los resultados del modelo. En este caso se tendrá que mostrar de una forma no técnica los resultados de los modelos y la identificación de las palabras claves en los grupos. Más estadística descriptiva que permita identificar tendencias de la forma de hablar de las personas.

Experimentación de los grupos de palabras: Para poder identificar de una forma más visible los tópicos con los cuales se están formando los grupos se hará una validación de cuáles son las palabras que conforman los grupos y si tienen coherencia entre sí, autores como Mowery et al.[4] proponen identificar cuáles son los temas relacionados directamente con depresión y cuáles son los temas relacionados a síntomas de depresión. Esto con el fin de poder identificar cuáles son los usuarios que necesitan un mayor apoyo antes de que los síntomas de ansiedad y depresión se desarrollen más.

Para la experimentación de los modelos tratados se harán los cambios controlados usando la base de datos de entrenamiento de los modelos. Cada experimento que se realice tendrá diferentes tratamientos los cuales fueron nombrados en los procedimientos anteriores. Esto con el fin de encontrar cuál de los tratamientos y características lingüísticas (tales como la eliminación de palabras vacías, o la identificación de los sujetos de la oración) son relevantes cuando se tratan de encontrar tendencias en los textos.

Contraste contra grupo control: Para validar de que el algoritmo haya encontrado tendencias diferenciales de las personas que sufren depresión y no de la población general se tendrá que comprar con los resultados del grupo control y comprobar que no son los mismos. Autores como Gamon [2] afirman que el comparar con un grupo control permite identificar cuáles eran los comportamientos normales de las personas y poder así comprobar que los resultados de los modelos son significativos.

Para esto se tomará la base general y se realizará dos tipos de modelos uno solo con los textos que posiblemente no estén relacionados con depresión y el otro con los textos que muy posiblemente si estén relacionados a este.

2. EXPLORACIÓN DE DATOS

En el evento de exploración de los datos, se ha utilizado mecánicas en las que se busca filtrar aquella información relevante para un estudio de NLP, estas técnicas incluyen los pasos desde la construcción de los Datasets basados en la descarga directa mediante API de los textos de las publicaciones en redes sociales hasta su limpieza, garantizando un mínimo de coherencia en los datos que nos permita realizar procesos más avanzados como el Word Embedding necesario para nuestro ejercicio analítico, veamos el detalle de los pasos a continuación:

2.1. Limpieza de los datos y construcción de los artefactos de datos analíticos.

Para efectos prácticos del ensamblado de la base de datos, se debió garantizar el acceso a las redes sociales mediante API para extraer la información necesaria para el análisis de las publicaciones: texto, fecha y hora, identificadores de usuario y criterio de búsqueda basado en los diccionarios LIWC [1][20]. Posterior a este primer paso se realiza el tratamiento de limpieza por medio de ETLs removiendo aquellas publicaciones duplicadas o retuiteadas, así mismo, removiendo las URLs y menciones de usuario que no tienen ningún significado lógico para un estudio de análisis de sentimiento y de emociones [12][13].

Una vez normalizados los datos, se exportan a formato binario (parquet o pkl) con el propósito que sean fácilmente legibles por un modelo de ejecución autocontenido y el cual debe ser computacionalmente ligero ya que la mayor parte del tiempo consumido en procesamiento se emplea en los eventos de análisis de lenguaje natural, este procesamiento incluye también el cálculo de los modelos analíticos de los Word Embedding y el Clúster necesarios para identificar las tendencias o estresores depresivos objeto de este proyecto de prueba de concepto[4].

Una vez creados los archivos de modelo (clústeres) y los artefactos de datos para visualización, esta información se despliega mediante archivos binarios o parquet para que sean utilizados bien sea con fines estadísticos, visualización de datos o en su defecto otros usos específicos de analítica.

2.2. Arquitectura del proyecto

Para entender mejor como se estructura los artefactos de analítica es necesario visualizar el proceso global de cómo se concibió el modelo base del ciclo de vida de los datos, es decir desde que se consumen y se organizan hasta cuando se despliegan para su uso:

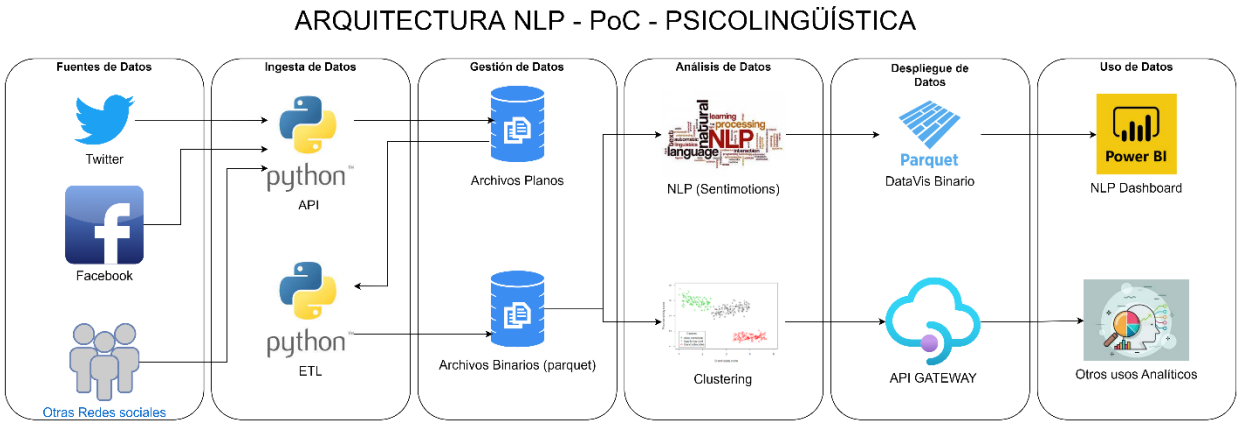


Gráfico 5 Arquitectura Base de PoC – Psicolingüística – Fuente Propia

Como ya explicamos en el numeral anterior (2.1), la arquitectura del proyecto tiene como fin la optimización de los recursos de cómputo, por ese motivo hay una notoria presencia de archivos binarios auto contenibles (parquet), esto ayuda a delegar los procesos de ejecución, dado que, como se memoriza dentro de esos archivos los estados de los objetos analíticos, no es necesario reentrenar modelos o desplegar los datos cada vez que se necesite su uso, como esto es un proceso delegado y de necesitarse un reentrenamiento sólo se ejecuta los pasos necesarios.

2.3. Exploración de la base de datos

2.3.1. Partes del lenguaje

Como parte de los procesos de análisis, dentro de la arquitectura se estimó el Procesamiento de Lenguaje Natural, el cual sirve para entender mejor el conjunto de información sobre el cual se realizó los análisis previos al modelado, este ejercicio consiste en descomponer en partes los textos (corpus) de las publicaciones que se trabajaron, como resultado se creó un tablero de control para el estudio descriptivo que muestra en proporciones de volumetría y componentes de tiempo el comportamiento de las Partes del Lenguaje (POS Tagging) sobre 620.139 publicaciones únicas en la red social de Twitter.

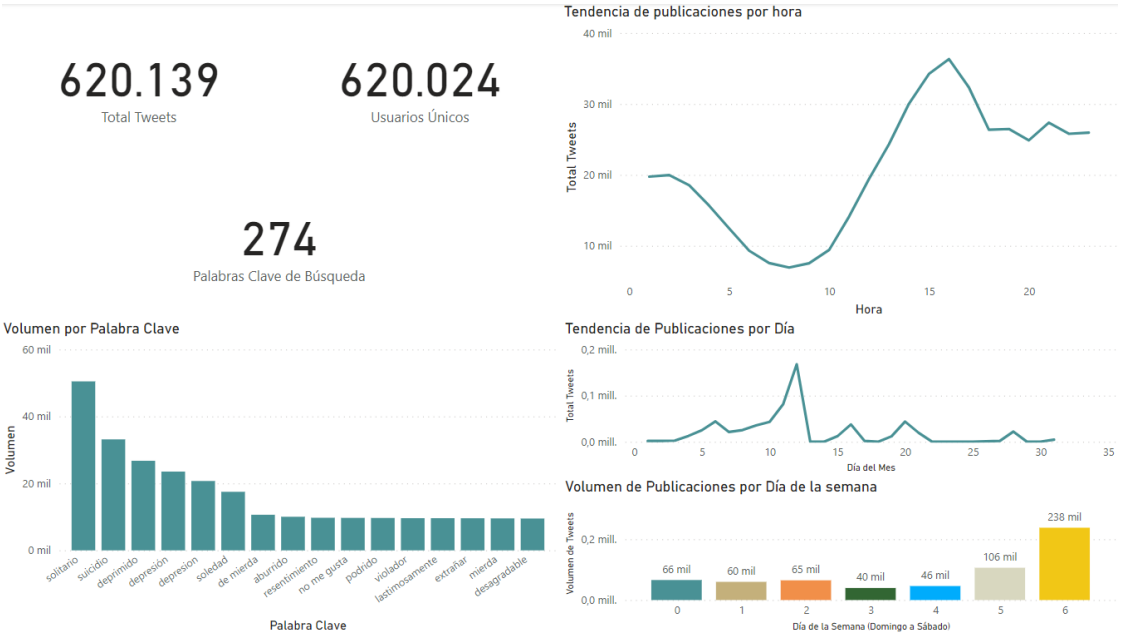


Gráfico 6 VOLUMETRÍA INICIAL – Publicaciones vs Tiempo – Fuente Propia

Posterior a haber identificado el universo de publicaciones a analizar, se procesó las partes del lenguaje y este nos muestra que el corpus tiene una predominancia entre 4 grandes grupos de tipos de palabra, SUSTANTIVOS, VERBOS, ADVERVIOS Y ADJETIVOS (CALIFICATIVOS), así mismo estos coinciden con los criterios de búsqueda, los cuales fueron definidos con antelación en el paso de ingesta de datos [13], en resumen, se obtuvo una lista reducida de 3.408 Fichas únicas (tokens o palabras), con el 44% de la base de datos perteneciendo a los SUSTANTIVOS y un global de 77.744 repeticiones totales de los tokens analizados.

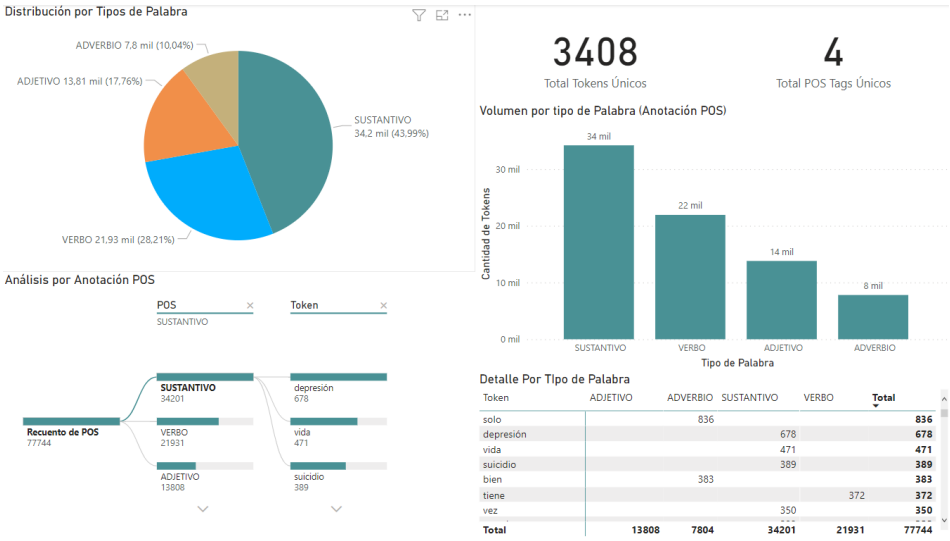


Gráfico 7 POS TAGGING –Partes del lenguaje (POS) y volumetría – Fuente Propia

2.3.2. Polaridad (Sentimiento)

En este apartado se muestra los resultados exploratorios sobre la polaridad hallada, la cual se confina a 3 posibilidades únicas: POSITIVA, NEGATIVA y NEUTRA; para este ejercicio fue necesario recurrir a una librería entrenada previamente basada en redes neuronales profundas y pre entrenadas denominadas BERT[22] que detecta de manera eficiente(98% de precisión en otras bases de datos relacionadas con tweets) tanto el sentimiento o polaridad además de la emoción impresa al momento de redactar el corpus, habiendo realizado el análisis descriptivo, las publicaciones presentan una tendencia hacia la negatividad (76% del volumen total de corpus), lo cual confirma de manera positiva lo expresado en trabajos similares a este [1][4][10][12], ahondando un poco más podemos ver que la hora de publicación es influyente, es decir, las publicaciones con sentimiento negativo en horas de la tarde y con decadencia notoria en horas de la mañana, esto es una hipótesis que previamente se había encontrado en la revisión de literatura en otro contexto.

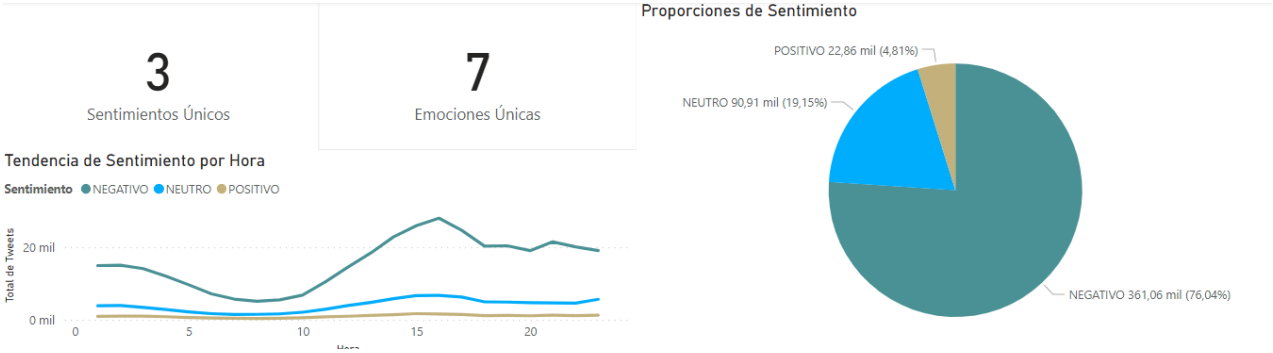


Gráfico 8 ANÁLISIS DE SENTIMIENTO –Proporcionalidad y tendencia de publicación– Fuente Propia

2.3.3. Emociones

Continuando con el análisis descriptivo de los datos (corpus), se valida que información es relativamente importante para el caso de estudio, de esto se desprende una representación de 7 emociones posibles detectadas en la data, la ira es la sensación humana más predominante con alrededor del 38% del volumen total de expresiones, sin embargo en el caso de otro tipo de emociones, el 37% corresponde a aquellas que no son parte del estudio y que podrían hacer parte de otros análisis en estudios posteriores, quizás con el fin de identificar posiblemente sarcasmo, ironía o elementos que también indicarían tendencia depresivas y que no son fácilmente identificables.

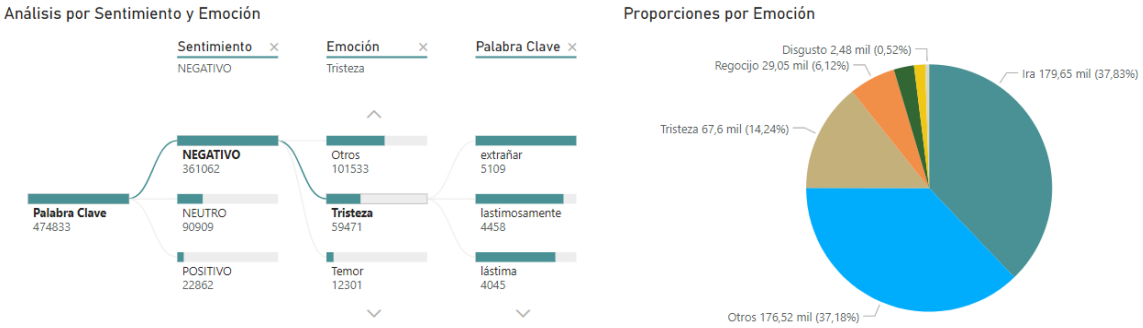


Gráfico 9 ANÁLISIS DE EMOCIONES – Proporciones – Fuente Propia

2.3.4. NLP Descriptivo

Inicialmente este apartado busca mostrar de una manera sencilla cual es la relación existente entre los tipos de palabra y los tokens (fichas), las cuales en este estudio descriptivo no han sido lematizadas o tratadas de alguna manera (exceptuando la limpieza inicial) con el fin de no perder la riqueza que puede aportar el análisis de los corpus[15][16].

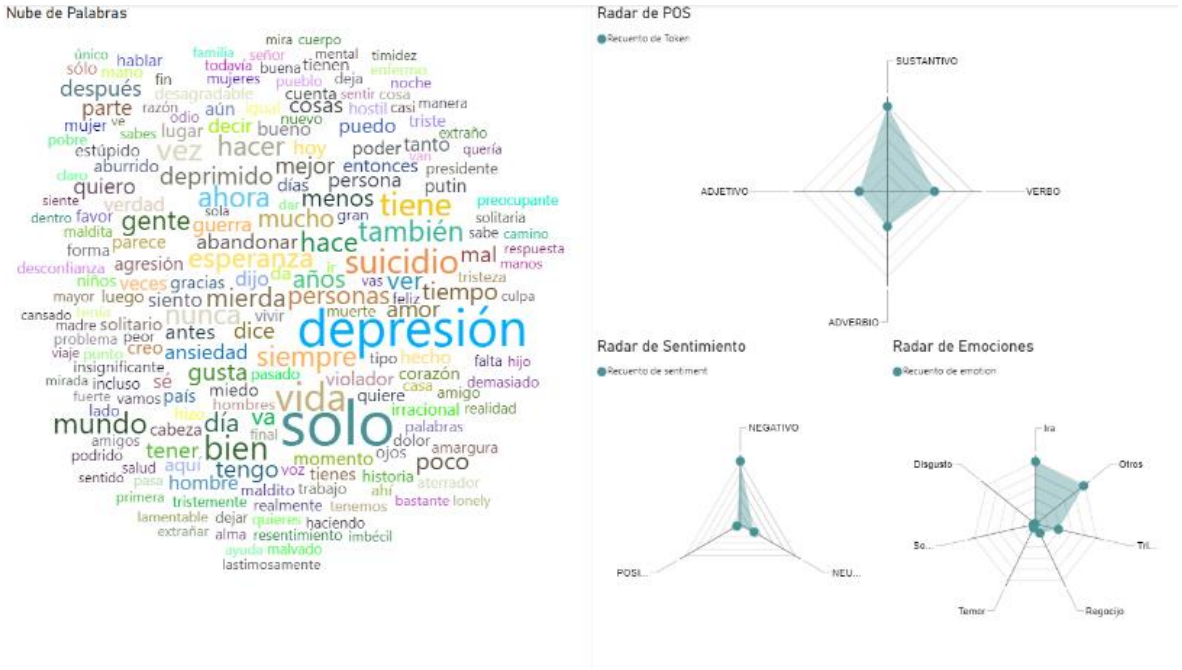


Gráfico 10 Nube de palabras relacionadas y elementos dimensionales – Fuente Propia

3. MODELAMIENTO Y EXPERIMENTACIÓN

3.1. Word Embeddings

Uno de los principales problemas que se tiene a la hora de hacer modelos de lenguaje natural es la representación que se hace de las oraciones en un espacio vectorial. Esto se debe a que cuando uno realiza una representación vectorial de las palabras no necesariamente se puede representar el contexto de un idioma y el significado de una oración se puede perder a volver las palabras números. Por ejemplo, como el autor Almeida [20] establece que uno de los casos más comunes de esto son las representaciones de bolsas de palabras o técnicas como TFIDF. Estas generan el espacio vectorial usando la existencia de palabras y la frecuencia de estas mismas en los textos encontrados. Lo cual generan matrices gigantes de valores las cuales están con valores cercanos a 0 y con una variedad muy compleja con la que se puede trabajar, además estos modelos no consideran las palabras que venían antes o después por lo cual no extraen un contexto de este. Por lo cual cuando se manejan textos informales como textos de redes sociales el contexto e intención de comunicación es importante. Por esta razón se necesitan modelos más sofisticados los cuales puedan describir mejor el contexto e intención de los textos en un espacio vectorial.

Los Word Embeddings es una metodología para poder generar estos usando más partes del contexto y que estos espacios vectoriales puedan extraer un poco más del contexto con el cual los textos fueron construidos. Esta técnica consta en usar redes neuronales no supervisadas que permitan tener un vector de ciertas dimensiones que represente a las palabras. Los principales hiper parámetros que tiene esta red neuronal son: las dimensiones del vector que representa la palabra, la ventana que va a usar para entender el contexto de las palabras, el umbral de palabras mínimas que se tomaran en cuenta para el modelo, las épocas que se usaran para entrenar la red neuronal y la arquitectura que se usara para representar las palabras. Las cuales son: CBOW (Continuous Bag Of Words) el cual trata de usar las palabras en la ventana de observación para predecir la palabra restante, sin embargo, está el modelo de SkipGram que consta en predecir las palabras de la venta de observación usando las palabras que actual. [20]

3.2. Latent Dirichlet Allocation

El algoritmo de Latent Dirichlet Allocation (LDA) es un modelo estadístico bayesiano que está constituido por tres niveles. Dentro de cada nivel cada documento del corpus se modela como una combinación sobre un conjunto de temas que esta embebido en el modelo. Luego cada tema se modela para determinar cuáles son las palabras que tienen mayor probabilidad de aparecen en un tema específico. El modelo representa los textos como una mezcla de estos temas debido a que están constituidos por varias palabras en sí. Finalmente hace la agrupación por estos temas. Debido a que el modelo está constituido por 3 capas. Donde la capa exterior representa los documentos y la capa interior representa las palabras que se tienen que se tienen que elegir para cada grupo, la tercera capa intermedia es la que regula el modelo la cual permite agrupar de diferentes maneras las palabras.

Los hiper parámetros que usualmente se usan para la experimentación de los temas son principalmente la beta que regula la distribución de palabras por tema y el alfa el cual regula la distribución de temas por documento. Estos parámetros se pueden incluir en el protocolo de experimentación, sin embargo, en este caso se dejará los predeterminados por la librería y se optará por hacer la experimentación sobre los datos de entrada.

3.3. Protocolo de experimentación

Para el protocolo experimental del proyecto se decidió seguir con la línea de proceso descrita en el proceso que esta descrito en la sección: *1.7.2. Flujograma de exploración*. En otras palabras, se realizó la experimentación variando las bases de datos de entrada en vez de centrarse en optimizar los hiper parámetros que se pueden sintonizar de los Word embeddings y el LDA. Esto con el fin de observar el impacto que diferentes componentes lingüísticos pueden impactar los modelos finales.

3.3.1. Medición de los modelos lingüísticos

Una de las dificultades que existen en la construcción de los Word embeddings es comprobar la calidad de este mismo. Por esta razón se generaron dos metodologías las cuales permiten medir de una forma más acertada la calidad de los Word embeddings, está la manera extrínseca e intrínseca.

La metodología extrínseca consiste en revisar la calidad de los vectores de palabras realizando tareas de clasificación con algoritmos simples para tener una métrica de clasificación que pueda representar la calidad de estos, el problema de esta metodología es que se necesita tener una base con etiquetas de acuerdo con un criterio lo cual en este caso no se tiene, por lo cual se descartan los métodos extrínsecos a la hora de medir la calidad de estos.

Por otro lado, está la metodología intrínseca autores como Gladkova[21] sugieren que este procedimiento se fundamenta principalmente en evaluar la similitud de las palabras y la relación que tiene estas con sus adyacentes o vecinos más cercanos, esto es una tarea que tiene que ser manual debido a que depende de la necesidad y objetivo con el cual fue construido el Word embedding. Por otro lado, también se pueden comparar los modelos creados entre sí, la calidad de los vectores la determinará un experto que en este caso será uno mismo.

3.4. Construcción del modelo

Para la construcción del modelo se decidió usar las técnicas de word2vec que provee la librería de Gensim, siguiendo el protocolo experimental que se mencionó anteriormente, para generar la agrupación de esto se usó un algoritmo de K means el cual permite generar grupos a partir de estos vectores. Para seleccionar el numero óptimo de grupos que se usaran se usaron dos metodologías, la primera es la técnica gráfica,

donde se mide la dispersión que los grupos pueden tener y la segunda es una técnica llamada GAP, la cual permite es otro método para visualizar el numero óptimo de clústeres.

Solo se realizó la agrupación con el Word embedding que tenía palabras similares a vista a partir de las palabras semillas que se usaron en la creación de la base de datos de entrenamiento. Además, para la evaluación de resultados se validó la dimensión de los Word embeddings usando dos tipos de dimensiones: 300 y 500 dimensiones por palabra, estos tamaños de vectores son debido a que los modelos ya existentes de Gensim y TensorFlow recomiendan estos tamaños para que no se genere sobreajuste en los pesos de estos modelos.

3.5. Evaluación de los resultados.

Los otros parámetros que contiene el Word embedding como la ventana de atribución o la selección de palabras claves se realizó una búsqueda de grilla para determinar los valores que tenían mejor desempeño usando como criterio subjetivo las analogías referentes a la palabra depresión y ansiedad ya que eran dos de las palabras semillas que se usaron para crear la base de datos consolidada. Para efectos prácticos solo se mostrará en la tabla 2. Los resultados de las mejores combinaciones encontradas para cada experimento y no se mostrará el detalle de todos los resultados ya que realizaron más de 5 experimentos por combinación de variables de texto. Los parámetros de oscilación en los cuales se experimentó fueron los mostrados en la tabla 3, donde el que menos vario fue la ventana de atribución ya que al ser textos tan cortos como los tweets las ventanas de atribución con gran cantidad de palabras no eran necesarias porque los textos lematizados tenían en promedio 8 palabras por tweet.

Parámetro	Espacio de búsqueda
Ventana de atribución	1,5,10
Tamaño del vector	300,500
Cantidad mínima valida por palabra	25,50,75,100
Épocas	50,100,250,500
Estilo de Word Embedding	CBOW o SkipGram

Tabla 2 Parámetros de búsqueda del Word Embedding.

3.5.1. Resultados

Variables de texto	Dimensiones del Word Embedding	Top 5 palabras similares relacionado a depresión
Sin Lematizar	300	perfil, imagen, sin, aterrador, no obstante, avergonzado
	500	risa, hundir, lastima, cara_sonriente, cara_corazones,

Lematizando, usando todas las palabras del POS	300	Castigar, pedir, nacional, ridículo, aceptar
	500	Dejar, motivo, interés, crisis motivo
Lematizando, usando solo las palabras claves de las oraciones (análisis de dependencia)	300	Nacer, política, iglesia, triste, pensar
	500	Triste, motivo, medico, iglesia, cara_tristeza
Lematizando, usando solo las palabras con POS de sustantivos, adjetivos, verbos y adverbios.	300	preocupar, medicina, consejo, ansiedad, complejo_inferioridad
	500	Negar, enfermo, detener, caro, necesidad

Tabla 3 Resultados del Word Embedding

Variables de texto	Coherencia máxima alcanzada
Sin Lematizar	0.05
Lematizando, usando todas las palabras del POS	0.15
Lematizando, usando solo las palabras claves de las oraciones (análisis de dependencia)	0.1
Lematizando, usando solo las palabras con POS de sustantivos, adjetivos, verbos y adverbios.	0.25

Tabla 4 Resultados del LDA

Como se puede observar en la tabla 3, los mejores resultados fueron del experimento del lematizado filtrando las palabras que tengan relación con los sustantivos, adjetivos verbos y adverbios. Esta combinación se resaltó debido a que cuando se realizó la parte descriptiva de la base de datos estos tipos de palabras resaltaron más en comparación a otro tipo palabras. Se puede observar que el embedding de 300 dimensiones por palabra es cierta medida más afinado que el de 500 dimensiones, con esto se puede concluir que al tener 200 dimensiones más la variabilidad entre palabras aumenta también, por lo cual el modelo está captando ruido con tantas dimensiones.

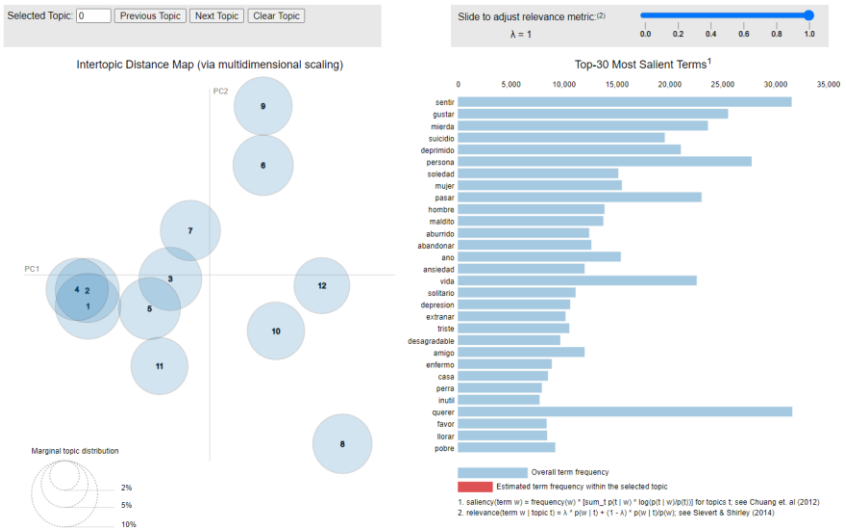
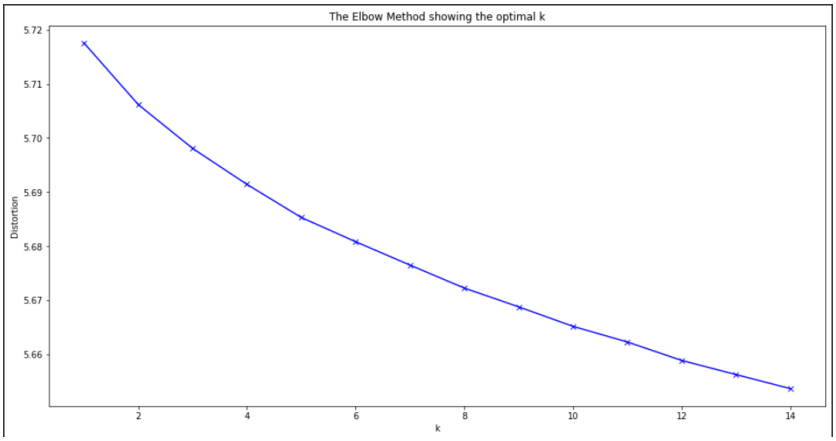


Gráfico 11 Resultados LDA

Asimismo, la mejor combinación presentada por el LDA es la que se ha lematizado las palabras y filtrado para que solo estén los sustantivos, verbos, adverbios y adjetivos. En la grafico 11 muestra la combinación de temas palabras usando la librería pylDAvis la cual permite visualizar el modelo de LDA. Como se puede apreciar el grupo 1,2,4 y el 3 y 5 están muy juntos entre si, por lo cual se pueden agrupar en un clúster único mientras que los demás si están distribuidos sobre las dos dimensiones del PCA.

Por otro lado, se puede apreciar que todos los modelos tienen palabras relacionadas con tristeza o depresión por lo cual se puede concluir la base de datos obtenida de los Twitter tiene tendencias a hablar negativamente lo cual valida que se obtuvo una base de datos pertinente y acorde con el objetivo por el cual fue creada. Asimismo, se puede apreciar que el experimento que no uso palabras sin lematizar las palabras está menos relacionado con la depresión y puede que les falte más contexto para poder relacionarse entre sí.



Número de clústeres	GAP
2	5.34
4	6.06
6	11.04
8	11.04
10	11.04
12	11.04

Gráfico 12 Gráfico del codo y distorsión de GAP

Luego, para la selección del número de clústeres usando el algoritmo K-Means se usaron dos métodos que miden la distancia que hay entre los puntos entre ellos relacionado con otros grupos. La medida del codo es un método grafico que se tiene que seleccionar el número de clústeres donde se vea que la pendiente cambia significativamente, no es recomendable este método porque es muy subjetivo por lo cual está el otro método que es el del GAP que prácticamente según el autor Tibhirani [22] esta medida sirve para medir la distorsión que se genera en las agrupaciones. Como se puede ver en el grafico 11, el número que sugieren los dos métodos es alrededor de 4 donde se genera una pendiente más empinada y donde la estadística del GAP es menor.

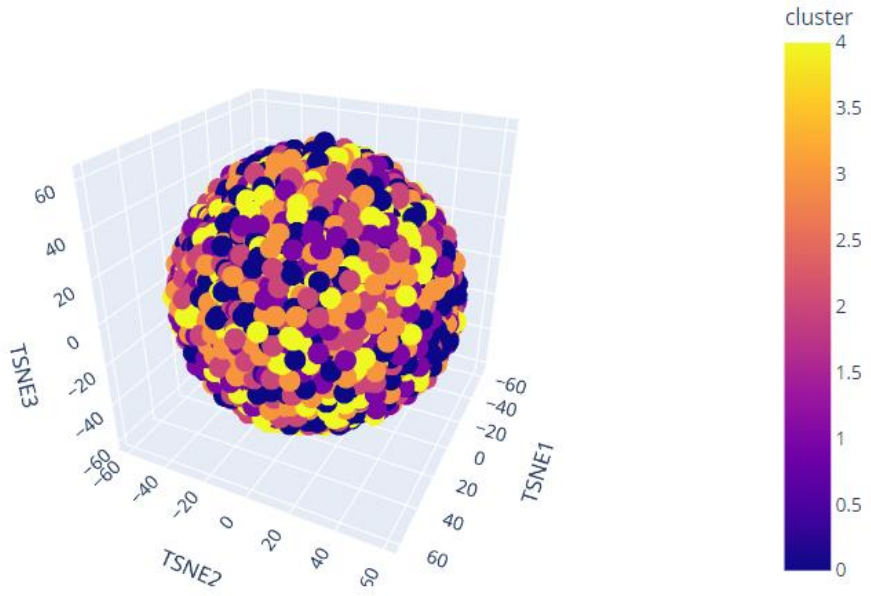


Gráfico 13 Visualización del Word Embedding usando TSNE

En la gráfica 12, se puede apreciar la representación en 3 dimensiones del mejor modelo seleccionado subjetivamente. Para poder visualizar el modelo de una forma general y poder ver como está distribuido los valores, se reduce las 300 dimensiones a 3 con una técnica llamada TSNE la cual principalmente sirve para la visualización de grandes volúmenes de datos con alta dimensionalidad. A partir de la imagen se puede concluir que las palabras tienen mucha variabilidad y que no son fáciles de separar usando un algoritmo de segmentación como lo es un K-means.

3.5.2. Conexión con el objetivo de negocio.

En cuanto a la relación que tiene con el objetivo de negocio de detectar síntomas de depresión, se puede ver que usando el modelo se puede identificar palabras que permitan explicar la forma de hablar de las personas que sufren estos síntomas. Además de esto usando el análisis exploratorio junto con el modelo se puede determinar que las personas con estos síntomas tienden un comportamiento determinado como lo es la hora de publicación y el mayor uso de palabras como lo son los verbos que autores como lo Salas-Zárate[1] lo había concluido en sus investigaciones.

3.6. Resultados tablero de control modelo

En este ejercicio de creación de artefactos de visualización se tuvo en cuenta aspectos en los que se realizaron pruebas experimentales, para ello, se creó varios tableros de control en los que se evidencia que los modelos basados en NLP como TFIDF pueden se prometedores, hay que tener siempre encuentra que la evaluación de los resultados en modelos no supervisados requiere de una validación de experto, la cual se analiza en el apartado 4.

Para los efectos prácticos, seguido presentamos los resultados de los experimentos de visualización para entrenamiento y predicción de clústeres.

3.6.1. Resultados del Tablero de Entrenamiento TFIDF

En el caso del modelo de representación de segmentos por TFIDF, se halló que en el mejor de los eventos un modelo con 4 Clústeres (85%), sin embargo, dentro de los experimentos se determinó que 3 Clústeres (76%) agrupan mejor la información, para ello se realizó con una muestra aleatoria de 60.000 tweets que constituye el 10% de la base de datos de publicaciones que se recabó en los estadios iniciales de este estudio

[0.6737989538197878, 2, 268.0247781662038], [0.7649123148761457, 3, 180.04777521343044], [0.8525829032486739, 4, 100.66000106346668]

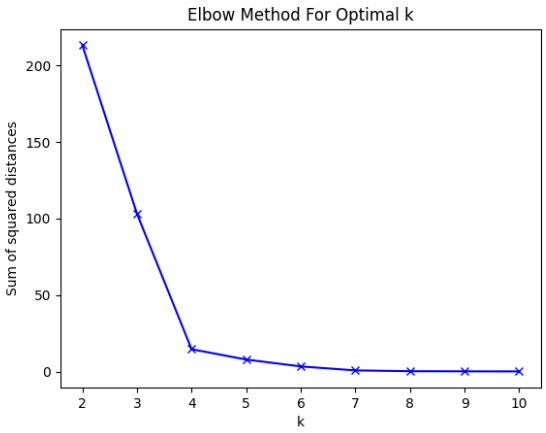


Gráfico 14 Método de la Silueta para segmentación o clústeres TFIDF

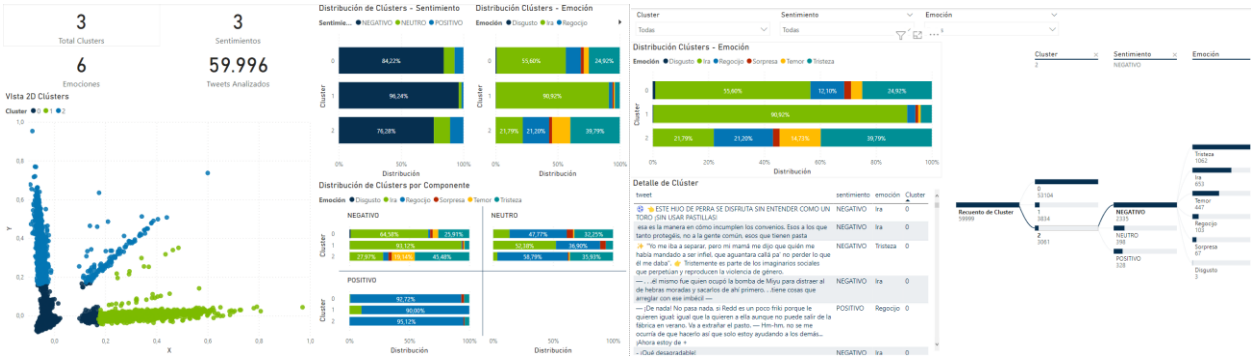


Gráfico 15 Dashboard clústeres TFIDF – Modo Entrenamiento

De acuerdo como se muestra en el gráfico 15, los eventos relacionados con la clasificación por clúster TFIDF con 3 grupos muestra una clara diferenciación entre los conjuntos de estudio, estos conjuntos se corresponden directamente y dentro del análisis es claro que el análisis por sentimiento no agrega valor, por otro lado, al diferenciar las emociones es claro que clústeres se diferencian.

a. Clúster 0: Representa un conjunto de publicaciones donde no hay una diferenciación de temas
b. Clúster 1: Representa un conjunto de publicaciones donde hay una notoria carga emocional tendiente a la IRA
c. Clúster 2: Representa un conjunto de publicaciones donde hay una correspondencia directa con los temas identificados en la bibliografía que mencionan tendencias depresivas, por tanto, este sería el clúster objeto de estudio. [1] [2] [4].

3.6.2. Resultados de Tablero de Entrenamiento LDA

Gráfico 16 Dashboard clústeres LDA – Modo Entrenamiento

Como se muestra en la gráfica 16, la diferenciación de los clústeres es menos notoria la diferencia, si se mira sólo al diferenciar gráficamente cada uno, puesto que no hay una distribución única, en este caso se deberá recurrir al modelo de validación por experto, el cual se contempla en el apartado 4.

3.6.3. Tablero de Predicción TFIDF a partir del modelo

En este ejercicio para el tablero de predicción para el tablero de TFIDF, se utilizó como referencia una base de 60.000 tweets no caracterizados previamente, aunque hacen parte de la base de datos construida a partir de la consulta con el api, estos no hicieron parte del modelo de entrenamiento, estos resultados se corresponden y complementan con el tablero de control de entrenamiento.

Gráfico 17 Dashboard clústeres TFIDF – Modo Predicción

Si se compara el gráfico 17 (predicción) y el 15 (entrenamiento), hay una similitud directa entre las distribuciones, tanto por volumetría como de la similitud de los textos en relación con los clústeres identificados por el modelo.

Cabe anotar que los Tableros aquí presentados, se alimentan en la práctica mediante la ejecución de la actualización de los data sets, estos conjuntos de datos son generados y actualizados mediante la API que se diseñó e implementó y la cual puede ser consumida mediante los tableros al momento de refrescar los datos.

4. VALIDACIÓN DE RESULTADOS.

Uno de los problemas de realizar algoritmos no supervisados es la validación de resultados. Los algoritmos de segmentación (clústeres) tienen métricas como la silueta o el GAP como se había nombrado en la sección 3.4.1 sin embargo cuando se tiene información de procesamiento de lenguaje natural las medidas de la silueta y de GAP se reducen significativamente por la dispersión que presentan los documentos cuando están vectorizados.

Existen métricas que modelo como LDA usan para medir la relación que existe entre grupos de palabras, estas medidas sirven para determinar si las palabras son coherentes y tienen relación entre sí, el problema con estas métricas es que dependen mucho de los datos de entrada y puede ser difícil compararlo con los valores de literatura debido a la naturaleza de los datos. En este caso la coherencia máxima obtenida por el modelo de LDA fue de 0.25, en comparación con otros artículos que hacen tareas similares en lenguas anglosajonas recomiendan que el valor sea superior a 0.7.

Por estas razones, para validar los resultados de los modelos obtenidos se necesitará recurrir a una validación cualitativa donde expertos en el tema de psicolingüística puedan comprobar la calidad del modelo. Para esto se realizó un instrumento de medición donde se mostraban los grupos de palabras y luego se hacían 3 preguntas, donde 2 de ellas median con una escala Likert y la tercera con opción abierta, las cuales eran:

1. De 1 a 5 donde 1 es nada relacionado y 5 muy relacionado, usted considera que palabras en el grupo están asociadas entre si
2. De 1 a 5 donde 1 es nada relacionado y 5 muy relacionado, usted considera que las palabras están asociadas con temas relacionados con trastornos de depresión o de ansiedad.
3. ¿Cómo llamaría al grupo de palabras?

Las preguntas se diseñaron basado se en las otras métricas y en los objetivos de la investigación en términos de detección de trastornos de depresión y ansiedad. La tercera pregunta era más para entender la percepción de los expertos referente a los grupos y de esta manera poder nombrar los grupos. Se considero usar los instrumentos como el TAM (Aceptación de modelo tecnológico), sin embargo, debido a que los usuarios que estaban contestando la encuesta no son los usuarios finales los a temas como percepción de usabilidad, percepción de facilidad de uso o intención de usabilidad no eran validos ya que la aplicación final no está diseñada para que ese tipo de profesionales la usen.

4.1. Evaluación de los resultados.

La encuesta fue llenada por el profesor de doctorado Javier Redondo, el cual es un experto en los temas relacionados con psicolingüística en el idioma español. Los psicólogos que se solicitaron por parte de la universidad Javeriana no respondieron la encuesta debido a que el proyecto propuesto no estaba alineado con la escuela de psicología que ellos seguían debido a que se estaba usando redes sociales y no se les estaba preguntando a la gente si sentían algún trastorno emocional.

Los resultados de las encuestas demostraron que los grupos de palabras dependen de contexto dado además si se mostraran los grupos de palabras solamente sin algún contexto, no se podría hacer una conexión directa con los trastornos de depresión o ansiedad que se tienen el objetivo de alcanzar. Esto se puede evidenciar ya que el profesor contexto todas las preguntas de escala Likert como 3/5 lo cual es el punto intermedio de la escala. Además, cuando se le preguntaron recomendaciones para la investigación y para el tablero de control, él solicito que aparte de mostrar el tablero de control presentado en el grafico 10, también agregar una parte que pueda visualizar fragmentos de la base de datos que representen a esos grupos, esto con el fin de que la persona pueda entender de una forma visual a que se está refiriendo.

5. CONCLUSIONES Y RECOMENDACIONES

a) Al realizar la exploración de los datos se pueden comprobar las hipótesis que se habían presentado en la revisión de literatura referente al uso de las redes sociales como Twitter. Se pudo comprobar que las personas con indicios de depresión y ansiedad tienden a publicar en horarios de la tarde o noche, luego de las 5 pm donde tienden a tener mayor tiempo libre para publicar. Además de esto, se también se validó la teoría de que las personas con estos aspectos tratan de hablar usando más verbos y adjetivos calificativos en comparación con otros tipos de usuarios. Como recomendación para futuros trabajos se recomienda poder hacer una mayor parte de experimentos e incluir variables de temporalidad en los modelos a realizar en el futuro.

b) La calidad de información y los textos de entrada en los modelos son primordiales para poder tener buenos resultados, incluso más que la búsqueda de hiper parámetros que permitan disminuir las métricas de error. Esto se pudo evidenciar en el protocolo experimental donde las principales mejoras que se presentaban en los modelos se debían en su mayoría por cambiar los datos de entrada en los modelos. Los mejores resultados de los modelos fueron al eliminar palabras que no fueran sustantivos, verbos, adjetivos y adverbios. Lo cual disminuyo mucho la variedad que podían presentar los datos a la hora de separarse en los grupos.

c) Los grupos encontrados usando el modelo de LDA son mucho más coherentes y significativos que los encontrados usando los Word Embeddings, debido a que éstos tienen mayor variedad al tener que separar en grupos los vectores de 500 dimensiones por texto en grupos. Además, los algoritmos de LDA al ser un modelo probabilístico bayesiano pueden tener mayor peso las palabras que este cerca a ellos y puede generar una agrupación más precisa de los factores latentes que puede existir en los textos.

d) A la hora de exposición de la herramienta frente a profesionales de psicología hay que resaltar que el modelo es una herramienta para diagnóstico de redes sociales mas no determina que las personas sufren depresión o ansiedad, esto se tiene que recalcar debido a que muchos profesionales pueden tener escuelas e ideales que no validen estas herramientas.

e) Es menester continuar con el estudio mediante el uso de algoritmos de NLP en este caso TFIDF, como se puede apreciar en la validación de los resultados y en el análisis descriptivo y predictivo se evidencia la posibilidad de realizar predicciones en tiempo real en la medida que se tome la data que coincida con los criterios de búsqueda especificados para los estudios psicolingüísticos.

6. SIGUIENTES PASOS

Como siguientes pasos para el modelo se recomienda consultar la aplicabilidad del modelo con profesionales calificados relacionados con la psicolingüística que puedan comprobar que los grupos de palabras si estén relacionados entre sí y con aspectos de depresión o ansiedad. Esto con el fin de validar con una muestra de mayor cantidad de personas que comprueben que efectivamente los grupos si estén cumpliendo con las expectativas dadas y que no esté sesgado por la opinión de algunos profesionales.

Por otro lado, se espera que el grupo de estudiantes de pregrado de la universidad javeriana puedan dar continuidad al proyecto y puedan continuar desarrollando la prueba de concepto de manera productiva, para que los modelos lingüísticos encontrados puedan ser de utilidad a las personas que necesiten determinar aspectos de este estilo, y que en un futuro pueda volverse un servicio brindado por la alianza CAOBA.

En el caso específico de artefactos de datos, se deja suficiente material implementado en la forma de tableros de control, código de fuente ejecutable y modelos de datos los cuales están disponibles para continuar con futuros análisis o experimentos, estos futuros. estudios pueden tomar esto como referencia de partida, este proyecto constituyó una prueba de concepto y los resultados obtenidos fueron suficientemente consistentes con el objetivo de identificar posibles tendencias relacionadas con la depresión en las publicaciones en redes sociales.

7. BIBLIOGRAFÍA

[1] M. P. Salas-Zárate, R. Valencia-García, and N. Aussenac-gilles, "A study on LIWC categories for opinion mining in Spanish reviews," 2014, doi: 10.1177/0165551510000000.

[2] M. Gamon and S. Counts, "Predicting Depression via Social Media Predicting Depression via Social Media," no. September, 2021.

[3] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "ScienceDirect Detecting depression and mental illness on social media : an integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017, doi: 10.1016/j.cobeha.2017.07.005.

[4] D. Mowery *et al.*, "Understanding Depressive Symptoms and Psychosocial Stressors on Twitter : A Corpus-BaMowery, D., Smith, H., Cheney, T., Stoddard, G., Coppersmith, G., Bryan, C., & Conway, M. (n.d.). Understanding Depressive Symptoms and Psychosocial Stressors on Twitter ," vol. 19, no. 2, pp. 1–17, doi: 10.2196/jmir.6895.

[5] A. E. Rajput and S. M. Ahmed, "Making a Case for Social Media Corpus for Detecting Depression," pp. 1–9, 2018.

[6] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, and F. Sanz, "Detecting Signs of Depression in Tweets in Spanish : Behavioral and Linguistic Analysis Corresponding Author :," vol. 21, 2019, doi: 10.2196/14199.

[7] D. L. Mowery, W. Way, C. Bryan, M. Conway, and W. Way, "Towards Developing an Annotation Scheme for Depressive Disorder Symptoms : A Preliminary Study using Twitter Data," pp. 89–98, 2015.

[8] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, "From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses," *2nd Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, CLPsych 2015 - Proceedings of the Workshop*, no. January 2015, pp. 1–10, 2015, doi: 10.3115/v1/w15-1201.

[9] V. v. Ramalingam, A. Pandian, A. Jaiswal, and N. Bhatia, "Emotion detection from text," *Journal of Physics: Conference Series*, vol. 1000, no. 1, 2018, doi: 10.1088/1742-6596/1000/1/012027.

[10] M. Birjali, A. Beni-Hssane, and M. Erritali, "A method proposed for estimating depressed feeling tendencies of social media users utilizing their data," *Advances in Intelligent Systems and Computing*, vol. 552, no. His, pp. 413–420, 2017, doi: 10.1007/978-3-319-52941-7_41.

[11] E. C. C. Kao, C. C. Liu, T. H. Yang, C. T. Hsieh, and V. W. Soo, "Towards text-based emotion detection: A survey and possible improvements," *Proceedings - 2009 International Conference on Information Management and Engineering, ICIME 2009*, pp. 70–74, 2009, doi: 10.1109/ICIME.2009.113.

[12] E. Cambria, "Affective Computing and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016, doi: 10.1109/MIS.2016.31.

[13] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” *Engineering Reports*, vol. 2, no. 7, pp. 1–24, 2020, doi: 10.1002/eng2.12189.

[14] A. Biradar and S. G. Totad, *Detecting Depression in Social Media Posts Using Machine Learning*, vol. 1037. Springer Singapore, 2019. doi: 10.1007/978-981-13-9187-3_64.

[15] F. T. Giuntini, M. T. Cazzolato, M. de J. D. dos Reis, A. T. Campbell, A. J. M. Traina, and J. Ueyama, “A review on recognizing depression in social networks: challenges and opportunities,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 4713–4729, 2020, doi: 10.1007/s12652-020-01726-4.

[16] X. Liu, “Full-Text Citation Analysis : A New Method to Enhance,” *Journal of the American Society for Information Science and Technology*, vol. 64, no. July, pp. 1852–1863, 2013, doi: 10.1002/asi.

[17] J. C. Eichstaedt *et al.*, “Facebook language predicts depression in medical records,” *Proc Natl Acad Sci U S A*, vol. 115, no. 44, pp. 11203–11208, 2018, doi: 10.1073/pnas.1802331115.

[18] M. A. Moreno *et al.*, “Feeling bad on facebook: Depression disclosures by college students on a social networking site,” *Depression and Anxiety*, vol. 28, no. 6, pp. 447–455, 2011, doi: 10.1002/da.20805.

[19] C. H. Tai, Z. H. Tan, Y. S. Lin, and Y. S. Chang, “Mental Disorder Detection and Measurement Using Latent Dirichlet Allocation and SentiWordNet,” *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, pp. 1215–1220, 2016, doi: 10.1109/SMC.2015.217.

[20] F. Almeida and G. Xex, “Word Embeddings: A Survey,” no. 1991, 2015.

[21] A. Gladkova and A. Drozd, “Intrinsic Evaluations of Word Embeddings: ¿What Can We Do Better?,” pp. 36–42, 2016, doi: 10.18653/v1/w16-2507.

[22] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of data clusters via the gap statistic,” *Journal of the Royal Statistical Society: Series B*, vol. 63, no. Part 2. pp. 411–423, 2001.