Revista Venezolana de Computación
ISSN: 2244-7040
http://www.svc.net.ve/revecom
Vol. 9, No. 1, pp. 10-19, Junio 2025
Selección de los Mejores Artículos de CoNCISa 2024
https://doi.org/10.63870/revecom.2025.09.01.002

Análisis de Sentimientos Migratorios en Publicaciones de Venezolanos en Medios Sociales

Livia Borjas, Mauricio Morales, Miguel Zamora
liborjas@ucab.edu.ve, mauricio1999morales@gmail.com, miguelzamora.23.97@gmail.com

Escuela Ingeniería de Informática, Universidad Católica Andrés Bello, Ciudad Guayana, Venezuela

Resumen: El presente trabajo tuvo como propósito realizar análisis de sentimientos y opinión migratoria en publicaciones de venezolanos en medios sociales, con la finalidad de apoyar el proceso de toma de decisiones en cuanto al fenómeno migratorio que sufre la nación venezolana. En primera instancia se llevó a cabo una revisión documental en aproximadamente ocho (08) artículos relacionados a los modelos de clasificación de textos, lo que permitió reconocer a los algoritmos de clasificación y el aprendizaje automatizado como los métodos más utilizados. En la etapa de desarrollo se definieron dos escalas de categorización, una que reflejó las opiniones de los usuarios y la otra los sentimientos asociados a su posición. Finalmente se procedió al proceso de construcción de los modelos de clasificación mediante la implementación de una aplicación móvil, donde los resultados obtenidos del análisis mostraron tendencias hacia la tristeza como sentimiento predominante y a la postura a favor de la migración, en cuanto a la opinión de los individuos.

Palabras Clave: Análisis de Sentimientos; Minería de Opinión; Clasificación de Textos; Algoritmos de Aprendizaje Automatizado; Fenómeno Migratorio.

Abstract: The purpose of this work was to conduct sentiment analysis and migration opinion analysis in posts by Venezuelans on social media, with the aim of supporting the decision-making process regarding the migratory phenomenon affecting the Venezuelan nation. Initially, a documentary review was carried out on approximately eight (08) articles related to text classification models, which allowed for the recognition of classification algorithms and machine learning as the most commonly used methods. In the development stage, two categorization scales were defined: one reflecting users' opinions and the other representing the feelings associated with their positions. Finally, the process of building classification models was carried out through the implementation of a mobile application, where the results obtained from the analysis showed trends towards sadness as the predominant sentiment and a favorable stance towards migration, regarding individuals' opinions.

Keywords: Sentiment Analysis; Opinion Mining; Text Classification; Machine Learning Algorithms; Migratory Phenomenon.

I. Introducción

El tema del análisis de sentimientos y minería de opinión ha tomado mucho interés en la actualidad. Muchas investigaciones han demostrado el creciente interés sobre el estudio de los fenómenos migratorios aplicando minería de datos con inteligencia artificial sobre publicaciones en redes sociales [1]. La presente contribución gira en torno al tema de analizar la opinión y sentimiento de los venezolanos expresados en sus redes sociales. El documento expone los primeros resultados de este estudio, organizando el aporte de la siguiente manera: la Sección II tiene el planteamiento del problema, la Sección III describe los antecedentes de la investigación, la Sección IV presenta una propuesta de metodología hibrida para aplicar en el caso en estudio, la Sección V describe los resultados, y finalmente la Sección VI presenta las conclusiones y recomendaciones.

II. Planteamiento del Problema

El continuo deterioro de la economía venezolana a lo largo de los años, ha generado escasez severa de alimentos y medicamentos básicos que, sumada a la crisis política e institucional, resultaron en un movido flujo migratorio de venezolanos, reseñándose la diáspora de mayor volumen en la historia venezolana. Según Arias, Arias, Morffe, Martínez y Carreño [2], en su artículo "Informe sobre la movilidad humana venezolana II" mencionan que la migración internacional venezolana es uno de los fenómenos con mayor impacto en Latinoamérica, donde no sólo se ve afectada la población venezolana, sino también los países receptores, ya que, si la cifra de migración mantiene el ritmo actual, creará cada vez más problemas presupuestarios, decaimiento de servicios básicos y la atención sanitaria en dichos países.

En el artículo "Refugiados y migrantes de Venezuela" de la Plataforma de Coordinación Interagencial para Refugiados y Migrantes de Venezuela en su reporte de septiembre del 2021 [3] afirma que, la cifra de venezolanos que han emigrado es de aproximadamente 5,66 millones de individuos, que representan alrededor de un 20% de la población total venezolana.

Para llevar a cabo estudios sobre estos flujos migratorios, el Observatorio Proyecto Migración Venezuela, en su artículo "Percepción en redes sociales sobre la migración venezolana" [4] plantea que el análisis de las redes sociales es una de las herramientas más rápidas y eficaces para obtener una visión general de la situación, ya que proporciona información que posteriormente facilita a los encargados del proceso de toma de decisiones, la materialización de sus objetivos.

Según Hütt [5] en su artículo "Las redes sociales: Una nueva herramienta de difusión" el impacto de las redes sociales tiene la capacidad de cambiar la percepción que los usuarios tienen sobre un tema en particular, y en una situación como la crisis migratoria de Venezuela, las ideas y creencias formadas por la opinión pública pueden entorpecer la integración de los migrantes, por ello es importante aclarar las posturas actuales que se tienen sobre este fenómeno.

Para realizar el análisis de mensajes en redes sociales, una de las técnicas más utilizadas es el Análisis de Sentimientos también conocida como Minería de Opinión, que Liu [6] en su artículo *"Sentiment analysis and opinion mining"* define como "el campo de estudio que analiza las opiniones de las personas, sentimientos, evaluaciones, apreciaciones, actitudes y emociones sobre entidades tales como productos, servicios, organizaciones, individuos, cuestiones, eventos, tópicos y sus atributos".

La relevancia del análisis de sentimientos radica en su capacidad para utilizar técnicas de procesamiento del lenguaje natural, sobre grandes volúmenes de información de forma automatizada, lo que facilita el reconocimiento de tendencias y patrones para la categorización de las opiniones y los sentimientos de la población.

A. Propuesta

Dada la relevancia del tema migratorio venezolano unido con el aporte que puede sumar el análisis de sentimientos en redes sociales, el objetivo del presente trabajo se enfoca en realizar un proceso de análisis de sentimientos y minería de opinión respecto al tema de migración sobre publicaciones de venezolanos en medios sociales.

Para alcanzar este propósito se aplicaron los siguientes pasos: examinar algunos antecedentes relacionados con el análisis de sentimientos y minería de opinión en redes sociales a fin de establecer la manera efectiva de aplicar este procedimiento en el tema migratorio, determinar los criterios sobre los cuales se llevará a cabo dicho análisis, diseñar modelos descriptivos o predictivos de sentimientos y de opinión que se ajusten a los criterios establecidos, para finalmente implementar dichos modelos mediante una aplicación móvil para su publicación, explotación y uso.

III. ANTECEDENTES

El avance de las tecnologías de big data y minería de textos, en la última década, han demostrado los aportes que pueden obtenerse mediante la aplicación de análisis de sentimientos. A continuación, se describen algunas investigaciones que proporcionan base para el desarrollo del trabajo a realizar.

Al momento de realizar el análisis de sentimientos u opiniones, uno de los mayores retos es el de la comprensión de los tonos lingüísticos en un mensaje, en ese apartado Salaz [7], en su tesis doctoral titulada "Detección de patrones psicolingüísticos para el análisis de lenguaje subjetivo en español", propuso un método para la detección de patrones psicolingüísticos en el análisis de sentimientos y la detección de la sátira en español.

Este método permite, a través de un enfoque automatizado supervisado, clasificar textos como positivo, negativo, neutro, muy positivo o muy negativo y como satíricos y no satíricos. Esta investigación demuestra las aplicaciones de los enfoques supervisados para reconocer aspectos particulares, en este caso el tono lingüístico en los mensajes.

Por su parte, Cestari [8] realizó una tesis de pregrado titulada "Propuesta para automatizar la asociación de emociones a textos en español", la cual consistía en automatizar la asociación de frases en español a emociones, mediante el análisis de sentimientos. Tomando en cuenta el modelo de emociones de Ekman, se construyó un prototipo capaz de asociar las frases a un conjunto de seis (6) emociones, basado en algoritmos de aprendizaje de máquina, aprendizaje profundo y redes neuronales.

Este estudio demuestra la aplicación del análisis de sentimientos y algoritmos de aprendizaje de máquina a textos en español, así como la flexibilidad que ofrecen estos modelos ya que el número de categorías pueden varían según el propósito para el cual esté dirigido.

Arango y Osorio [9] realizaron un artículo de investigación titulado "Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning", donde se planteó analizar los sentimientos subyacentes de los comentarios de Twitter relacionados con el aislamiento, identificando los temas y palabras frecuentemente utilizados en el contexto, a través de un algoritmo de machine learning. Se obtuvo como resultado la identificación del miedo como el sentimiento predominante durante todo el periodo de confinamiento.

Este artículo de investigación demuestra que el uso del análisis de sentimientos puede ir más allá que solo clasificar los mensajes en grupos particulares, sino que también permite identificar los temas y palabras características de cada uno de los grupos resultantes.

Abordando más afondo el fenómeno migratorio, se llevó a cabo una investigación sobre crisis migratoria basada en análisis de sentimientos [10] donde se plantea que la dinámica migratoria se ha producido a razón de crisis económicas, políticas y humanitarias. Por medio de un acercamiento desde el análisis de sentimientos se obtuvieron que las preocupaciones radicaban mayormente en las subcategorías de los derechos humanos, seguridad y desempleo. En base a estos resultados se puede validar que el análisis de sentimientos permite obtener opiniones públicas sobre un tema o situación

determinada, para así tomar las medidas preventivas o correctivas correspondientes.

Adicionalmente, al aplicar la técnica de revisión documental, sobre aproximadamente más de ocho (08) referencias, [5][6][8][9][10][11][12][16], cuyas fechas de publicación estén vigentes con una década de diferencia al 2021 (fecha que se realizó la primera fase del estudio), en los cuales se aplica las técnicas de aprendizaje automático para realizar análisis de sentimientos. En este análisis documental se pudo apreciar que todos los trabajos abordaron un procedimiento metodológico basado en el entrenamiento de clasificadores de textos, aplicando las siguientes etapas:

- Etapa de recolección de datos
- Pre procesamiento de textos
- Entrenamiento de los modelos
- Despliegue de los resultados

Para poder llevar a cabo la fase de recolección de datos, en su mayoría se basaron en mensajes de la red social Twitter, y para su extracción se utilizaron, el software estadístico R y en otras ocasiones lo hicieron a través de la API de Twitter con el lenguaje de Python.

De la misma forma para llevar a cabo la construcción de los clasificadores de textos se seleccionaron diferentes algoritmos basados en aprendizaje automático supervisado, entre los más utilizados se encontraron los siguientes algoritmos:

- Naive Bayes
- Máquina de Soporte Vectorial
- Regresión Logística
- Árboles de decisión
- K-vecinos

Para evaluar el desempeño de los modelos construidos, se destacó el uso de las siguientes métricas de rendimiento:

- Precisión
- Recall
- f1-core

En resumen, para llevar a cabo un estudio de análisis de sentimiento y opinión es importante tener presente:

- El uso de la plataforma de Twitter u otra red social como fuente de información, si es adecuado para el tema a indagar
- El uso del lenguaje Python o R para la extracción de textos de las redes sociales
- Las diferentes etapas que constituyen la construcción, entrenamiento y evaluación de modelos de clasificación de textos

El enfoque más utilizado es el de aprendizaje automático, más en concreto los basados en algoritmos de clasificación tradicionales.

IV. METODOLOGÍA

El análisis de sentimientos y minería de opinión son temas que vienen en ascenso debido al impacto que pueden tener a nivel comercial o social. Debido a esto se decidió abordar la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining o Proceso Estándar Para la Extracción de Datos en Todos los Sectores) [14][15], pero teniendo en consideración que el presente estudio tratará un tipo particular de minería de datos complejos, específicamente mensajes cortos de textos, se aplicará una propuesta de adaptación de autoría propia de la metodología CRISP-DM al caso de datos complejos de textos.

Esta adaptación consistirá de dos etapas generales: una etapa de pre-procesamiento y una etapa de descubrimiento. La primera etapa, está formada por la comprensión del problema, la creación del corpus y la transformación de los textos a algún tipo de representación estructurada o semiestructurada que facilite su posterior procesamiento, mientras que la segunda etapa, está constituida por el proceso de modelado para la clasificación de mensajes, la evaluación de rendimiento y finalmente el despliegue de los conocimientos obtenidos. En la Figura 1 se pueden apreciar las etapas para la metodología CRISP-DM adaptadas al proceso de minería de textos.

A continuación, se describen las etapas que se llevan a cabo, siguiendo la metodología CRISP-DM adaptada al proceso de análisis de sentimientos, como se puede observar en la Figura 1.

A. Comprensión del Negocio

En esta primera etapa, se lleva a cabo la etapa de revisión documental relacionada con el análisis de sentimientos, las técnicas de procesado de lenguaje y la construcción de modelos de clasificación, para fortalecer los conceptos fundamentales presentes en el análisis.

B. Definición de los Criterios para el Análisis de Sentimientos

En esta etapa se establecen los parámetros que se tendrán en cuenta para aplicar el análisis de sentimientos, así como las premisas que deben cumplir cada uno de ellos para la selección de sus elementos.

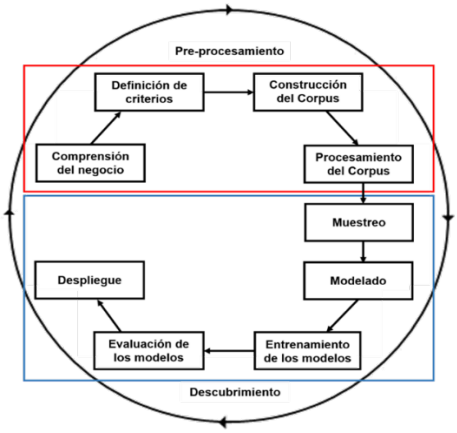

Figura 1: Adaptación de la Metodología CRISP-DM al Proceso de Análisis de Sentimientos

C. Construcción del Corpus

En esta etapa se recopilan la máxima cantidad de textos posibles que cumplan con la serie de criterios establecidos

previamente y se asocia cada uno de ellos con alguna de las categorías de la escala de sentimientos.

D. Procesamiento del Corpus

En esta instancia se llevan a cabo un conjunto de técnicas para simplificar los mensajes a representaciones estructuradas, con la intención de mejorar la interpretación del análisis automatizado y así obtener mayor precisión al momento de entrenar y evaluar los modelos de clasificación.

E. Muestreo del Corpus

En esta etapa se define la proporción de los mensajes del corpus que formarán parte de los datos de entrenamiento, utilizados en la construcción del modelo, y los datos de prueba, utilizados en la etapa de evaluación del modelo.

F. Modelado y Entrenamiento de los Clasificadores

En esta etapa se lleva a cabo el desarrollo de los modelos, a partir de los algoritmos de clasificación seleccionados y la data de entrenamiento establecida.

G. Evaluación de los Modelos de Clasificación

En esta etapa se mide el desempeño de los modelos de clasificación haciendo uso de los datos de prueba. Este proceso se logra mediante el estudio de las métricas de rendimiento que evalúan el índice de éxito del modelo para diferentes situaciones.

H. Despliegue de los Modelos de Clasificación

Finalmente, según los resultados obtenidos, los modelos pueden ser desplegados para su visualización o utilizados como base para otros trabajos con relación al área.

V. Resultados

Los resultados de los objetivos más resaltantes, después de aplicar la metodología, se describen a continuación:

A. Definición de Criterios para el Análisis de Sentimientos

Entre los criterios a tomar en cuenta para el proceso de recolección de textos, que constituye la primera fase para llevar a cabo un análisis de sentimientos, se encuentran:

1) Selección de la Plataforma. La plataforma seleccionada para la extracción de los textos de estudio fue Twitter debido su facilidad de acceso a la API para el momento del estudio, año 2021, los formatos al realizar una extracción de datos y como indica Mathews et al. (2016) citado por Saura, Palos y Ríos [16], la red social Twitter, además de generar gran cantidad de datos y mensajes, permiten vincularlos a un evento o característica y posteriormente segmentar audiencias al recopilar mensajes en torno a un hashtag, las cuales son etiquetas ubicadas dentro de los textos para asociarlos a algún tema en particular.

2) Selección de la Población Objetivo. En esta etapa se seleccionaron los perfiles de twitter, de los cuales se extraerán los mensajes para el análisis. Para definir a la población se utilizaron los criterios mencionados por Arias, Villasis y Miranda [1] donde establecen que para definir una población es necesario tener en cuenta las características de homogeneidad, la temporalidad y los límites espaciales. Teniendo en cuenta estos factores, para la selección de la población objetivo se definieron las siguientes premisas:

- Homogeneidad. Los perfiles deben pertenecer a individuos mayores a 18 años de edad. Los perfiles deben tener alguna participación sobre la migración
- Temporalidad. Los perfiles deben haber tenido participación en el año 2021, año donde se realiza la recolección de los datos a través del API
- Definir los límites espaciales. Los perfiles deben pertenecer a individuos venezolanos

Para llevar a cabo esta tarea se utilizó la herramienta web FollowerSearch, la cual es una herramienta de análisis de twitter, especializada en la búsqueda de perfiles al suministrar ciertos parámetros. Para este estudio, el criterio de selección aleatoria, de perfiles de venezolanos y mayores de edad. Haciendo uso de ella se obtuvo un total de 128 usuarios de twitter que cumplían con las características requeridas.

3) Definición de la Escala de Opiniones y Sentimientos. Para obtener la clasificación más precisa sobre los mensajes a analizar se determinó necesario el uso de dos escalas de categorización. La primera se basa en una escala de Likert, que según Matas [17], es uno de los instrumentos más utilizados para la medición de opiniones, en este caso se determinó el uso de tres ítems, ya que este plantea que no se encuentran diferencias significativas en las versiones de tres, cinco y siete alternativas, y para esta temática no resulta relevante la subclasificación de las posturas:

- A favor de la migración
- Neutral ante la migración
- En contra de la migración.

La segunda escala hace referencia al sentimiento asociado a la opinión emitida, en esta investigación se tomó en cuenta la clasificación propuesta por Jack, Garrod y Schyns [18] de la Universidad de Glasgow, donde sugieren que las categorías utilizadas para medir las emociones estarán formadas por las siguientes categorías:

- Felicidad
- Tristeza
- Ira
- Miedo.

4) Enfoque de Análisis. En el momento del análisis, los métodos de aprendizaje automático tenían mayor relevancia debido a que utilizan las capacidades de un algoritmo de clasificación para que encuentre por sí mismo a qué polaridad pertenecen las palabras y oraciones, a partir de un conjunto de datos de entrenamiento, debido a esto, la investigación se centrará en este tipo de enfoque.

5) Algoritmos de Aprendizaje Automático

Con el fin de aplicar los modelos de aprendizaje automático es necesario realizar la selección de los algoritmos de clasificación que se tomarán en cuenta al momento de construir el modelo ya que cada uno de los algoritmos de clasificación tiene una forma distintiva para llevar a cabo la categorización de los datos. En la Tabla I, se pueden observar

los algoritmos seleccionados y una breve descripción de cada uno de ellos.

Tabla I: Algoritmos de Clasificación para el Análisis de Sentimientos

Algoritmo	Descripción
Naive Bayes (NB)	Es un modelo basado en la identificación de características para cada una de las categorías. Se basa en la predominancia de un grupo de características al realizar una predicción.
Maximum Entropy (MaxE)	Es un modelo basado en regresión logística, utiliza el principio de la máxima entropía, para determinar las características mutuamente dependientes que identifican una categoría.
Árboles de Decisión (DT)	Es un modelo basado en el uso de una serie de reglas simples extraídas de los datos para obtener así una predicción verificando el cumplimiento de alguna de las obtenidas.
Máquina de Vectores de Soporte (SVM)	Es un modelo basado en la identificación de planos que le permitan delimitar las diferentes categorías

B. Construcción del Corpus

Para recopilar los mensajes de la plataforma de Twitter, se utilizó la herramienta Twint, que es una librería del lenguaje Python para el scraping de tweets, la cual tuvo como resultado 5000 tweets, donde se incluían los más recientes de cada uno de los individuos de la población objetivo. Estos mensajes fueron almacenados posteriormente en un archivo .CSV, debido a la facilidad para operar sobre él con las herramientas disponibles en el lenguaje Python.

Posteriormente se llevó a cabo la limpieza del corpus que engloba múltiples etapas, la primera de ellas, es la selección de las columnas que serán utilizadas en el análisis las cuales son las siguientes:

- Fecha de publicación
- Usuario que lo publicó
- Contenido del tweet

Luego se procedió a la eliminación de aquellos tweets que no aportan valor al desarrollo de este trabajo como son los siguientes casos:

- Tweets repetidos
- Tweets que no tienen como enfoque la migración venezolana

Como resultado se descartaron un total de 1254 mensajes, resultando una muestra de 3746 tweets que conformaron el corpus inicial. Finalmente, para concluir esta etapa, se procedió a categorizar cada uno de los mensajes, en las escalas de sentimientos anteriormente descritas. En las Tablas II y III se muestran la distribución de los tweets para cada una de las clasificaciones.

Tabla II: Distribución de los Mensajes Según su Opinión sobre la Migración

Postura sobre la migración	Número de tweets
A favor de la Migración	2702
Neutral ante la Migración	191
En contra de la Migración	853

Tabla III: Distribución de los Mensajes según la Emoción Asociada

Emociones	Número de Tweets
Felicidad	871
Tristeza	913
Ira	878
Miedo	1084

C. Pre-procesamiento del Corpus

Después de obtener el corpus inicial se llevó a cabo el procesamiento del lenguaje natural, aplicando técnicas de análisis sobre los textos de entrada que permiten mejorar la capacidad de interpretación de los algoritmos, en este estudio se llevaron a cabo las siguientes etapas:

1) Tokenización

Se utilizó el paquete de nltk.tokenize para separar cada una de las palabras y signos de puntuación que conforman un mensaje y así analizar cada una de estas por separado en las siguientes etapas.

2) Eliminación de Ruido

En esta etapa haciendo uso del paquete nltk.corpus, se removieron los siguientes casos que afectaban el análisis

- Palabras de Parada
- Espacios múltiples
- Enlaces web
- Menciones a otros usuarios
- Emoticones

3) Normalización

En esta etapa haciendo uso del paquete ntlk.stem se aplicaron las siguientes técnicas de normalización

- Conversión de todos los tokens extraídos a letras minúsculas
- Stemming, donde cada una de las palabras es transformada a su raíz.

D. Muestreo del Corpus

Con la finalidad de evaluar cada uno de los modelos de clasificación seleccionados se realizaron tres pruebas, con base en el corpus creado, las cuales fueron variadas en el porcentaje de mensajes utilizados para su entrenamiento. En la Tabla IV, se pueden apreciar las pruebas realizadas y la distribución de los mensajes utilizados en cada apartado.

Tabla IV: Distribución del Corpus para las Pruebas de los Modelos

N° Prueba	Entrenamiento (%)	Entrenamiento (Totales)	Prueba (%)	Prueba (Totales)
1	70	2622	10	375
2	80	2997		
3	90	3371		

Los mensajes que forman parte de la categoría de entrenamiento son aquellos a partir de los cuales se creará el modelo de clasificación, mientras los que pertenecen a la categoría de prueba son aquellos utilizados para evaluar el desempeño de los modelos posteriormente.

E. Entrenamiento de los Modelos de Análisis de Sentimientos

Una vez construido el corpus y desarrollados los algoritmos de clasificación, se procedió a realizar la etapa de entrenamiento del algoritmo, en la cual se les suministraron a los algoritmos los datos ya procesados y etiquetados, para así obtener los modelos clasificadores que serán los encargados posteriormente de realizar las predicciones y categorizar nuevos elementos.

F. Evaluación de los Modelos de Análisis de Sentimientos

En esta etapa se llevaron a cabo la serie de pruebas descritas en la sección del muestreo del corpus, que permitieron estimar el desempeño de los modelos de clasificación, las cuales variaron el porcentaje de mensajes utilizados para su entrenamiento. Al evaluar los resultados de las métricas de rendimiento descritas en la Tabla V.

Tabla V: Métricas de Rendimiento para Evaluar los Modelos de Clasificación

Métrica	Descripción
Precisión	Representa el porcentaje de mensajes predichos de una categoría que pertenecen a ella realmente
Sensibilidad	Representa el porcentaje de mensajes reales de una categoría que fueron correctamente asignados en la predicción
Valor F1	Es un valor que representa la precisión y la sensibilidad en un único atributo
Exactitud	Representa el porcentaje de casos que el modelo ha acertado de todas las categorías

En la Tabla VI, se muestra el resumen de los resultados obtenidos en la evaluación de rendimiento para cada uno de los modelos de clasificación.

Tabla VI: Rendimiento de los Modelos de Clasificación en la 3era Prueba

N° Prueba	Precisión (%)	Sensibilidad (%)	Valor-F1 (%)	Exactitud (%)
Naive Bayes	79.23	76.56	77.27	77.07
Maximum Entropy	80.27	79.5	79.76	79.73
Arboles de Decisión	76.96	76.02	76.35	76.27
Máquinas de Vectores de Soporte	83.08	82.74	82.77	82.67

El rendimiento de los modelos de clasificación fue mejorando con el incremento de la data de entrenamiento, lo cual se cumplió para cada uno de los algoritmos a lo largo de las tres pruebas realizadas. Por ello, para el análisis de resultados se tomó como referencia principal la última prueba que utilizó la mayoría de los datos de entrenamiento.

El modelo que obtuvo mejores resultados bajo estas condiciones fue el basado en máquinas de vector de soporte, ver Tabla VII, que obtuvo una tasa de éxito en el 82.67% de los casos, que para la prueba realizada representó un total de 310 predicciones correctas de 375 textos seleccionados. Debido a esto, el modelo de clasificación que se utilizó principalmente en la aplicación desarrollada fue el modelo basado en máquinas de vector de soporte. Sin embargo, también se contempló la posibilidad de utilizar los otros modelos para observar los distintos comportamientos.

Tabla VII: Resultados para la Escala de Sentimientos con la Data de Entrenamiento al 90%

Categoría	precision	recall	f1-score	confusion matrix
Felicidad	85	78.16	81.44	68 2 5 12
Ira	89.89	90.91	90.4	2 80 1 5
Miedo	87.25	81.65	84.36	2 4 89 14
Tristeza	70.19	80.22	74.87	8 3 7 73
average	83.08	82.74	82.77	accuracy 82.67

Desde el punto de vista de la escala de sentimientos y opiniones, se puede determinar el rendimiento general que tuvieron cada una de las categorías a través de los múltiples modelos de clasificación, y así reconocer las tendencias generales de los modelos. En la Figura 2 podemos apreciar que la opinión a favor de la migración fue la que tuvo el mejor desempeño, mientras que las otras posturas no lograron un buen rendimiento, esto viene dado principalmente a la diferencia en la cantidad de mensajes de entrenamiento disponibles en el corpus inicial, donde hay una predominancia de textos asociados a la categoría a favor de la migración lo que permite un modelo más refinado para detectar este tipo de textos y genera una predisposición a su clasificación.

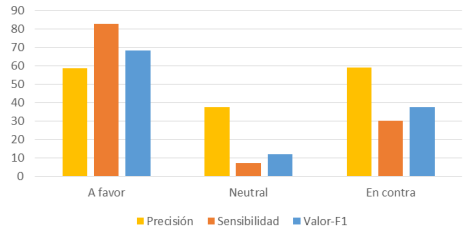

Figura 2: Rendimiento General según la Escala de Opiniones

En la Figura 3 por otro lado, podemos observar que las categorías de la escala de sentimientos se encuentran todas con un desempeño similar, esto se debe principalmente, a que la cantidad de datos de entrenamiento suministrados fueron similares para cada uno de los sentimientos.

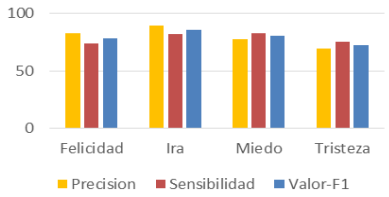

Figura 3: Rendimiento General según la Escala de Sentimientos

G. Resultados del Análisis de Sentimientos

Para obtener una visión general de la situación actual, se decidió analizar la mayor cantidad de registros de data, en vista que mejoraba el comportamiento del modelo, por lo cual se utilizó un total de 17.252 mensajes disponibles referentes al fenómeno de migración. Haciendo uso del modelo de clasificación basado en máquinas de soporte vectorial, que fue

el que obtuvo el mejor de desempeño de los modelos construidos. Los resultados fueron los siguientes:

2. Distribución de Sentimientos General

Con la finalidad de conocer la predominancia de alguna de las categorías de la escala de opiniones y sentimientos, en las Figuras 4 y 5 se disponen la distribución de los resultados obtenidos en la clasificación.

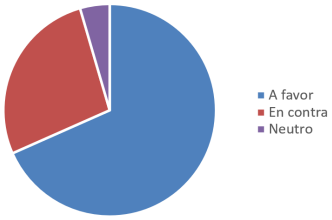

Figura 4: Distribución de la Escala de Opiniones

En la Figura 4 se puede observar que existe una tendencia de opinión clara hacia la postura a favor de la migración, el modelo clasificó aproximadamente un 64.94% de los mensajes a esta categoría, mientras que un 25.77% fueron asignados a la postura en contra y tan solo un 4.28% a la postura neutral.

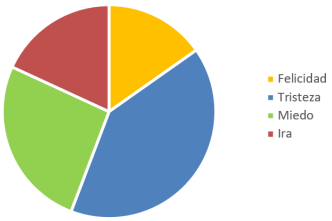

Figura 5: Distribución de la Escala de Sentimientos

Por otro lado, en cuanto a la escala de sentimientos, en la Figura 5 se puede observar una distribución más equilibrada, que se decanta ligeramente hacia la emoción de tristeza con un total de 40.55% de los mensajes, mientras que la distribución de los mensajes asignados a las categorías de miedo, ira y felicidad fueron similares con un total de 26.07%, 18.14% y 15.14% de los textos respectivamente.

3. Temas y Palabras Frecuentes

Los resultados de este estudio nos brindaron una mayor perspectiva sobre cuáles son los pensamientos de la población venezolana teniendo en cuenta su postura, para su obtención se evaluaron dos tipos de variables, los Unigramas y los N-gramas. Los Unigramas, en este caso, están representados por palabras, por ende, este método nos permite conocer aquellos términos que tienen mayor aparición en los textos, en la Figura 6 se pueden observar una lista con los Unigramas más relevantes y su número de apariciones. Por otro lado, los N-gramas están representados en este ámbito por un conjunto de palabras, que proporcionaron una perspectiva más detallada, en la Figura 7 se pueden observar los N-gramas que mayor aparición tuvieron y su número de apariciones.

Para identificar las relaciones que tienen estos términos con cada una de las categorías de la escala de opiniones, en las Figuras 8, 9 y 10 se presentan los N-gramas de las posturas a favor, en contra y neutral ante la migración respectivamente, con la intención de comparar los valores característicos de cada uno.

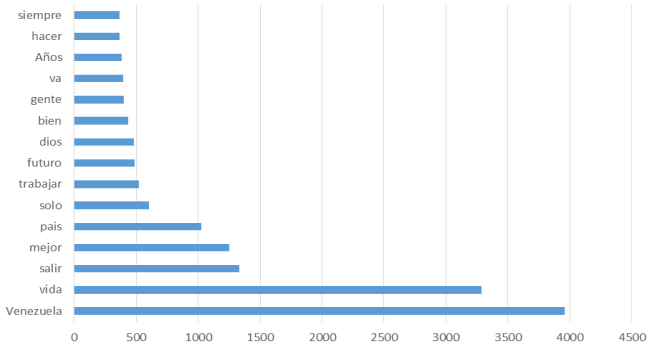

Figura 6: Unigramas con Mayor Aparición en el Análisis de Sentimientos General

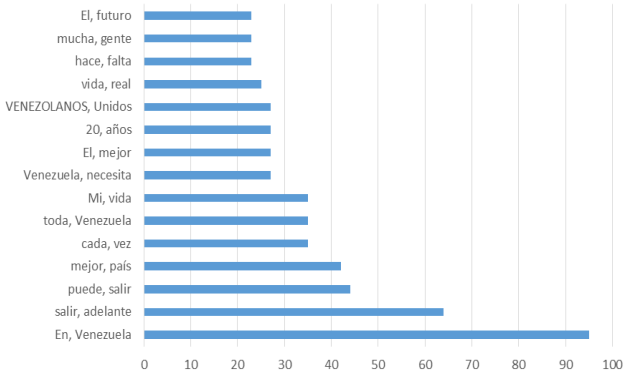

Figura 7: N-gramas con Mayor Aparición en el Análisis de Sentimientos General

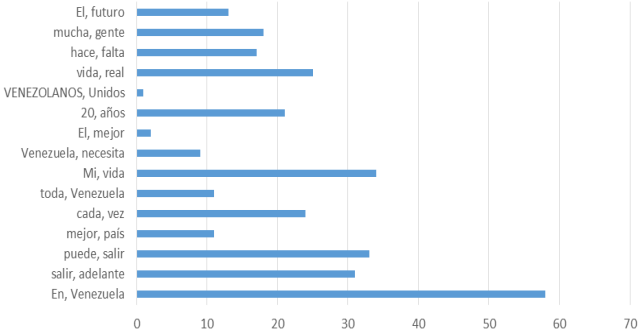

Figura 8: N-gramas con Mayor Aparición en Textos a favor de la Migración

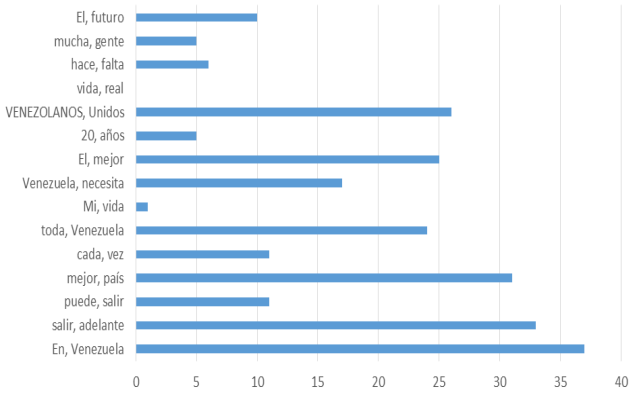

Figura 9: N-gramas con Mayor Aparición en Textos en Contra de la Migración

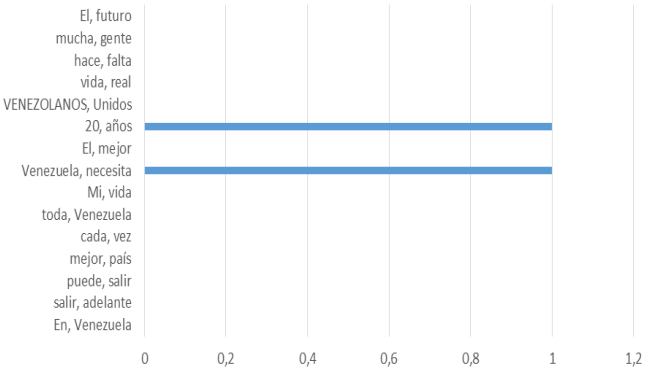

Figura 10: N-gramas con Mayor Aparición en Textos Neutrales ante la Migración

Se puede observar que los términos más relevantes de la postura a favor de la migración se enfocan a mencionar Venezuela, la vida que estos tengan en el país y la posibilidad de salir del mismo. En cuanto a la postura en contra de la migración, esta presenta un enfoque más centrado en la mejora de Venezuela y la contribución del pueblo venezolano para lograrlo. Y finalmente la postura neutral ante la migración solo hace presencia en un par de palabras debido a que el mayor número de los mensajes se ven identificados en las 2 posturas anteriores.

B. Desarrollo de la Aplicación

Para el desarrollo del aplicativo se llevó a cabo una serie de procesos, entre ellos se destacan, la creación de la base de datos, que fue la estructura encargada de almacenar los datos necesarios para la generación y evaluación de los modelos de clasificación, la construcción de una API, que fue la encargada de conectar la aplicación móvil con el servicio de base de datos y finalmente la aplicación que utilizó ambas herramientas para mostrar los resultados del análisis obtenido.

1) Base de Datos

En primera instancia, se diseñó una base de datos capaz de almacenar los valores necesarios para la etapa de entrenamiento y evaluación de los modelos de clasificación, debido a la simple naturaleza de los datos utilizados, el manejador de base de datos seleccionado fue PostgreSQL, una base de datos de tipo relacional, la cual provee un gran manejo y facilidad de acceso desde Python.

2) Application Programming Interface (API)

Para la comunicación entre la aplicación móvil y el servidor de base de datos, se desarrolló un servicio web que serviría de intermediario en este proceso. A través de la Tabla VIII, se describen las rutas para atender a las peticiones de la aplicación.

Tabla VIII: Rutas del Servicio Web Intermediario entre la Aplicación y el Servidor de Base de Datos

Ruta	Método HTTP	Descripción
/tweets	GET	Petición que devuelve todos los tweets de data de entrenamiento o los ya evaluados que estén almacenados en la Base de datos.
/tweets	POST	Petición que permite añadir un registro en la Base de Datos.
/tweets/scraping	GET	Petición que ejecuta la función para poder recolectar los datos de twitter y los procesa para ser almacenados en un archivo csv.
/tweets/classification	GET	Petición que ejecuta los modelos de predicción usando los datos de ejemplo para demostrar el funcionamiento de los mismos.
/tweets/classification/message	POST	Petición que recibe un mensaje y el nombre del usuario, para poder ejecutar los modelos predictivos sobre ese mensaje.
/tweets/fill	GET	Petición que solicita que se vacíen y rellenen la tabla de la base de datos.
/tweets/ngrams	GET	Petición para recolectar la frecuencia de las palabras que estén en los datos de prueba.
/statistics /general_probability_distribution	GET	Petición que devuelve el porcentaje de cuantos mensajes fueron clasificados en cada una de las escalas de sentimientos y los de opinión.
/statistics /distribution_by_date	GET	Petición que devuelve la cantidad de ocurrencias de cada una de las escalas de sentimiento y los de opinión pero esta vez bajo el parámetro de los meses del año.
/statistics /distribution_monograms	GET	Petición que devuelve las 15 palabras más frecuentes con sus datos de sentimientos y opinión según en los mensajes que haya aparecido.
/statistics /distribution_ngrams	GET	Petición que devuelve los 15 ngramas (Bigramas o Trigramas) más frecuentes con sus datos de sentimientos y opinión según en los mensajes que haya aparecido.

3) Aplicación Móvil

Para el desarrollo de la aplicación móvil se utilizó la herramienta de Flutter ya que como expone Herazo [19] es un marco moderno y sencillo para crear aplicaciones móviles, que tiene su propio motor de renderizado y permite ver los resultados en tiempo real.

El desarrollo de la aplicación se dividió en dos secciones principales, el desarrollo de la sección de las estadísticas del análisis de sentimientos y opiniones y la sección de predicciones utilizando los modelos de clasificación desarrollados. Para ambas secciones se hace uso de la API descrita en la Tabla VIII que permite la interacción con el servidor de base de datos.

Para la representación de las gráficas, se utilizó la librería *charts_flutter*, la cual cuenta con un conjunto extensivo de gráficas para datos estadísticos.

VI. Conclusiones y Recomendaciones

A partir del análisis del problema en perspectiva de los resultados obtenidos en el trabajo, se extrajeron las siguientes conclusiones:

- La revisión del estado del arte indica que, para llevar a cabo el proceso de análisis de sentimientos y minería de opinión, la plataforma Twitter constituye una gran fuente de data, pues es de las redes sociales más utilizadas para la publicación de textos, los cuales pueden ser extraídos y procesados mediante las herramientas que brinda el lenguaje de Python, para la construcción del corpus inicial y posteriormente los modelos de clasificación.
- En la determinación de los criterios para el análisis de sentimientos y minería de opinión se identifican los parámetros que apoyan al proceso de recolección y análisis de los textos, entre ellos se destacan las características de la población objetivo, la escala de los sentimientos y opiniones que categorizan los textos y las métricas de rendimiento para evaluar el desempeño de los modelos que clasifican los textos relacionados al tema de la migración venezolana.
- Para diseñar modelos predictivos en base a textos publicados en redes sociales, es necesario la construcción de un corpus de palabras presentes en las publicaciones de venezolanos para luego determinar los sentimientos y opiniones referentes al tema de la migración con el uso de las técnicas de procesamiento del lenguaje natural y los algoritmos de aprendizaje automatizado. La calidad de los datos recolectados afecta de manera proporcional la calidad y validez de los modelos extraídos.
- Esta investigación muestra el ventajoso uso de los algoritmos de clasificación y la selección de categorías para la escala de sentimientos y opiniones para la construcción de modelos de clasificación que abarquen todas las perspectivas de la población venezolana sobre el fenómeno migratorio. Una vez entrenado el clasificador, se lleva a cabo el proceso de evaluación, haciendo uso de las métricas de rendimiento que permiten determinar su desempeño al categorizar textos relacionados a la migración venezolana.
- El resultado de la evaluación de los modelos de clasificación muestra que el basado en el algoritmo de Máquina de Soporte Vectorial es el que posee el mejor desempeño, obteniendo una exactitud general del 82.67% al momento de clasificar los mensajes de venezolanos sobre el fenómeno migratorio. El uso de este modelo para realizar el análisis de sentimientos y minería de opinión sobre la población venezolana en el año 2021 respecto al fenómeno migratorio venezolano, destaca que el sentimiento predominante es la tristeza, con un porcentaje de aparición en textos del 40.55%, y en cuanto a las opiniones, la postura a favor de la migración prevalece en el 64.94% de los casos.
- La implementación de los diferentes modelos predictivos mediante una aplicación móvil resulta ser útil, ya que permite la publicación de los resultados del análisis de datos, como la distribución general de la escala de sentimientos y opiniones, su distribución a través del tiempo y las palabras clave que identifican a cada una de las categorías, mediante una plataforma potencial de uso masivo, que apoya el proceso de toma de decisiones, suministrando información oportuna, para afrontar el fenómeno migratorio.
- El resultado obtenido en el presente trabajo evidencia la factibilidad del uso del análisis de sentimiento y minería de opinión de publicaciones en redes sociales, para la construcción de un modelo predictivo sobre el fenómeno migratorio venezolano, demostrando la razón del aumento de este tipo de investigaciones para el descubrimiento y explotación de opiniones y sentimientos de una comunidad respecto a un evento en un momento determinado.

En base a los resultados obtenidos en la presente investigación, a continuación, se enumeran una serie de recomendaciones a tomar en cuenta para trabajos futuros:

- Mejorar los modelos de clasificación, explorando la aplicación y/o combinación de otras técnicas de análisis de datos, como otras técnicas de agrupamiento, reglas de asociación, entre otras.
- Refinar el comportamiento de los modelos de clasificación descubiertos, utilizando un conjunto de datos más elevado y equitativamente distribuido entre todas las categorías.
- Explorar el uso de otros medios sociales para la recolección de datos, ya que pueden tener una representación distinta de los diferentes grupos de la sociedad venezolana.
- Experimentar el uso de otras arquitecturas, modelos de entrenamiento, librerías y entornos de desarrollo, con la finalidad de ampliar el conocimiento sobre otras opciones de análisis de sentimientos, y así elegir la más adecuada para un problema en específico.
- Ampliar el alcance de la aplicación móvil implementada, para abarcar la interacción entre usuarios, a modo de red social, permitiendo una mejora en la exposición de información para la toma de decisiones.

REFERENCIAS

[1] J. Arias, M. Villasís, y M. Miranda. *El Protocolo de Investigación III: La Población de Estudio.* Revista Alergia México, vol. 63, no. 2, pp. 201-206. México. Abril, 2016.

[2] R. Arias, N. Arias, M. Morffe, C. Martínez, y T. Carreño. *Informe sobre la Movilidad Humana Venezolana II Realidades y Perspectivas de Quienes Emigran (8 de abril al 5 de mayo 2019).* ISBN: 978-980-7879-02-6, San Cristóbal, Venezuela, Junio 2019.

[3] Plataforma de Coordinación Interagencial para Refugiados y Migrantes de Venezuela. *Refugiados y Migrantes de Venezuela*. 2021. https://www.r4v.info/es/refugiadosymigrantes.

[4] Observatorio Proyecto Migración Venezuela. *Percepción en Redes Sociales sobre la Migración desde Venezuela*. 2020.

[5] H. Hütt. *Las Redes Sociales: Una Nueva Herramienta de Difusión.* Reflexiones, vol. 91, no. 2, pp. 121-128. Universidad de Costa Rica San José, Costa Rica. 2012. https://www.redalyc.org/pdf/729/72923962008.pdf

[6] B. Liu. *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers. Chicago, United States. 2012.

[7] M. Salaz. Detección de Patrones Psicolingüísticos para el Análisis de Lenguaje Subjetivo en Español. Tesis Doctoral del Programa Oficial de Doctorado en Informática. Universidad de Murcia. España 2017. https://rua.ua.es/dspace/bitstream/10045/74617/1/PLN_60_10.pdf

[8] A. Cestari. *Propuesta para Automatizar la Asociación de Emociones a Textos en Español.* Trabajo de Grado. Escuela de Ingenieria Informática, Universidad Católica Andrés Bello, Puerto Ordaz, Venezuela, 2019.

[9] C. Arango y C. Osorio. *Aislamiento Social Obligatorio: Un Análisis de Sentimientos Mediante Machine Learning.* Suma de Negocios, ISSN-e 2027-5692, ISSN 2215-910X, vol. 12, no. 26, pp. 1-13, 2021.

[10] M. Arguedas, J. Beita, F. Rodríguez, J. Umaña y M. Vaca. *Crisis Migratoria en Colombia y Costa Rica: Una Visión desde el Análisis de Sentimientos.* Revista humanidades ISSN electrónico: 2215-3934, vol. 10, no. 2, Junio 2020. https://revistas.ucr.ac.cr/index.php/humanidades/article/view/42238

[11] C. González, *Clasificador de Texto Mediante Técnicas de Aprendizaje Automático*. Trabajo de Grado en Ingeniería Informática, Escuela Técnica Superior Universidad Politécnica de Valéncia, España 2020.

[12] A. Mundalik, *Aspect Based Sentiment Analysis Using Data Mining Techniques Within Irish Airline Industry*. MSc Research Project Data Analytics. 10.13140/RG.2.2.13637.40165. Abril 2018. http://norma.ncirl.ie/3413/1/aishwaryamundalik.pdf

[13] E. Páez y A. Monroy, *Implementación de un Modelo de Análisis de Sentimientos con Respecto a la JEP basado en Minería de Datos en Twitter.* Trabajo de Grado. Universidad Católica de Colombia. Facultad de Ingeniería. Programa de Ingeniería de Sistemas. Bogotá, Colombia 2020.

[14] P. Chapman, J. Clinton, R. Keber, T. Khabaza, T. Reinartz, C. Shearer, and R.Wirth. CRISP-DM 1.0 Step by Step BIguide. Edited by SPSS. 2000.

[15] F. Peralta, *Proceso de Conceptualización del Entendimiento del Negocio para Proyectos de Explotación de Información.* Revista Latinoamericana de Ingeniería de Software, vol. 2, no. 5, pp. 273-306, ISSN 2314-2642, 2014.

[16] J. Saura, P. Palos-Sánchez y M. Ríos. *Un Análisis de Sentimiento en Twitter con Machine Learning: Identificando el Sentimiento sobre las Ofertas de BlackFriday*. Revista Espacios, vol. 39, no. 42, 2018.

[17] A. Matas. *Diseño del Formato de Escalas Tipo Likert: Un Estado de la Cuestión.* REDIE ISSN 1607-4041, vol. 20, no. 1, pp. 38-47. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1607-40412018000100038

[18] R. Jack, O. Garrod, and P. Schyns. Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. Current Biology, Vol. 24, Issue 2, Pages 187-192, ISSN 0960-9822,2014. 2014. https://www.sciencedirect.com/science/article/pii/S0960982213015194

[19] L. Herazo. *¿Qué es Flutter y por qué Utilizarlo en la Creación de Apps Móviles?*. 2021. https://anincubator.com/que-es-flutter-y-por-que-utilizarlo-en-la-creacion-de-apps-moviles/