



Tipo de artículo: Artículos originales
Temática: Inteligencia artificial
Recibido: 10/07/2022 | Aceptado: 02/07/2022 | Publicado: 30/09/2022

Identificadores persistentes:
ARK: [ark:/42411/s9/a66](https://nbn-resolving.org/urn:ark:/42411/s9/a66)
PURL: [42411/s9/a66](https://nbn-resolving.org/urn:purl:42411/s9/a66)

Clasificación de tutoriales en YouTube basándonos en el análisis de sentimientos realizados a sus comentarios

Classification of tutorials on YouTube based on the analysis of feelings made to your comments

Valeria Alejandra Goyzueta Torres ^{1*}, Ronald Fabricio Centeno Cardenas ²[\[0000-0001-6639-8603\]](https://orcid.org/0000-0001-6639-8603), Victor Andre Ranilla Coaguila ³

¹ Universidad La Salle. Arequipa, Perú. vgoyzuetat@ulasalle.edu.pe

² Universidad La Salle. Arequipa, Perú. rcentenoc@ulasalle.edu.pe

³ Universidad La Salle. Arequipa, Perú. vranillac@ulasalle.edu.pe

* Autor para correspondencia: vgoyzuetat@ulasalle.edu.pe

Resumen

El flujo de información surge día a día mediante internet de manera continua gracias a las constantes interacciones presentes entre los usuarios, estas interacciones presentan sentimientos que pueden ser positivos o negativos. Esto ayuda mucho a los creadores de contenido de las redes sociales a comprender cuan útil es lo que ellos hacen para sus seguidores, y es que, si estos son un gran número, un análisis hecho por una sola persona no es suficiente. Para ello es necesario el uso de herramientas que operan con grandes cantidades de datos como BERT, que es un modelo que ayuda al análisis de sentimientos y clasificación de comentarios basados en lo que expresa uno de estos. En este trabajo se usará este modelo para la clasificación de comentarios de YouTube y clasificación de videos de esta misma plataforma, valorando estos videos según su contenido y ayudando a los espectadores a elegir los videos si es que estos lo ayudarán con respecto a lo que se encuentran buscando. Se harán además uso de métricas y de sugerencias futuras para la propuesta mencionada en este trabajo.

Palabras clave: Análisis de sentimientos, Comentarios de Youtube, Clasificación de videos.

Abstract

The flow of information arises day by day through the Internet in a continuous way thanks to the constant interactions between users, these interactions present feelings that can be positive or negative. This helps social media content creators a lot to understand how useful what they do is for their followers, and if these are a large number, an analysis done by a single person is not enough. For this, it is necessary to use tools that operate with large amounts of data, such as BERT, which is a model that helps analyze sentiments and classify comments based on what one of them

expresses. In this work, this model will be used for the classification of YouTube comments and the classification of videos on this same platform, evaluating these videos according to their content and helping viewers to choose the videos if they help them concerning what is expected. find searching. This work will also use future metrics and suggestions for the proposal.

Keywords: *Sentiment Analysis, Youtube Comments, Video Ranking.*

Introducción

La capacidad de adaptación de la tecnología sobre diferentes contextos, no deja atrás a ámbitos como la enseñanza y el aprendizaje, pues permite tanto a estudiantes como a maestros ser partícipes del intercambio de conocimiento sin importar la distancia ni las limitaciones físicas presentes. Este conocimiento viaja de plataforma en plataforma y está siempre presente en donde más interacción entre personas ocurre, siendo en la actualidad las redes sociales.

La vida en las redes sociales es muy amplia y compleja de entender, pero puede resumirse como la interacción entre diversos usuarios que intercambian información, intereses y opiniones sobre un tema en específico de manera remota. Esta interacción se lleva en mayor medida dentro de los comentarios de las publicaciones que realizan otros usuarios, llevando a dar una visión general sobre una opinión de un tema en específico que muchas veces no es muy acertada, pues la cantidad de información dentro de este mismo es muy diversa.

Dentro de las plataformas donde más se realiza este intercambio de opiniones se encuentran: Facebook, Twitter y YouTube, siendo esta última considerada la plataforma más grande de videos en la red, donde se realizan subidas e intercambio de contenido, tanto educativo como de entretenimiento. En ella cada minuto se sube un aproximado de 500 horas de videos y más de un billón de videos dentro de la misma son visualizados en diversos lugares del globo [2], dentro de los cuales se encuentran los videos tutoriales.

Los videotutoriales son una herramienta que ayudan al fortalecimiento de conocimiento, como la aclaración de dudas sobre un tema en específico, que puede ser adquirido en un entorno presencial o simplemente sea un nuevo concepto que quiere ser aprendido. Esta clase de material tiene un mismo formato: Un tutor enseñando acerca de un tema que domina, haciendo uso de herramientas para facilitar la comprensión de sus instrucciones y convirtiendo su video en un material reconocido y adecuado de enseñanza y aprendizaje. Gracias al alcance de internet, la propagación de estos

materiales se realiza rápidamente y más en YouTube, pues es mucho más sencillo transmitir conocimientos mediante videos que haciendo empleo de otros sentidos. Cada video que es subido a YouTube es categorizado basándose en su contenido, y calificado según ese mismo sobre la base de los usuarios que interactúan con él. Todo esto gracias a tres herramientas brindadas por la misma plataforma: los botones de like, dislike y los comentarios. La calidad de los videos es determinada por estos indicadores y marcan una reputación sobre quien subió este material. Las principales fuentes que determinaban esta calidad eran las cantidades que eran brindadas por los botones previamente mencionados, pues estos determinaban los niveles de aprobación que poseía un video y cuán útil es con respecto al ámbito donde este se enfocaba. Esto hasta el año 2021 se deshabilitó la visualización de la cantidad de estos botones, dejando a los usuarios con una vaga idea de la utilidad de un video a simple vista. Pero haciendo un análisis más profundo, los comentarios dentro de los videos son los que también determinan la reputación de cada creador y la calidad de cada video, pues cada comentario contiene información valiosa que puede ayudar a la clasificación de un video y su relevancia en la plataforma.

Cada comentario contiene palabras clave que ayudan a identificar una emoción asociada a una respuesta de reacción al video, que puede ser tanto positiva como negativa. El hecho de identificar estas palabras asociadas a emociones en sencillo, pero cuando la cantidad de datos es exponencial, pues estamos hablando de plataformas globales, es necesario el uso de algoritmos y métodos de procesamiento de lenguaje natural. Al realizar la segmentación de comentarios en dos categorías, basados en palabras clave relacionadas con sentimientos de usuarios, indica la relación directa con un área del procesamiento del lenguaje natural (NLP) llamada Análisis de Sentimientos (SA). Como su nombre lo indica, su principal objetivo es la extracción de sentimientos dentro de comentarios, para esto los comentarios deben ser clasificados con base en información puntual u opiniones subjetivas [5]. El análisis de sentimientos normalmente está compuesto de 4 fases: Extracción de información, Procesamiento de data, Clasificación de sentimientos y Presentación de la salida, dentro de los cuales el Procesamiento de data es el paso que más esfuerzo requiere, pues envuelve procesos como: Preprocesamiento de texto, Feature Extraction y Feature Selection. Una vez completada esta fase, la clasificación hace uso de algoritmos de Machine Learning (ML) para realizar una clasificación de polaridad de comentarios, que ayudará a la clasificación de videos según su utilidad. Este estudio se enfoca en la obtención de opiniones basadas en los comentarios que son manifestados a manera de reacción a los videos tutoriales que son subidos a la plataforma de YouTube. Para poder determinar su utilidad y realizar la clasificación según esta misma (útil, inútil), además de representar las clasificaciones de comentarios en variables numéricas. Para ello, primero se realizará un preprocesamiento de información, dejando aquellos comentarios que presenten sentimientos positivos o negativos

explícitamente involucrados luego haciendo uso de algoritmos de procesamiento de lenguaje natural, determinar el porcentaje predominante y concretar la primera tarea enunciada, la tarea de clasificación. Este trabajo está organizado de la siguiente manera: La introducción, motivación del trabajo, los trabajos previos que han sido ejecutados en el área del procesamiento del lenguaje natural, el marco teórico, la propuesta que involucra parte de la implementación, los resultados y finalmente las conclusiones y recomendaciones aplicables en trabajos posteriores relacionados a este sector computacional.

II. Motivación

El siguiente trabajo busca desarrollar una herramienta de análisis de sentimientos de comentarios que surgen como respuesta a videos tutoriales que son subidos a la plataforma de YouTube, para enseñar las palabras más predominantes que indiquen la utilidad de un video a base de su contenido y la opinión popular que es generada por los usuarios. Las preguntas que se intentan resolver, mediante la elaboración de este trabajo, son las siguientes:

- P1: ¿Qué palabras son las más frecuentes cuando un video es considerado útil por los usuarios?
- P2: ¿Qué palabras son las más frecuentes cuando un video es considerado inútil por los usuarios?
- P3: ¿Qué porcentaje de comentarios presentan una posición 'neutral' o 'indefinida'?

III. Trabajos relacionados

Para la elaboración de este trabajo hemos recolectado trabajos previos, cuya intención se centra en el análisis de sentimientos con base en comentarios emitidos como reacción a videos publicados en YouTube: Hanif et al. [3] elaboran un modelo basado en NLP, que se encarga de retornar a los usuarios los videos más relevantes y populares dependiendo de los comentarios de las personas. Esta propuesta hace uso de herramientas de preprocesamiento de texto encargadas de extraer aquellos caracteres y conjuntos de palabras que son totalmente irrelevantes, como enlaces, símbolos, caracteres, emoticones y aquellos comentarios que no se encuentran en el idioma al que está orientada la propuesta para poder eliminarlos, además de los signos de puntuación.

Una vez eliminados los caracteres que no sirven para la propuesta, se procede a generar un dataset limpio, con la forma singular de las palabras que conforman los comentarios. A todos los adjetivos que conforman los comentarios se les aplicó un POS Tagger, para generar un segundo dataset. Sobre la base de estos dos datasets generados, se realizó el

análisis de sentimientos, que presenta como limitación o como condición determinante del rendimiento, la forma en la que los comentarios y las palabras que los conforman son procesadas y el análisis semántico de su contenido.

Otra de las propuestas que hacen análisis de sentimientos dentro de comentarios hacia material audiovisual es la propuesta elaborada por Obadimu et al. [11] que si bien no hacen referencia a si los comentarios son positivos o negativos como tal, hace énfasis en el análisis de las palabras que representan a un sector, que en este caso es la toxicidad dentro de los comentarios de YouTube sobre la opinión de una corriente política. Ellos toman en consideración 5 tipos de toxicidad. Haciendo uso de una CNN, se logró determinar cuándo un comentario era tóxico dentro de una discusión entre usuarios. Para tokenizar las palabras que forman parte de los comentarios se hizo uso de la librería de Python: NLTK. Que es ampliamente utilizada en el área del procesamiento del lenguaje natural. El trabajo elaborado por Obadimu et al. presenta sugerencias de sanción frente a esta clase de comentarios identificados dentro de Youtube. Cunha et al. [1] nos presenta una manera de clasificar los comentarios de un video con base en la influencia que tienen sobre los usuarios, la relevancia del video y la calidad visual de este mismo.

Cada clasificación presenta tres posibles posiciones: positivo, negativo y neutral. Esta propuesta hace uso de heurísticas de preprocesamiento de texto para luego aplicar Deep Learning en la predicción de las reacciones de los usuarios a ciertos videos manifestadas dentro de los comentarios. Los autores sugieren que un preprocesamiento más exhaustivo puede ser necesario para poder mejorar la efectividad del modelo propuesto. Singh y Tiwari [12] nos presentan una forma de realizar el análisis de comentarios de Youtube haciendo uso de diferentes técnicas de Machine Learning y Deep Learning. Además de librerías de Python como: SciKitLearn que ayuda a la conversión de la data textual a numérica para poder interactuar con ella a manera de vectores. Ellos hicieron uso de tareas clásicas del procesamiento del lenguaje natural: Lemmatisation y la remoción de caracteres que no aportarían nada al análisis de sentimientos, como los signos de puntuación. Para la parte de clasificación, se usaron diversos algoritmos, entre ellos los 6 más conocidos: Bayes Naives, Support Vector Machine, etc. Haciendo una comparación entre ellos, se llegó a la conclusión de que el algoritmo que arroja mejores resultados es el de Random Forest, y el que resultados más bajos logró obtener fue el de Naive Bayes, que necesitó menos preprocesamiento de texto. Muhammad et al. [10] presenta una clasificación a los comentarios de los videos de YouTube combinando los métodos de Naïve Bayes y Support Vector Machine (NBSVM) con un enfoque de Clasificación Binaria. El uso de estos métodos fue elegido por ellos porque Naïve Bayes es muy bueno en la clasificación de textos con un pequeño número de datos, mientras que Support Vector es muy bueno en la clasificación de textos con un número relativamente alto de datos. Los resultados obtenidos muestran que la

combinación de Naïve Bayes y Support Vector Machine produce un mejor nivel de precisión y un mayor rendimiento. Sin embargo, la combinación de varios clasificadores no siempre aumenta la precisión de las clasificaciones.

IV. Marco teórico

Inteligencia Artificial

La inteligencia artificial hace referencia a sistemas informáticos con la capacidad de hacer predicciones o realizar acciones basándose en los patrones de los datos disponibles y poder aprender de sus errores para ser más precisos [13]. Una inteligencia artificial avanzada procesa la información nueva con suma rapidez y precisión, es por ello que generalmente se puede asociar el entendimiento humano a la computadora por medio de este tipo de tecnologías.

Procesamiento de lenguaje Natural

El procesamiento del lenguaje natural es un enfoque de la inteligencia artificial que ayuda a interpretar el lenguaje humano a través de algoritmos de análisis de texto y reconocimiento de texto [13], haciendo uso de elementos de ciencia y lingüística computacional, para que el lenguaje humano sea procesado bajo una correcta comprensión por parte del computador. Su importancia radica en ayudar al entendimiento entre un computador y una persona a través del lenguaje humano, cooperando así en la realización de múltiples tareas basadas en reconocimiento de voz, interpretación y análisis, además de la medición del sentimiento.

Medición de sentimiento

El análisis de sentimiento hace referencia al uso de NLP, por medio de herramientas basadas en lingüística computacional y análisis de texto, para reconocer y sustraer información relacionada con los recursos analizados. La medición se realiza basándonos en un tratamiento enfocado en relaciones estadísticas y de asociación lingüística que repara en la creación de conclusiones referidas a encontrar una meta u objetivos[4]. Para realizar un análisis o medición de sentimiento se utilizan Datasets basados en reseñas, opiniones y comentarios, que brindaran una idea determinada al enfoque de análisis impuesto.

Datasets

El término DATASET se refiere a un archivo que contiene uno o más registros de información. Estos registros son seleccionados y clasificados a base de al enfoque de su aporte. Muchas veces estos registros de información

se utilizan para almacenar información que necesitan las aplicaciones o el propio sistema operativo; Al final se catalogan basándonos en el tipo de información al cual se enfocan. Al término de su preprocesamiento y su clasificación, son procesados con métodos y técnicas de inteligencia artificial basados en algoritmos CNN, BOW o BERT , para analizar y concluir la correcta interpretación de los datos procesados.

Convolutional Neural Network

Son un tipo de redes neuronales artificiales donde las neuronas corresponden a campos receptivos de una manera muy similar a las neuronas en la corteza visual primaria de un cerebro biológico. Este tipo de red es una variación de un perceptrón multicapa, sin embargo, debido a que su aplicación es realizada en matrices bidimensionales, son muy efectivas para tareas de visión artificial, como en la clasificación y segmentación de texto e imágenes, entre otras aplicaciones [13].

Las CNN aprenden por lo general a reconocer una diversidad de objetos dentro de imágenes [13], además de interpretar y entender textos por medio de oraciones y palabras específicas, pero para ello necesitan entrenarse con una cantidad importante de muestras, las cuales pueden ser muchas veces brindadas por un gran banco de información, generalmente conocidos como Big data.

Big data

Big data se refiere a conjuntos de datos que son demasiado grandes o complejos para ser tratados por el software de aplicación de procesamiento de datos tradicional. Los desafíos del análisis de big data incluyen la captura de datos, el almacenamiento de datos, el análisis de datos, la búsqueda, el intercambio, la transferencia, la visualización, la consulta, la actualización, la privacidad de la información y la fuente de datos [13].

El uso actual del término big data tiende a referirse al uso de análisis predictivos, análisis del comportamiento del usuario u otros métodos avanzados de análisis de datos que extraen valor de los grandes datos, y rara vez a un tamaño particular de Datasets[4].

Herramientas más populares en el uso de análisis de sentimientos

El análisis de sentimientos por lo general consiste en valorar y estimar la disposición de un usuario, en relación con sus opiniones; con la finalidad de obtener información que permita comprender su postura y reacción respecto a un servicio o producto en específico. Por ello se utilizan diversas herramientas de análisis de

sentimiento relacionadas con tecnologías avanzadas de inteligencia artificial, entre las que se encuentran el enfoque de procesamiento del lenguaje natural, análisis de frases o textos y data science [7], pero entre las que destacan son el uso de estos enfoques a través de redes neuronales recurrentes.

Es por ello que la mayoría de estas herramientas son desarrolladas de forma modular y su entrenamiento es desarrollado por la comunidad open source, obteniendo así diversos métodos de trabajo basados en el análisis del lenguaje natural para la extracción de datos y el desarrollo analítico de sintaxis basándonos en entidades, la detección de sentimiento y la clasificación de contenido.

Entre las herramientas más reconocidas para el desarrollo de NLP con redes neuronales recurrentes encontramos:

a) Google CLOUD Platform - Natural Language IA

Como su nombre lo refiere, Google creó una herramienta para obtener información de textos no estructurados mediante el aprendizaje automático del algoritmo de Google; todo ellos con el objetivo de realizar un análisis de texto perspicaz con el aprendizaje automático que extrae, analiza y almacena texto, entrena modelos personalizados de aprendizaje automático sin una sola línea de código con AutoML y aplique la comprensión de lenguaje natural a las aplicaciones con API de lenguaje natural. En síntesis, esta herramienta utiliza el análisis de entidades para encontrar y etiquetar campos dentro de un documento, luego las analiza para comprender las opiniones de los usuarios, encontrando información procesable sobre servicios y productos.

b) Open IA GPT3

Open IA desarrolló una ampliación de los modelos de análisis de lenguaje natural que mejora el rendimiento de desarrollo de tareas y pocos intentos. Desarrollo GPT3, un modelo de lenguaje autorregresivo con millones de parámetros que ayuda a determinar una respuesta sugerente a la determinación de sentimientos en el desarrollo de análisis de opiniones y predicciones[6].

c) BERT

Bert significa, representaciones de codificador bidireccional de Transformers y es un modelo de aprendizaje automático utilizado para tareas NLP. Fue entrenado con Wikipedia en inglés y BookCorpus;

actualmente existen dos variaciones de BERT preentrenadas, el modelo base de 12 capas neuronales y otro de 24. En rendimiento, BERT es realmente superado por GPT3 de Open IA, pero el acceso limitado a GPT3 obliga a utilizar el enfoque BERT.

V. Metodología

Propuesta

Para este trabajo, un dataset de comentarios de videos de YouTube que son catalogados y juzgados por su contenido ha sido utilizado [9]. Para continuar con las tareas propuestas, un proceso de limpieza ha sido necesario para poder obtener la información en un formato más legible y entendible, además de ordenado al momento de realizar la tarea de clasificación. La tarea de limpieza se llevará a cabo con librerías de procesamiento de lenguaje natural y Machine Learning, eliminando aquellos caracteres que no son de utilidad e interrumpe el entendimiento del comentario. Para luego obtener los comentarios filtrados y realizar un análisis de sentimientos que permitirá la clasificación de comentarios en alguna de las tres clases: positivo, neutral o negativo. Y con base en esta clasificación poder clasificar un video según la cantidad de registros presentes en una clase Figura 1.

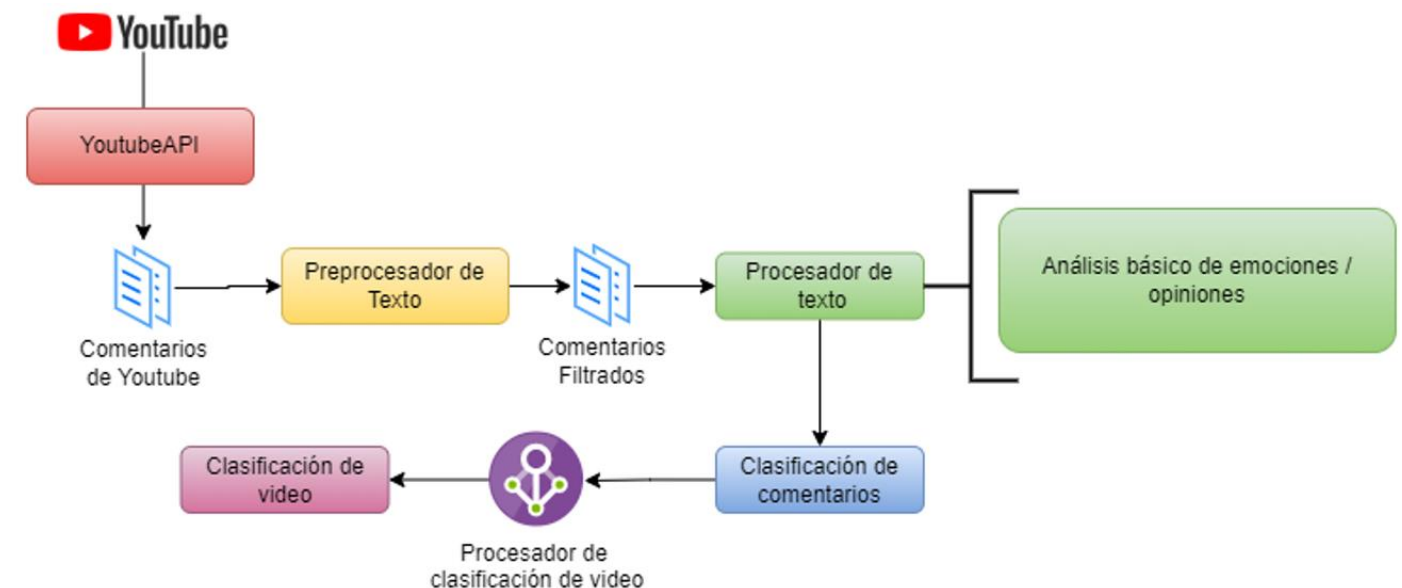


Figura 1, Pipeline de la propuesta.

Descripción de la data

Hemos usado este dataset que se encuentra disponible desde el año 2020, y cuya última actualización se llevó a cabo ese mismo año. Estos comentarios han sido extraídos de la Play Store de Google, y se enfocan en comentarios de crítica a aplicaciones, que en general suelen ser más de 82 billones de aplicaciones. La Tabla 1, enseña los atributos de dataset.

Tabla 1. Descripción de atributos

Atributos	Descripción
reviewID	ID del comentario
UserName	Usuario
UserImage	Perfil de Usuario
Content	Contenido de la aplicación
thumbsUpCount	'likes' hacia el comentario
at	La aplicación
replyContent	Respuestas de un comentario
repliedAt	Comentario respondido

El archivo pesa un aproximado de 3.45Mb, conteniendo más de 12000 registros dentro de el, muestra además una clasificación basa en números del 1 al 5, representado en la Figura 2.

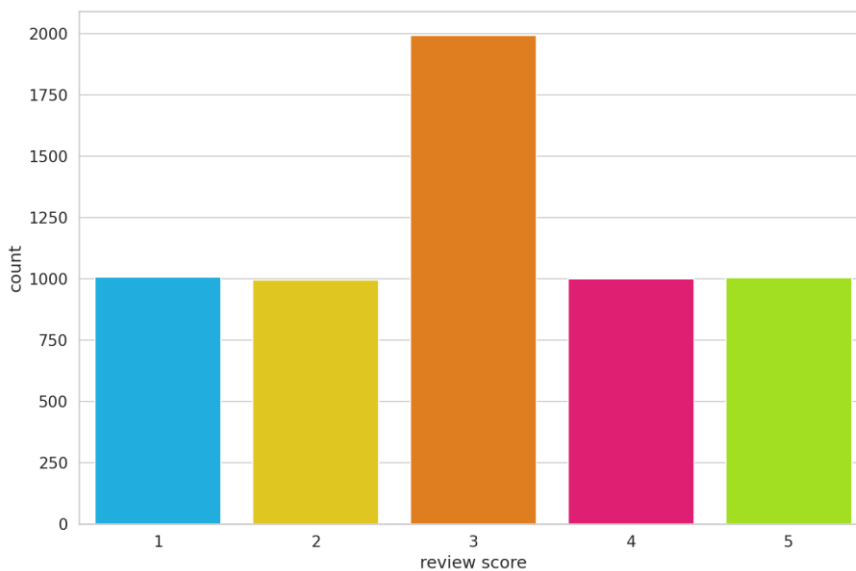


Figura 2, Clasificación numérica de los comentarios acerca de una aplicación.

Herramientas

Para la implementación de la propuesta mencionada anteriormente se ha decidido hacer uso de las siguientes herramientas:

1. BERT: Ya mencionado anteriormente, la funcionalidad de este modelo bidireccional de aprendizaje es de gran utilidad, pues entre las principales tareas que realiza se encuentra la de clasificación de texto basada en los sentimientos que se encuentran dentro de los comentarios. Con esta herramienta, dentro de la implementación de esta propuesta se ha usado de manera que podamos clasificar los comentarios en tres clases, obteniendo la clase predominante y clasificando de esta manera el videotutorial que nosotros hemos seleccionado. Siendo mucho más específicos la manera en la que BERT es usado en la implementación de la clasificación de comentarios y, por tanto, de videos es en la división de comentarios con base en su contenido que ayuda a la clasificación de videos como tales.

2. API de YouTube: Esta herramienta es de gran utilidad, pues ayuda a conseguir los comentarios, de manera que haciendo uso del resto de herramientas podemos clasificarlos. Podemos decir que esta herramienta es la fuente principal de datos que ayudan a poner a prueba lo que se ha implementado, haciendo que la propuesta mencionada dentro de este trabajo sea aplicable dentro de entornos en la vida real.

3. Python: Lenguaje de programación interpretado de tipado fuerte y dinámico, que soporta la programación orientada a objetos, la programación funcional y la programación imperativa, usado en áreas como el Machine Learning, Deep Learning, Reconocimiento Facial, y otras tareas más. Se usa para la construcción de variedad de aplicaciones usadas en cualquier contexto.

Librerías

- **Numpy:** Librería de Python que ayuda en la creación de vectores y matrices de grandes proporciones, acompañada de una gran cantidad de funciones de alto nivel que ayudan en la interacción con estas estructuras numéricas.
- **Pandas:** Biblioteca usada para el análisis y manipulación de datos. Es una extensión de Numpy. Ayuda en la definición de nuevas estructuras de datos que pueden ser accedidas mediante índices o lo que pueden ser

nombres para las columnas y ubicaciones para las filas. Permite además la operación con estos datos de manera rápida y eficiente.

- **Matplotlib:** Es una librería para la generación de gráficas que surgen a base de arrays o listas definidas dentro del lenguaje de programación Python. Contiene variedad de gráficas y ayudan a la expresión de métricas de manera visual, funcionando de esta manera como un complemento para el resto de librerías cuyo centro principal es la manipulación directa con información en grandes volúmenes.
- **Seaborn:** Similar a Matplotlib, es una librería que se basa en la ayuda de representación visual de información en gráficas, el gráfico más usado al utilizar esta librería es el histograma.
- **Torch:** Es una librería open-source enfocada en el Deep Learning y aprendizaje automático que acelera el camino desde la creación de prototipos de investigación hasta el despliegue de las aplicaciones en un entorno de producción.
- **Transformers:** Proporciona APIs para descargar y entrenar fácilmente modelos preentrenados de última generación. El uso de estos puede ayudar en la reducción de costo computacional, la huella de carbono y ahorrarle tiempo en lugar de realizar un entrenamiento desde cero de un modelo que no ha sido entrenado con anterioridad.
- **SciKitLearn:** Es una librería de aprendizaje automático que soporta algoritmos de clasificación, regresión y clustering: SupportVector Machines, Random Forests, Gradient Boosting, K-means y DBSCAN). Está diseñada para el poder interactuar con las librerías Numpy y Scipy en el ámbito del aprendizaje profundo y automatizado.

VI. Resultados y Comparativa

Al hacer uso de dataset ya mencionado, la división que se realiza entre la data para el entrenamiento y la data que va a ser evaluada basándonos en el entrenamiento se encuentra en una proporción del 80 % dejando el resto de data para la evaluación de la efectividad del modelo. Para poder verificar la calidad y el éxito de este modelo se utilizarán métricas que permiten evaluar la efectividad de la propuesta de este trabajo:

- **F1-Score:** Es la medida de precisión que tiene un test, hace uso de métricas previas como lo son la precisión y la exhaustividad del modelo. Opera con los falsos negativos, falsos positivos, verdaderos positivos y verdaderos negativos. Suele ser empleado en la fase de prueba de algunos algoritmos, sobre todo en aquellos de clasificación (Ecuación 1)

$$F1 - Score = \frac{2 * precision * recall}{precision + recall} \quad (1)$$

- **Accuracy:** O precisión, mide la cantidad de predicciones correctas en relación con el total de predicciones realizadas (Ecuación 2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

- **Matriz de confusión:** Indica la cantidad de elementos que han sido clasificados en sus respectivas clases en forma de matriz. De allí su nombre.

Se ha definido el espacio de trabajo con un aproximado de 6000 registros, que tienen comentarios variados, pertenecientes a las tres clases que se han mencionado antes. Como cada comentario posee un puntaje que puede estar entre los números 1 a 5, se ha colocado estos indicadores numéricos en indicadores textuales, para poder hacer la tarea de clasificación más sencilla y basada enteramente en clases que puedan ser entendidas y no ambiguas:

- 1-2: Negativos
- 3: Neutrales
- 4-5: Positivos

La distribución de registros se visualiza en la Figura 3.

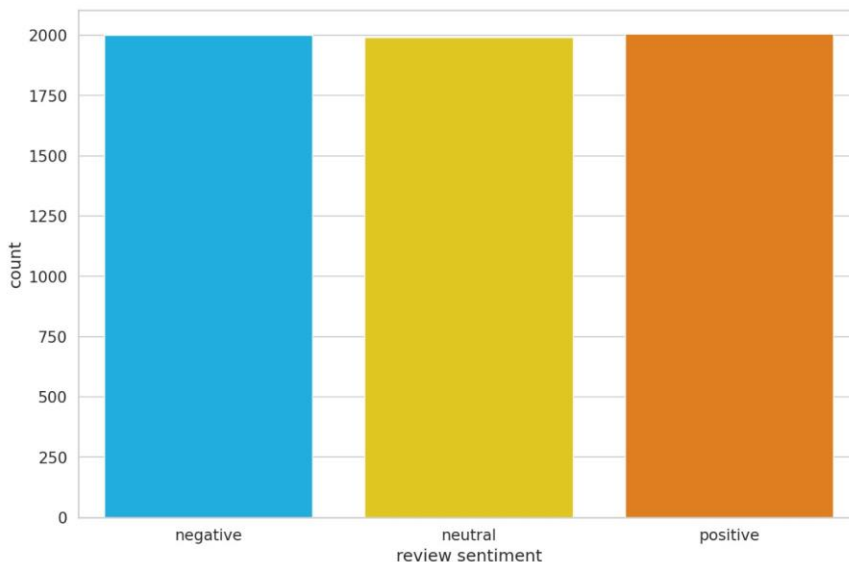


Figura 3, Distribución de registros en tres clases textuales.

Una vez establecidas las tres clases, se procede a hacer la medición del accuracy del modelo, cuyo valor numérico es de 0.73 o representado como el 73 %. Y finalmente se procede a calcular el F1-score que como se ha mencionado hace uso del accuracy y el recall, donde el puntaje es: 0.72 o representado como el 72 %. De la misma manera, la matriz de confusión presenta un alto índice de predicción correcta en aquellos comentarios neutrales, como se refleja en la Figura 4.

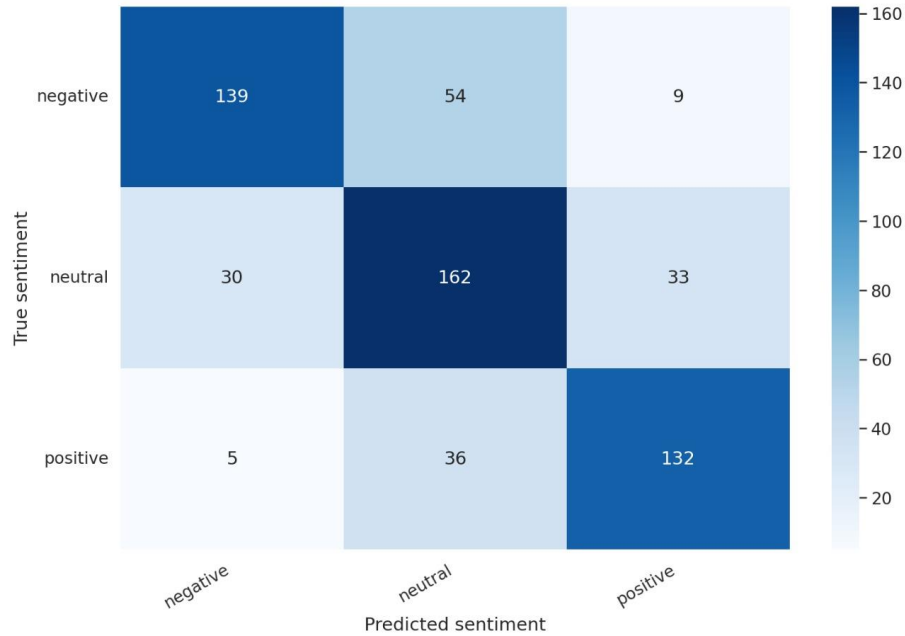


Figura 4, Matriz de confusión del modelo

Los datos obtenidos han sido comparados con [8], el cual nos muestra una tabla de resultados obtenidos usando dos métodos de extracción de caracteres, Term Frequency-Inverse Document Frequency (TFIDF) y vectores de documentos (Doc2vec), y usando métodos de clasificación como Random Random Forest (RF), Bootstrap Aggregating (Bagging), Gradient Boosting (GBT), Naïve Bayes (NB), Ridge Regression y Linear Support Vector Machine (SVC). Estos datos han sido comparados con nuestro resultado obtenido y la comparación se muestra en la Tabla 2.

Tabla 2. Resultados de diferentes tipos de clasificación.

Algoritmo	Accuracy	
TF-IDF	RF	71.49 %
	Bagging	71.57 %
	GBT	77.19 %
	NB	65.53 %
	Ridge	60.29 %
	SVC	64.61 %
Doc2Vec	RF	63.83 %
	Bagging	64.07 %
	GBT	69.23 %
	NB	59.58 %
	Ridge	69.84 %
	SVC	70.11 %
BERT	72 %	

Aplicación en los comentarios de Youtube

Una vez entrenado en modelo, procederemos a aplicar este mismo en la clasificación de comentarios, para la posterior clasificación de los videos, basándose en la cantidad de comentarios predominantes que posea un video. La predominancia de los comentarios ayudará a clasificar un video como útil, parcialmente útil y no útil, pues hablamos de tutoriales y estos deben de poseer utilidad que será aplicada por el resto de usuarios que miren el video. Para la obtención de los comentarios se ha usado una API de YouTube como ya se ha mencionado, esto nos ayudará gracias a la implementación que hemos realizado a que obtengamos específicamente solo los comentarios y analicemos la cantidad de sentimientos que estos poseen de trasfondo. Para esta prueba se han usado tres videos relacionados a la computación y que contienen variedad de comentarios que se encuentran dentro de las clases que hemos establecido con anterioridad. Los tres videos que vamos a usar son los siguientes:

- Video 1: Screensaver Not Working in Windows 10 FIX [Tutorial]
- Video 2: How to Enable keyboard in BIOS. 100 % working (HD)
- Video 3: Windows 11 Blue Screen Error Critical Process Died FIX [Complete Solution]
- Video 4: React Hooks Course - All React Hooks Explained

Cuyos resultados se encuentran reflejados en la Tabla 3, 4, 5 y 6.

Tabla 3. Resultados del Video 1

Tipo de comentario	Cantidad
Positivo	14
Negativo	16
Neutral	19

Tabla 4. Resultados del Video 2

Tipo de comentario	Cantidad
Positivo	10
Negativo	5
Neutral	7

Tabla 5. Resultados del Video 3

Tipo de comentario	Cantidad
Positivo	33
Negativo	30
Neutral	82

Tabla 6. Resultados del Video 4

Tipo de comentario	Cantidad
Positivo	267
Negativo	39
Neutral	195

Dejando a los videos con la siguiente clasificación:

- Video 1: Parcialmente útil.
- Video 2: Parcialmente útil.
- Video 3: Parcialmente útil.
- Video 4: Útil.

Conclusiones

Gracias al uso de herramientas que interactúan con el procesamiento del lenguaje natural como lo es BERT se pueden obtener los sentimientos involucrados en una serie de comentarios y que gracias a estos se pueden determinar la utilidad de un material audiovisual publicado en la red. Este trabajo se enfoca en la clasificación de comentarios haciendo uso del algoritmo BERT clasificando videotutoriales basándonos en la clasificación de sus comentarios. Con los resultados obtenidos, además de las métricas, encontramos una motivación para un desarrollo futuro de la posible futura de la propuesta brindada en este Trabajo utilizando nuevos métodos de clasificación o incluso mejorando el propuesto.

Trabajos Futuros

Dentro de la elaboración de este trabajo existen variedad de ideas que consideramos pueden ser integradas en propuestas futuras que mejoren la precisión de modelo planteado en este escrito. Este trabajo se enfoca principalmente en la

clasificación de videos sobre la base de sus comentarios, que colocan a los sentimientos en tres posibles clases. Las ideas que consideramos pueden ser agregadas en trabajos futuros son las siguientes:

- El uso de un dataset enfocado en comentarios neutrales, pues estos pueden parecer positivos al comienzo, pero su neutralidad puede afectar la clasificación de estos, dando resultados que puedan afectar la precisión del modelo.
- Enfoque del modelo en videos específicos, haciendo uso de términos especiales dentro de un contexto, por ejemplo: Iteraciones o referencias en un contexto de programación.
- El uso de más clases que puedan considerar los posibles giros que posean los comentarios neutrales.
- La forma en la que el modelo es construido puede ser diferente, pues en esta propuesta se hace uso de BERT, puede que se use otra herramienta para poder llevar a cabo la clasificación de comentarios de manera más efectiva y mejore las métricas, de la misma manera generando una matriz de confusión más grande pero más entendible y mucho más clara.

Referencias

- [1] Melissa Carvalho Costa Alexandre Ashade Lassance Cunha and Marco Aurelio C. Pacheco. Sentiment analysis of youtube video comments using deep neural networks. In *Lecture Notes in Computer Science*, pages 561–570, 2019.
- [2] Salman Aslam. Youtube by the numbers: Stats, demographics & fun facts. Omnicore, March 14, 2022, <https://www.omnicoreagency.com/youtubestatistics/>.
- [3] Hanif Bhuiyan, Jinat Ara, Rajon Bardhan, and Md Rashedul Islam. Retrieving youtube video by sentiment analysis on user comment. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 474–478. IEEE, 2017.
- [4] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. . O'Reilly Media, Inc.", 2009.
- [5] Marouane Birjali, Mohammed Kasri, and Abderrahim Beni-Hssane. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *KnowledgeBased Systems*, 226:107134, 2021.
- [6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

- [7] Navoneel Chakrabarty. A machine learning approach to comment toxicity classification. In Computational intelligence in pattern recognition, pages 183–193. Springer, 2020.
- [8] Jin Ding, Hailong Sun, Xu Wang, and Xudong Liu. Entity-level sentiment analysis of issue comments. In Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering, pages 7–13, 2018.
- [9] ENOIT DURAND. 500+ programming ytb comments. Kaggle, <https://www.kaggle.com/datasets/bdok /774-programming-ytb-commentsdataset/code?resource=download>.
- [10] Abbi Nizar Muhammad, Saiful Bukhori, and Priza Pandunata. Sentiment analysis of positive and negative of youtube comments using naïve bayes–support vector machine (nbsvm) classifier. In 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), pages 199–205. IEEE, 2019.
- [11] Adewale Obadimu, Esther Mead, Muhammad Nihal Hussain, and Nitin Agarwal. Identifying toxicity within youtube video comment. In International conference on social computing, Behavioral-cultural modeling and prediction and behavior representation in modeling and simulation, pages 214–223. Springer, 2019.
- [12] Ayushka Tiwari Ritika Singh. Youtube comments sentiment analysis. International Journal of Scientific Research in Engineering and Management (IJSREM), 2021.
- [13] Jalaj Thanaki. Python natural language processing. Packt Publishing Ltd, 2017.

Roles de Autoría

Valeria Alejandra Goyzueta Torres: Conceptualización, Análisis formal, Investigación, Metodología, Redacción - borrador original. **Ronald Fabricio Centeno Cardenas:** Curación de datos, Investigación, Metodología, Software, Validación, Redacción - borrador original. **Victor Andre Ranilla Coaguila:** Curación de datos, Investigación, Metodología, Software, Validación, Redacción - borrador original.