**Camilo A. Medina V.**

**Clase:** Ingeniería de Prompt y la Inteligencia Artificial

**Profesor:** Juan Carlos Correa

**Tema:** Implementación de herramientas de IA para el análisis de textos por medio de técnicas de ingeniería de prompt.

**Objetivo:** Desarrollar herramientas que permitan cuantificar y valorar retroalimentaciones escritas, por medio de la implementación de modelos de análisis de sentimientos y de IA generativa, con el fin de proporcionar a las empresas un método practico que sea útil al momento de evaluar las retroalimentaciones escritas de sus empleados.

**Introducción**

La idea de desarrollar un método de evaluación de retroalimentaciones escritas, nace debido a que en las empresas se llevan a cabo evaluaciones hacia los empleados por parte de distintos grupos (clientes, gerencia, pares, entre otros), y usualmente parte de la evaluación contiene un texto de retroalimentación de connotación cualitativa. A través de un análisis de sentimiento es posible cuantificar los textos asignando una puntuación especifica de acuerdo a la polaridad del texto, entre mayor positividad tenga mayor será la calificación. Lo anterior nos permite conocer, por ejemplo, cual es el estado actual de percepción de un cliente hacia un grupo determinado de empleados.

Además de la evaluación de sentimientos, es posible también hacer uso de la IA para extraer diferentes revelaciones (“insights”) que puedan dar un valor adicional a áreas de talento humano o gerenciales, en el presente trabajo por medio de la IA generativa se identificará sobre una lista de predefinida de competencias organizacionales[[1]](#footnote-1), que competencia organizacional esta ligada al texto de retroalimentación y una justificación de la misma.

**MODELOS ANÁLISIS DE SENTIMIENTOS**

Descripción base de datos usada:

3 columnas:

1. Feedback (STRING). Contiene una muestra de 113 retroalimentaciones.
2. Position (STRING). Describe el rol de la empresa.
3. Date (DATE). Fecha de las retroalimentaciones.
4. Label (STRING). Etiqueta que asigna a cada texto una polaridad (Positive, Negative, Neutral)

**Modelo Roberta**

Referencia: <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>

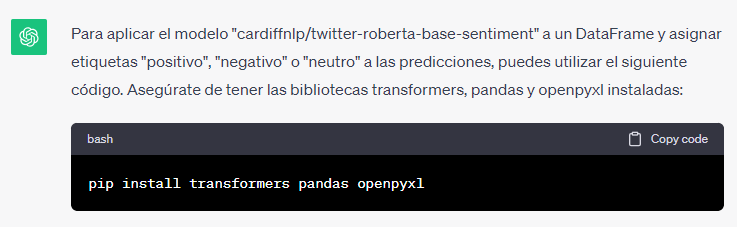
Es un modelo de Lenguaje Natural y su principal propósito es determinar la polaridad de un sentimiento entre positivo, negativo o neutral, fue entrenado con 58 millones de tweets.

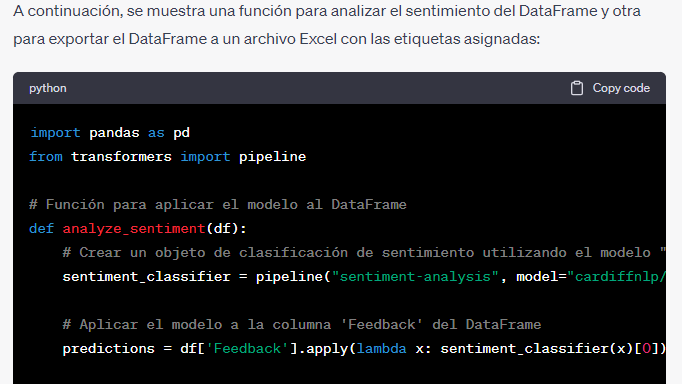
Este modelo permite identificar polaridades de frases o párrafos enteros identificando el contexto en que fue escrito, difiere de los modelos tipo diccionario los cuales determinan la polaridad de acuerdo a un etiquetado previo de la polaridad de cada palabra.

El código utilizado en la implementación del fue generado a través del envío de Prompt a chat GPT (motor Default (GPT-3.5)).

Chat GPT Prompt:

Dame una funcion python para aplicar el modelo "cardiffnlp/twitter-roberta-base-sentiment" a un dataframe df = pd.read\_excel('https://raw.githubusercontent.com/kmiloR2/Maestria/main/feedback-examples.xlsx'), las etiquetas tienen que ser "positivo", "negativo" o "neutro", dame también la función para exportar el df a un Excel.





**2nd Model**: Utilización de API OpenAI

Modelo pre-entrenado y tuneado con millones de registros, que permite a partir de proporcionarle un contexto y un “prompt” suministra una respuesta concreta.

En nuestro caso se utilizó para dar como respuesta un objeto json para ser exportado a un dataframe.

**Modelo BERT**

Referencia: <https://huggingface.co/blog/sentiment-analysis-python>

Existen alrededor de 215 modelos públicamente disponibles en Hugging Face Hub para realizar análisis de sentimiento, la comunidad ha desarrollado un “pipeline” que facilita la implementación de los modelos a través de 5 líneas de código:

pip install -q transformers

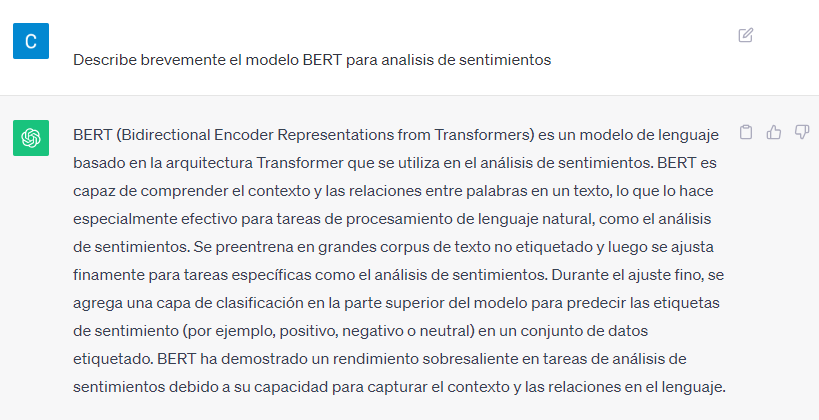
from transformers import pipeline

sentiment\_pipeline = pipeline("sentiment-analysis")

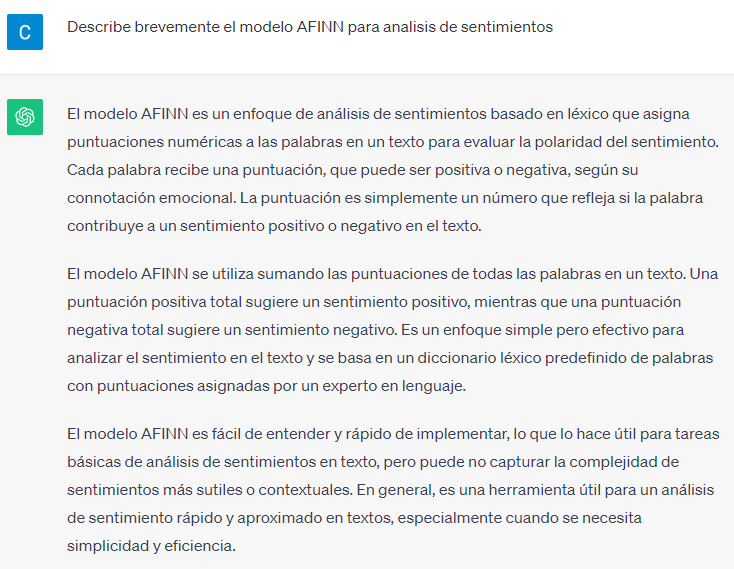
data = ["I love you", "I hate you"]

sentiment\_pipeline(data)

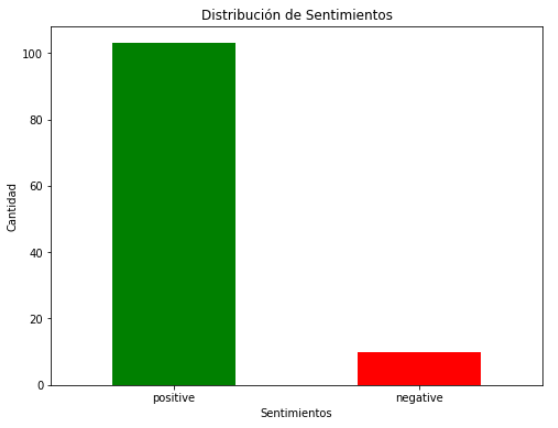
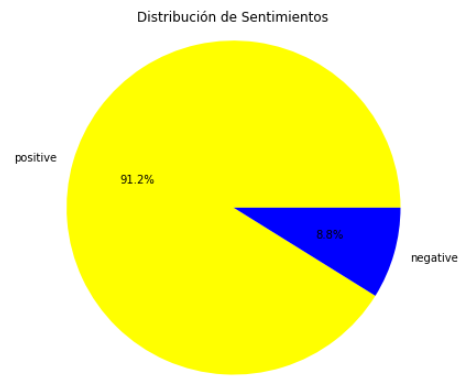
Para el presente trabajo hemos utilizado el modelo “finiteautomata/bertweet-base-sentiment-analysis”.



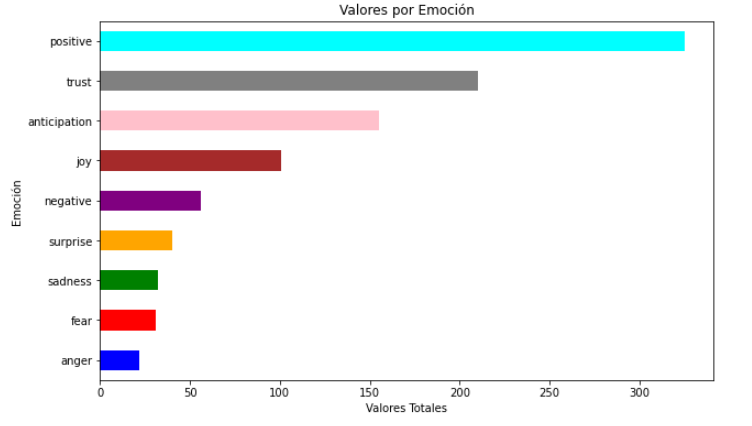
**Modelo AFINN**

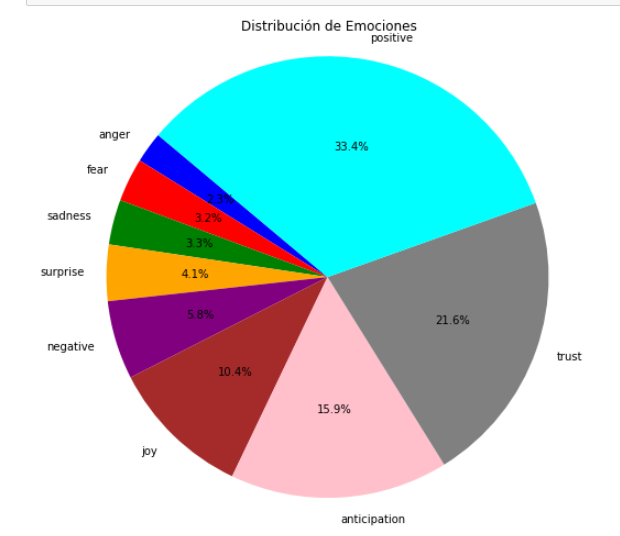


Este modelo tiene una funcionalidad adicional para extraer emociones (enojo, alegría, entre otros), lo que agrega valor al análisis. También incorpora la función "score" que nos da una puntuación numérica para cada texto según la polaridad, más palabras positivas más puntuación.

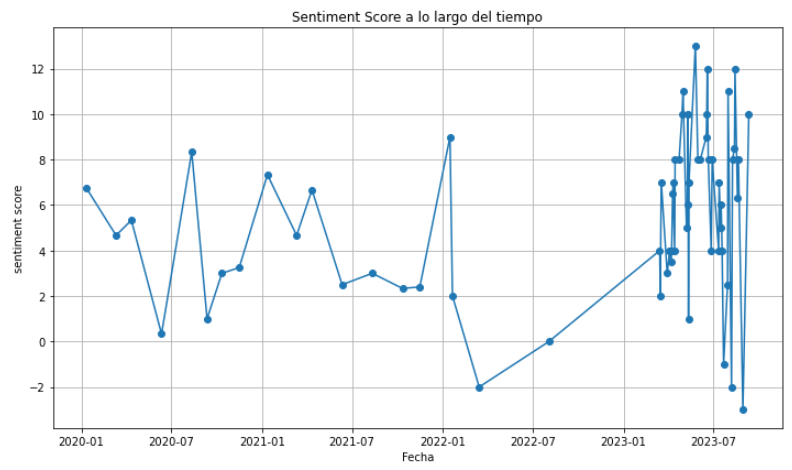


Los siguientes gráficos nos muestran las emociones predominantes del total de retroalimentaciones por conteo de palabras:



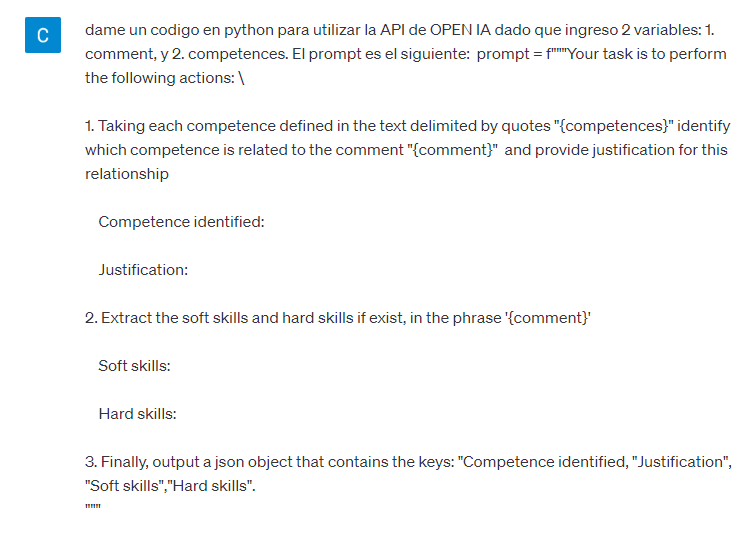


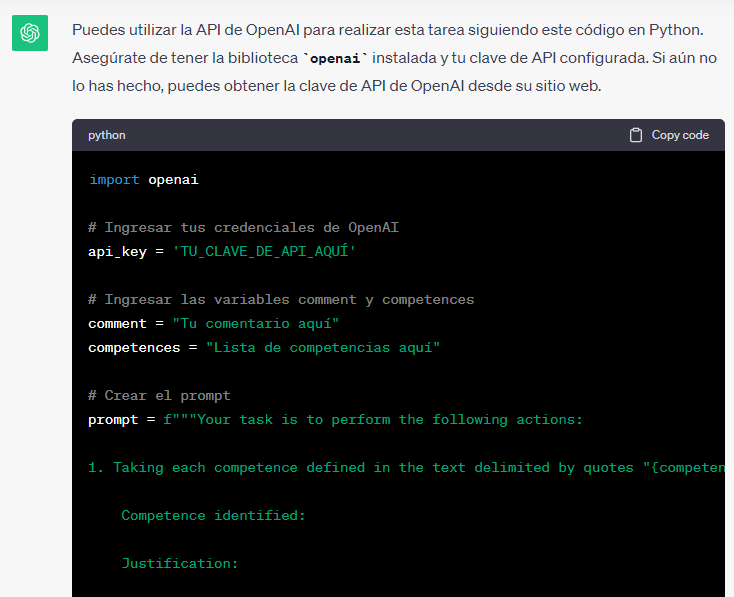
El siguiente gráfico nos muestra a lo largo del tiempo como se ha comportado el score de sentimientos:



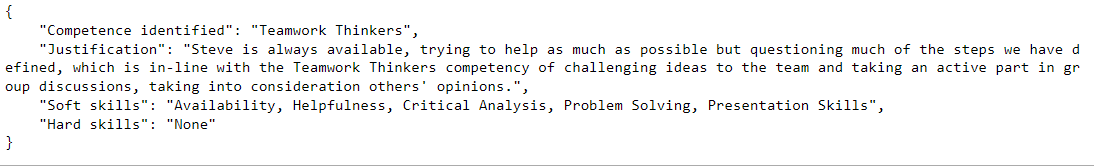
**Modelo OPEN IA**

Con la implementación de este modelo se busca ingresar las retroalimentaciones (variable comment) y a partir de ellas identificar la competencia organizacional relacionada y también la justificación de la misma, además de las principales Soft y Hard skills de cada individuo.





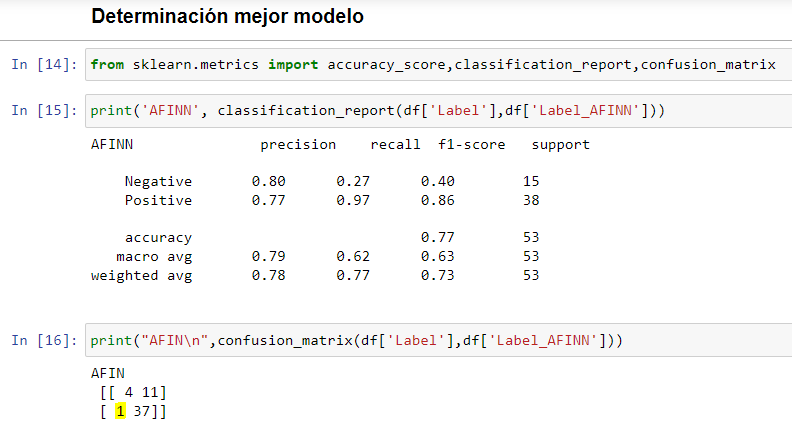


El código como resultado el siguiente formato:

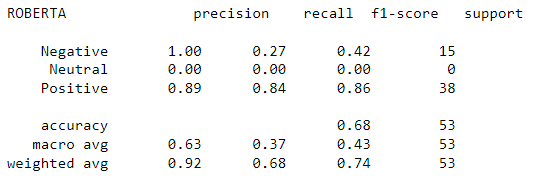
**Comparación entre modelos**

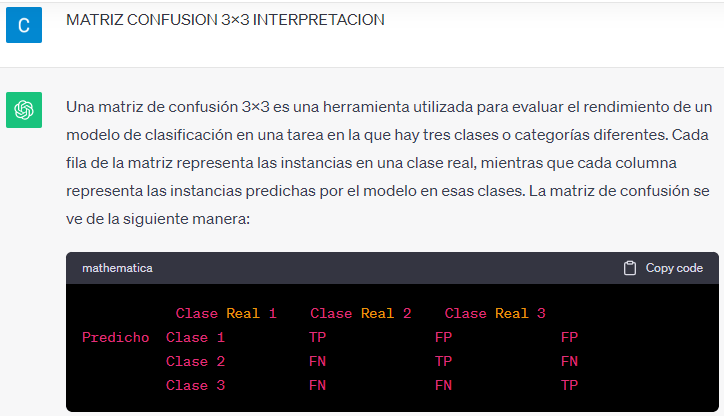
Para nuestro propósito en concreto se buscaría disminuir la cantidad de falsos negativos, dado que una retroalimentación negativa calculada erróneamente puede tener un mayor impacto en el momento de evaluar el desempeño de una persona.

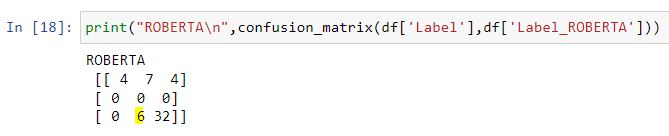
Por ello se ha realizado por medio del uso de la matriz de confusión la evaluación para determinar el mejor modelo.

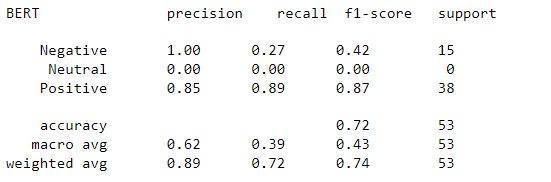


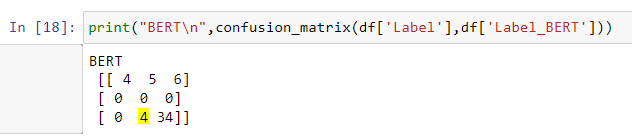
Para el modelo ROBERTA y BERT se generaría una matriz 3x3











Como conclusión el mejor que tiene menos errores en los falsos negativos sobre la etiqueta “Positve”, es el modelo AFINN.

**Conclusiones**

El uso de distintos modelos de Lenguaje Natural en combinación con IA generativa nos permitió sacarle un mayor provecho a los textos de retroalimentación de personas, proporcionando una mayor información que la que tuviera su sola lectura. Lo anterior se logró por medio de una evaluación cuantitativa de los textos, ya que al asignarles un valor numérico se puede lograr hacer comparaciones entre personas, evaluaciones de percepción de clientes y evaluación de rendimiento a lo largo del tiempo. Se consolida entonces como una herramienta para áreas de talento humano, clientes, gerencia u otros grupos interesados, que permite dar un mejor seguimiento tanto de las capacidades como el rendimiento para cada individuo o a nivel de área si se llegasen a agrupar los resultados. Además, la da como valor agregado el conocimiento de las competencias que puede tener cada persona.

Como trabajo futuro, los modelos pueden ajustarse a otros idiomas según sea la necesidad, para este trabajo los datos se procesaron únicamente con el idioma inglés. También es posible optimizar el prompt de la IA generativa para sacar otro tipo de revelaciones (insights) de los textos, esto según el interés especifico de cada empresa o persona.

1. La lista de definiciones organizacionales corresponde a la empresa Publicis GD. [↑](#footnote-ref-1)