

Análisis de sentimientos orientado al estudio de la probabilidad de fuga

Javier Fernández Alonso

Estudiante de Ingeniería Informática en Gijón.

CONTACTO: j.fernandezalonso7@gmail.com / UO269427@uniovi.es



Cátedra TotalEnergies de Analítica de Datos e Inteligencia Artificial

Introducción

En una sociedad cambiante y cada vez más globalizada por el uso de internet, la competencia entre las compañías que ofertan servicios a los ciudadanos no hace más que ir en aumento.

Las estrategias para atraer clientes o retenerlos pasan muchas veces por lanzar ofertas con rebajas de precio o medidas similares pero el trato cercano está en una clara tendencia a la baja, siendo cada vez menor.

La fidelidad de los clientes no sólo depende de factores económicos, en una gran medida se ve afectada también por el trato que reciban, y esto es algo que se está perdiendo de vista.


Objetivos del estudio

Tomando esto como pretexto inicial, el análisis del sentimiento de los clientes en base a los comentarios de los mismos podría llegar a ser de vital importancia a la hora de mantener a un cliente con la empresa. Permitiría ajustar el trato al estado de ánimo que cada cliente muestre y añadir este factor a un modelo que estudie la probabilidad de fuga de los mismos.

Métodos


Tras una serie de pruebas, se decidió aplicar Transfer Learning para el desarrollo más avanzado y detallado del proyecto. Y, más concretamente, se ha tomado como punto de partida el modelo BERT de Google.

Este enfoque permite realizar en base a los comentarios un análisis de positividad/negatividad mediante Fine-Tuning. Permite incluso estudiar, utilizando modelos ya existentes, cuán representativos son sentimientos como la alegría, el enfado o la tristeza dentro de un comentario concreto.




POSITIVE

"Great service for an affordable price. We will definitely be booking again."



NEUTRAL

"Just booked two nights at this hotel."

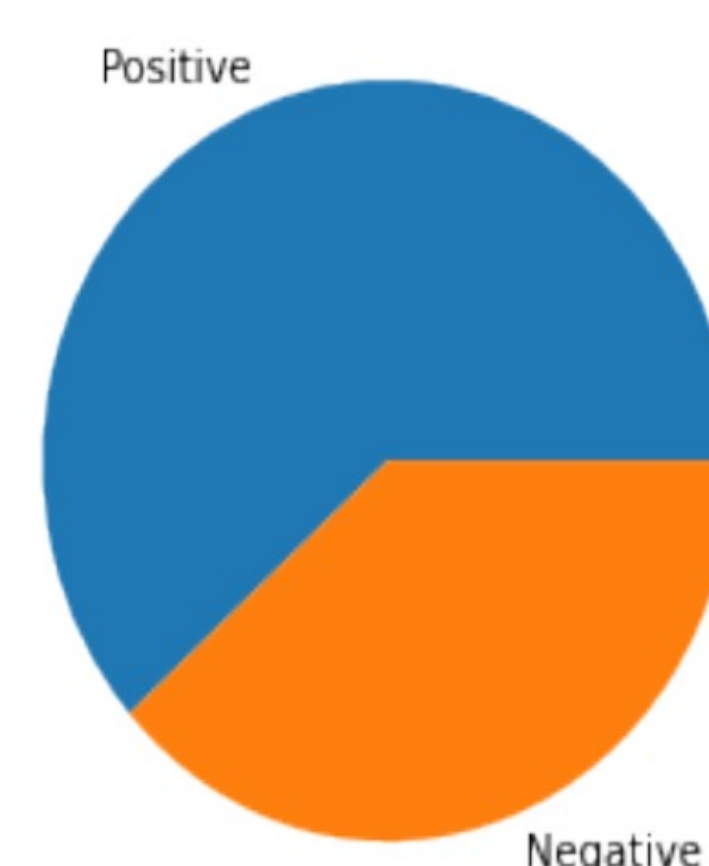


NEGATIVE

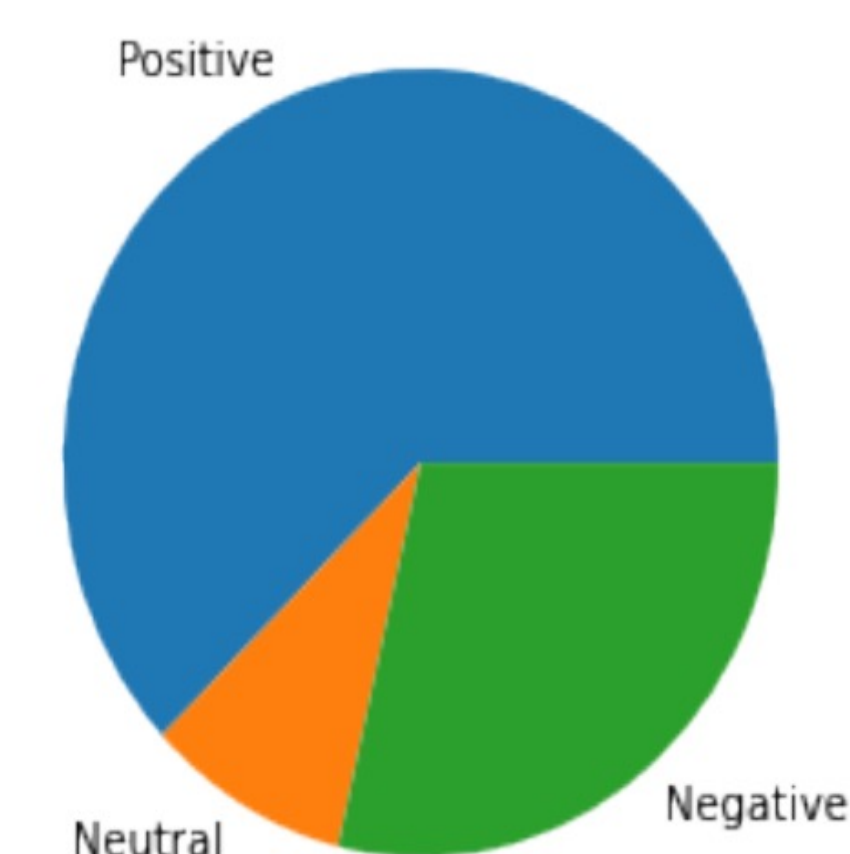
"Horrible services. The room was dirty and unpleasant. Not worth the money."

Figuras y Resultados

Values distribution for predicted data

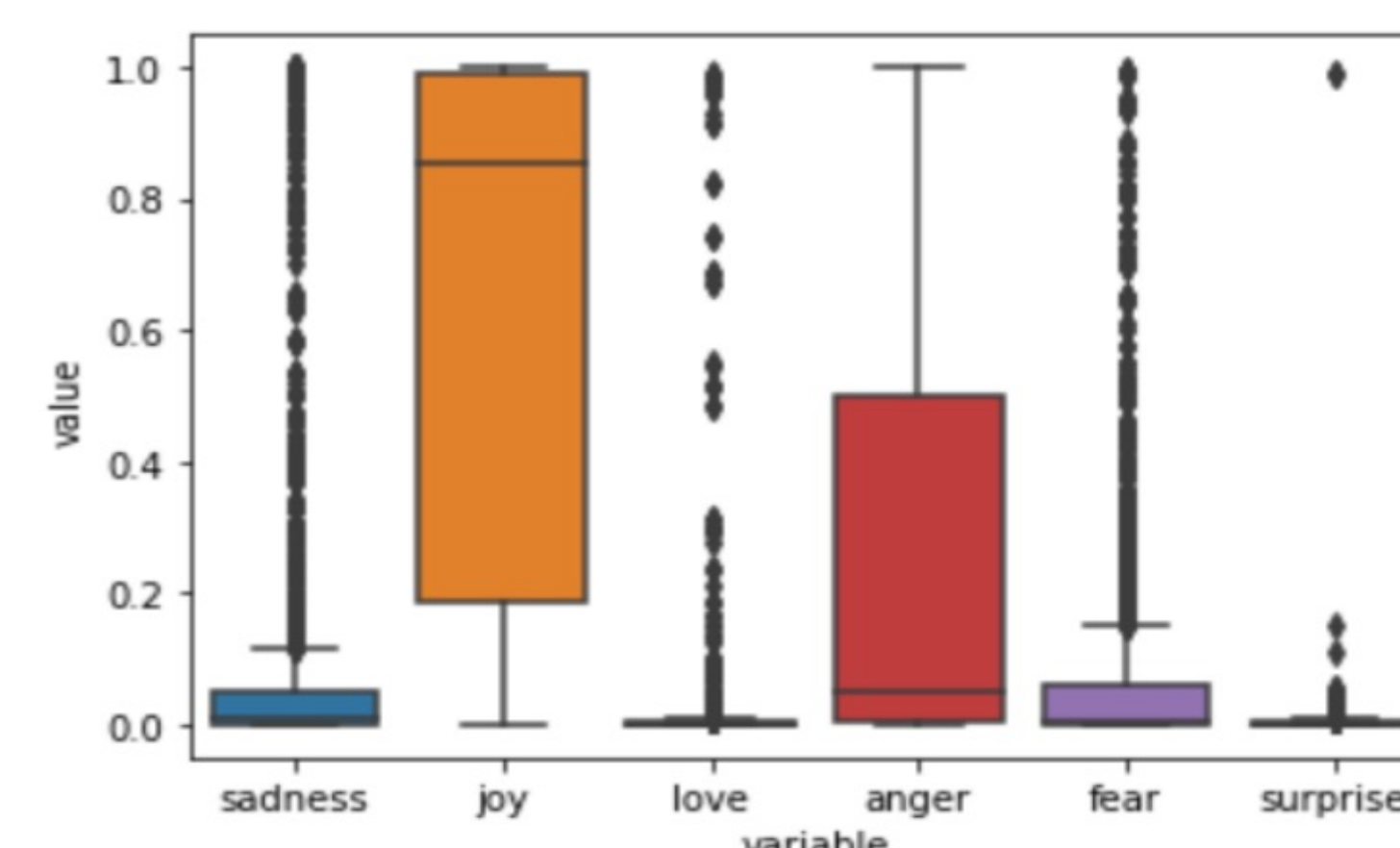


Values distribution for users' data



Siendo la referencia la gráfica de la derecha, cuyo criterio para asignar la positividad es que la nota dejada junto con el comentario sea mayor que 5, para la neutralidad que sea un 5, y para la negatividad que sea menor que 5, podemos ver que la estimación llevada a cabo mediante Fine-Tuning es bastante cercana a la realidad ya que presentan una distribución casi idéntica.

La precisión obtenida con las estimaciones es por tanto bastante elevada y podría incluso llegar a mejorarse utilizando un etiquetado de los datos propios o refinando más el modelo para proporcionarle un mayor nivel de contexto acerca del sector empresarial.



Texto	Tristeza	Alegría	Amor	Enfado	Miedo	Sorpresa
Comentario 1	0,0012	0,9925	0,001	0,0043	0,0006	0,0004
Comentario 2	0,0711	0,333	0,0036	0,528	0,0617	0,0027
Comentario 3	0,0132	0,0485	0,007	0,8315	0,0968	0,003
Comentario 4	0,0008	0,9966	0,0007	0,001	0,0007	0,0003
Comentario 5	0,0173	0,9708	0,0015	0,0052	0,0041	0,001

La gráfica muestra cómo se distribuye la puntuación de cada sentimiento en concreto a lo largo del corpus de comentarios con los que se contaba al realizar el estudio. La tabla, ejemplos concretos acerca de cómo se hace esa distribución. Esta alternativa podría permitirnos afinar más acerca de factores como la ironía o el sarcasmo, contrastando todas las emociones dentro de una frase y generando un resultado en base a ellas.

Conclusiones

Los resultados obtenidos hasta el punto actual de desarrollo del proyecto son muy interesantes y presentan un elevado potencial. El análisis del sentimiento de los textos, disciplina propia del procesamiento del lenguaje natural, en sí misma es un área con mucho potencial y en periodo de desarrollo.

Salta a la vista que, con los recursos existentes, hay muchísimos aspectos por explotar en el análisis de ciertos factores que pueden determinar de manera importante cuál es el estado de ánimo de un cliente con respecto a la empresa y, en general, qué sentimientos tienen estos clientes. ¿Cuánto influye realmente un comentario negativo? ¿Cómo de determinante es el sentimiento de un cliente de cara al riesgo de fuga del mismo? ¿Tienen suficiente contexto los algoritmos aplicados para el análisis?

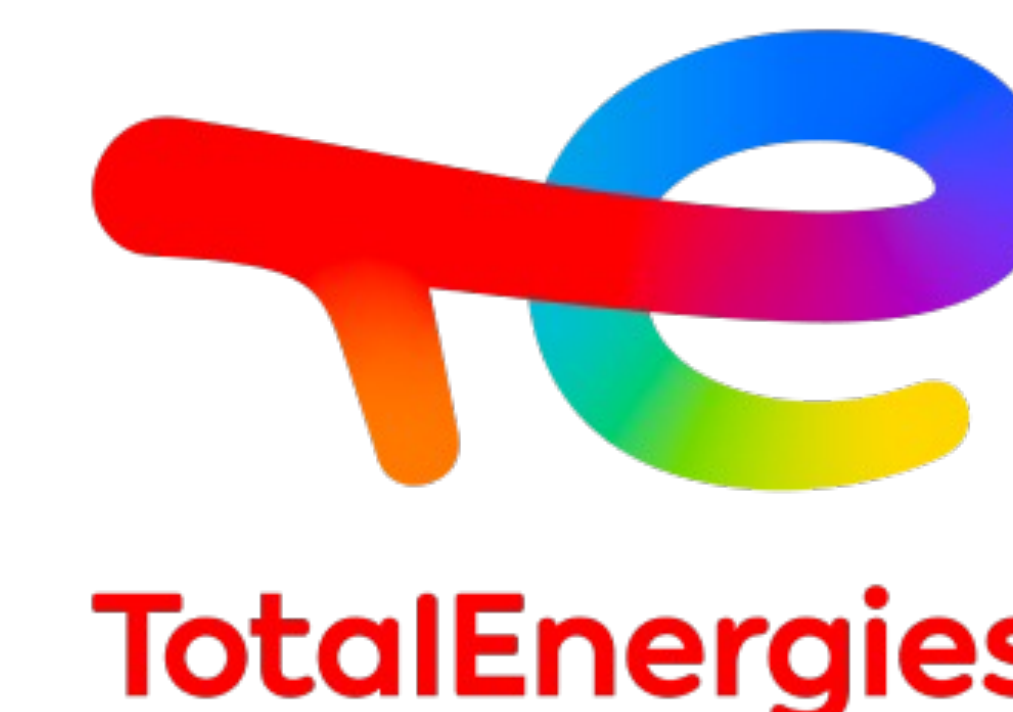
Lo que si sabemos con certeza es que estos resultados fruto del período de prácticas en TotalEnergies dejan a la vista un problema de lógica borrosa, con una serie de sentimientos asociados a los comentarios a la probabilidad de la recomendación de los clientes, acompañados de una valoración positiva o negativa de los mismos. Esto es el punto de partida fundamental a la hora de poder contestar a esas preguntas e ir más allá con el proyecto.

Esto nos deja por tanto ante un producto totalmente escalable que podría ser aplicado en diversas disciplinas adaptándolo a una serie de necesidades y, además, estaríamos planteando un posible sistema que supondría un enorme ahorro en recursos humanos para el análisis de la situación de los clientes.

Trabajo futuro

El estudio inicial se ha realizado tomando como base para el Fine-Tuning un modelo con datos de opiniones en IMDB, una plataforma acerca de películas y series. Aunque el uso de estos datos permite obtener unos resultados bastante certeros, carecen del contexto con el que sería ideal contar, por lo que sería adecuado o bien realizar un etiquetado de los datos propios o bien utilizar otros conjuntos de datos que permitan compensar esto.

El siguiente paso que se afrontará es la resolución del problema de lógica borrosa que plantean los resultados obtenidos a día de hoy. Con él, se generará una variable para el entorno que permita estimar la probabilidad de fuga de un cliente.



Universidad de Oviedo