

Análisis de Sentimientos

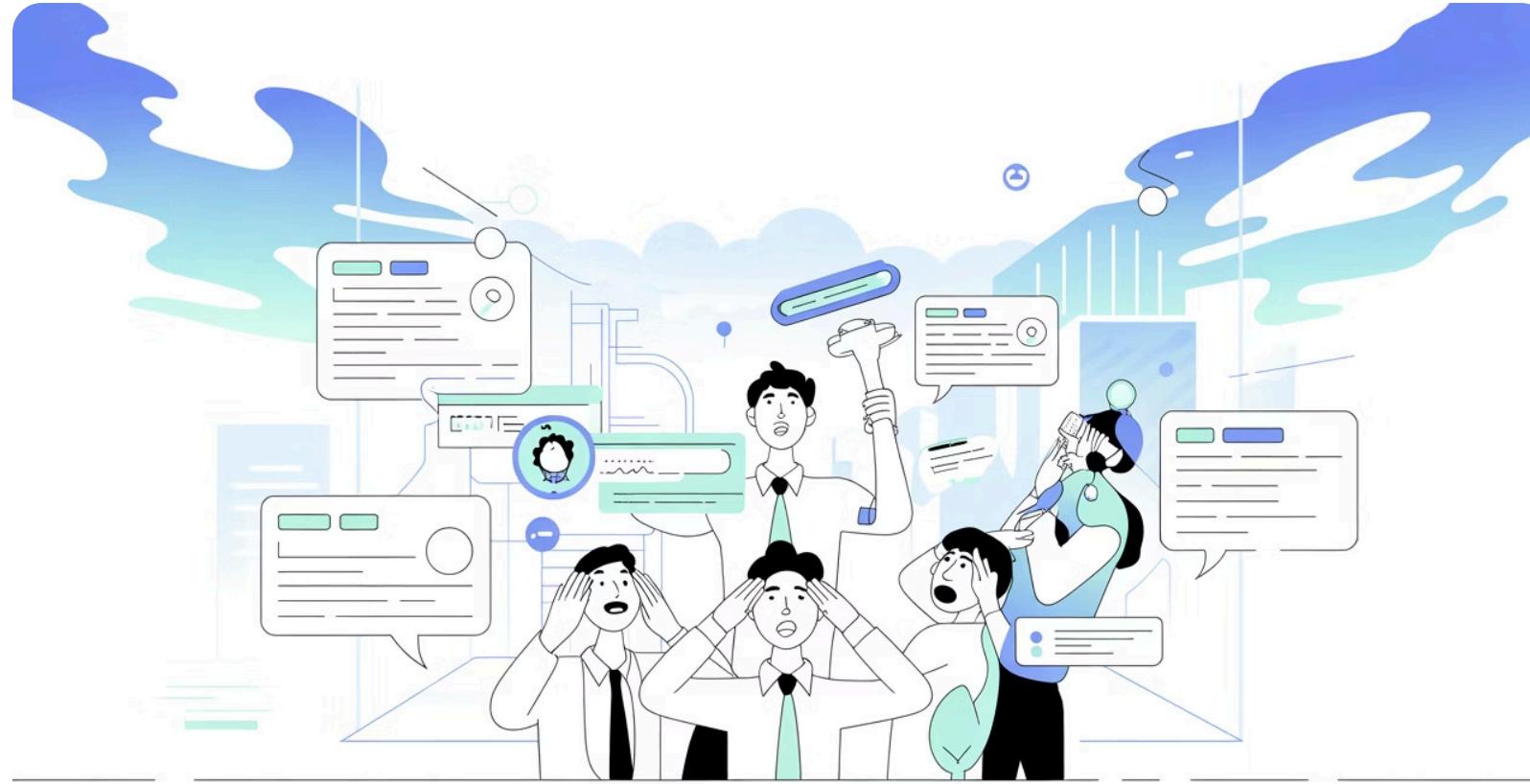


Entendiendo la Voz del Cliente en Google Play

Francisco V. | Proyecto Final de Machine Learning

El Problema y la Oportunidad

Millones de reseñas de usuarios se publican diariamente. Es imposible para un equipo humano leerlas todas. Se pierden quejas valiosas, elogios importantes y feedback crucial.



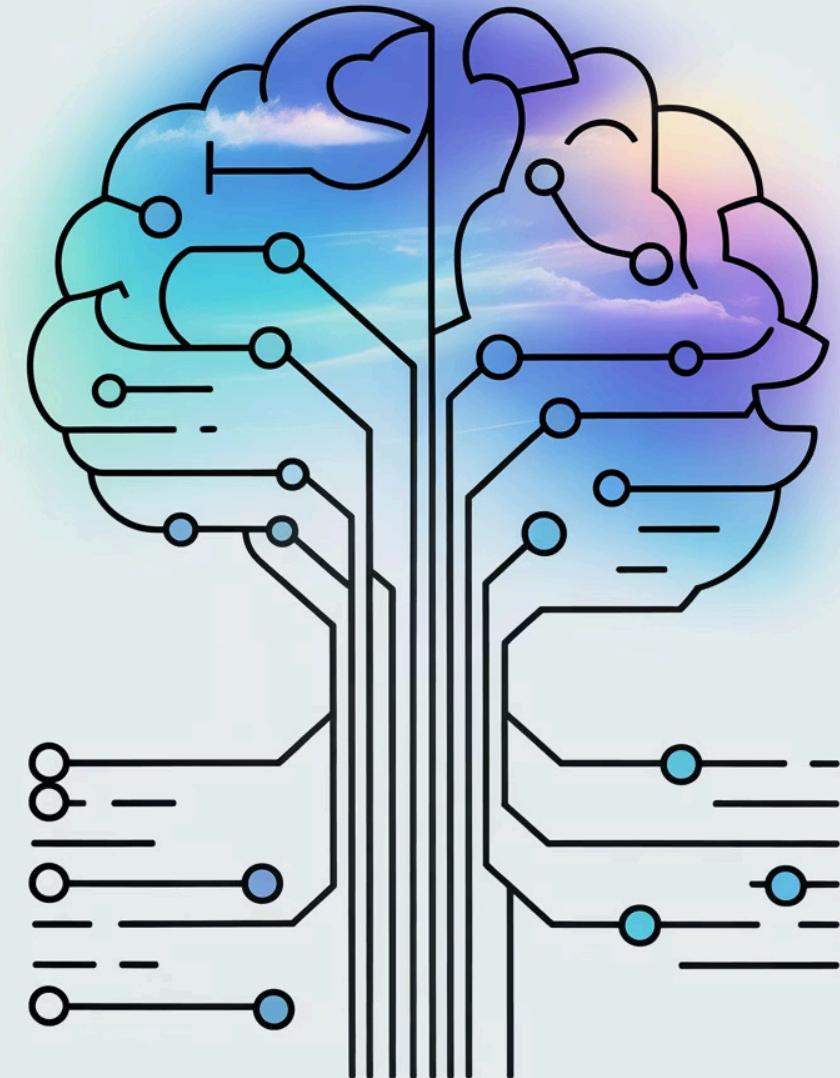
La Oportunidad

¿Cómo podemos "escuchar" a todos los usuarios a la vez?

¿Podemos automatizar la clasificación de millones de comentarios para tomar decisiones de negocio en tiempo real?

La Solución: Una IA que Lee por Ti

Construimos un modelo de Machine Learning que lee y clasifica automáticamente cada reseña como **Positiva**, **Negativa** o **Neutral**, permitiendo un análisis instantáneo de la opinión del cliente.





Explorando los Datos (EDA)

¿De qué Hablan los Usuarios?

El primer paso fue explorar los datos (EDA) para entenderlos.

Generamos una Nube de Palabras para visualizar los términos más frecuentes que la gente usa en sus reseñas. Esto nos da una idea clara de los temas más comunes y las palabras clave en el conjunto de datos.

Pude percatarme que las palabras mas populares dentro de los comentarios, son precisas hacia el sentir del publico, como por ejemplo: Good, great, love, work y really

¿Cómo Funciona? (Parte 1: El Proceso)



1. Limpieza de Texto

Primero, "limpiamos" el texto. Eliminamos puntuación, convertimos todo a minúsculas y quitamos palabras irrelevantes. Esto deja solo la información esencial para el modelo.

2. Vectorización (TF-IDF)

La IA no lee palabras, lee números. Usamos una técnica llamada **TF-IDF** para convertir las palabras limpias en valores numéricos. Esto le da más peso a palabras importantes (como 'excelente' o 'falla') que a palabras comunes (como 'y' o 'el').

¿Cómo Funciona? (Parte 2: El Modelo)

El Modelo (Random Forest)

Usamos un "Bosque Aleatorio". Este modelo funciona creando cientos de pequeños "árboles de decisión".

Cada árbol "vota" por un sentimiento (Positivo, Negativo, Neutral). El modelo elige la opción más votada.

Validación

Separamos los datos en conjuntos de "entrenamiento" y "prueba" para asegurar que el modelo pudiera predecir con precisión reseñas que nunca había visto antes.

Ajuste Fino (Tuning)

No nos conformamos con la primera versión. Realizamos un "ajuste de hiperparámetros" (GridSearchCV) para calibrar el modelo y encontrar la combinación de ajustes que nos diera la máxima precisión posible.



Resultados: ¿Qué tan Preciso Es?

91,2%

Official (Accuracy)

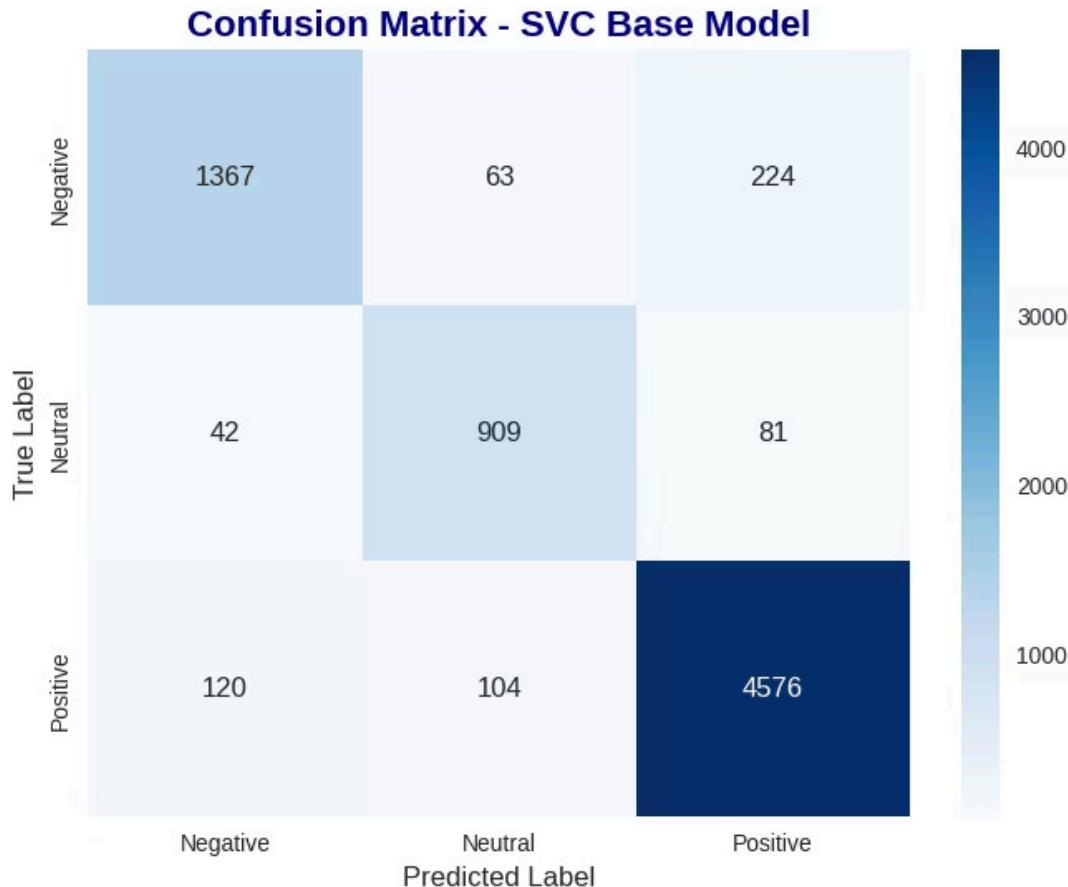
¡9 de cada 10 Veces Acierta!

El resultado del ajuste fino fue un modelo con una precisión global del **91.2%**. Esto significa que es altamente confiable para identificar la opinión de los usuarios.

También obtuvimos un **F1-Score (Macro) de 88.58%**, lo que confirma que el modelo es robusto y balanceado al clasificar las tres categorías (Positivo, Negativo y Neutral).

Resultados: ¿Dónde se Confunde?

La Matriz de Confusión



Esta gráfica es el "mapa" de aciertos y errores del modelo. La línea diagonal (de arriba-izquierda a abajo-derecha) muestra todos los aciertos.

Podemos ver que el modelo es extremadamente bueno identificando reseñas Positivas (acertó 4800 veces).

Esta matriz de confusión muestra cómo el sistema ha clasificado correctamente o incorrectamente diferentes sentimientos en un conjunto de datos. La mayoría de las veces, acierta y clasifica bien los sentimientos, pero también comete algunos errores, confundiendo ciertos sentimientos positivos, negativos o neutrales entre sí.

El Modelo en Acción: La API



Entrada (JSON)

La API está lista para recibir una reseña en inglés (ya que así fue entrenado).

```
{ "review": "This app is excellent!" }
```

Salida (JSON)

Y devuelve la predicción y el porcentaje de confianza.

```
{ "prediction": "Positive", "confidence": "96.20%" }
```

En la Nube (Ngrok)

Esta API está desplegada en internet y lista para ser consultada desde cualquier aplicación.

<https://flavory-adell-perinephral.ngrok-free.dev>
<https://flavory-adell-perinephral.ngrok-free.dev>
<https://flavory-adell-perinephral.ngrok-free.dev>
<https://flavory-adell-perinephral.ngrok-free.dev/>

Principales Aprendizajes

La Limpieza es Clave

El preprocessamiento y la limpieza del texto (EDA) son, sin duda, la parte más crítica del proyecto para lograr una buena precisión.

El Poder de TF-IDF

La vectorización no es magia, es una técnica poderosa para traducir la complejidad del lenguaje humano a matemáticas que el modelo puede entender.

Del Colab a la Realidad

Un modelo no sirve de nada si se queda en el notebook. El despliegue (serialización .pkl, Flask API, Ngrok) es el puente crucial hacia una aplicación real.



Conclusión y Siguientes Pasos

Hemos creado con éxito una herramienta funcional que convierte el "ruido" de miles de reseñas en información valiosa y accionable. El modelo no solo funciona, sino que es preciso (91.2%) y está listo para ser consumido (vía API).

- **Siguientes Pasos:** Re-entrenar el modelo periódicamente, probar arquitecturas más complejas (Redes Neuronales), y desplegar en un servidor permanente.

¿Preguntas?

Muchas gracias por su atención