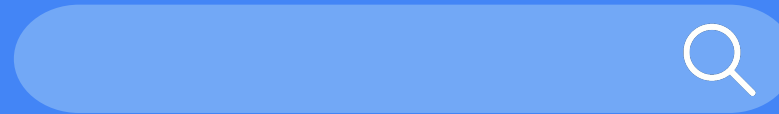




# Google Play



Bootcamp **UDD**  
Ciencia de datos e  
inteligencia artificial

Luis Garrido Cornejo



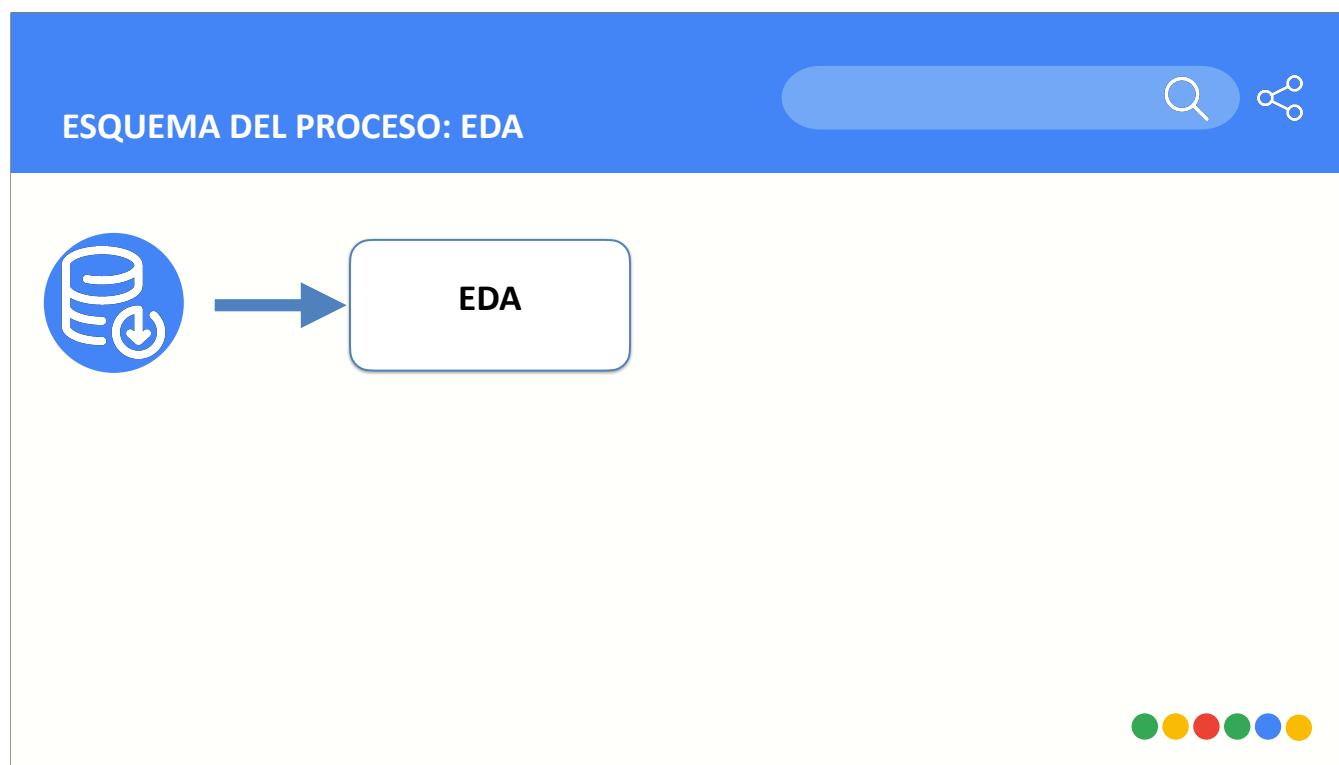
¿Te has preguntado qué opinan realmente los usuarios sobre las aplicaciones en Google Play Store?

¿Y si pudiéramos desentrañar las complejidades de esas opiniones para prever los sentimientos que generan?

En este proyecto de aprendizaje automático, exploraremos el vasto océano de reseñas de aplicaciones en Google Play para descifrar el tono emocional detrás de cada comentario.

La pregunta central que buscamos responder es:

¿Cómo podemos predecir con precisión el sentimiento de los usuarios hacia una aplicación en función de sus reseñas? ¡Acompáñanos en este viaje a través de los datos para descubrirlo!



Para este proyecto, se usó un set de datos de Google Play, que contiene entre varias características:

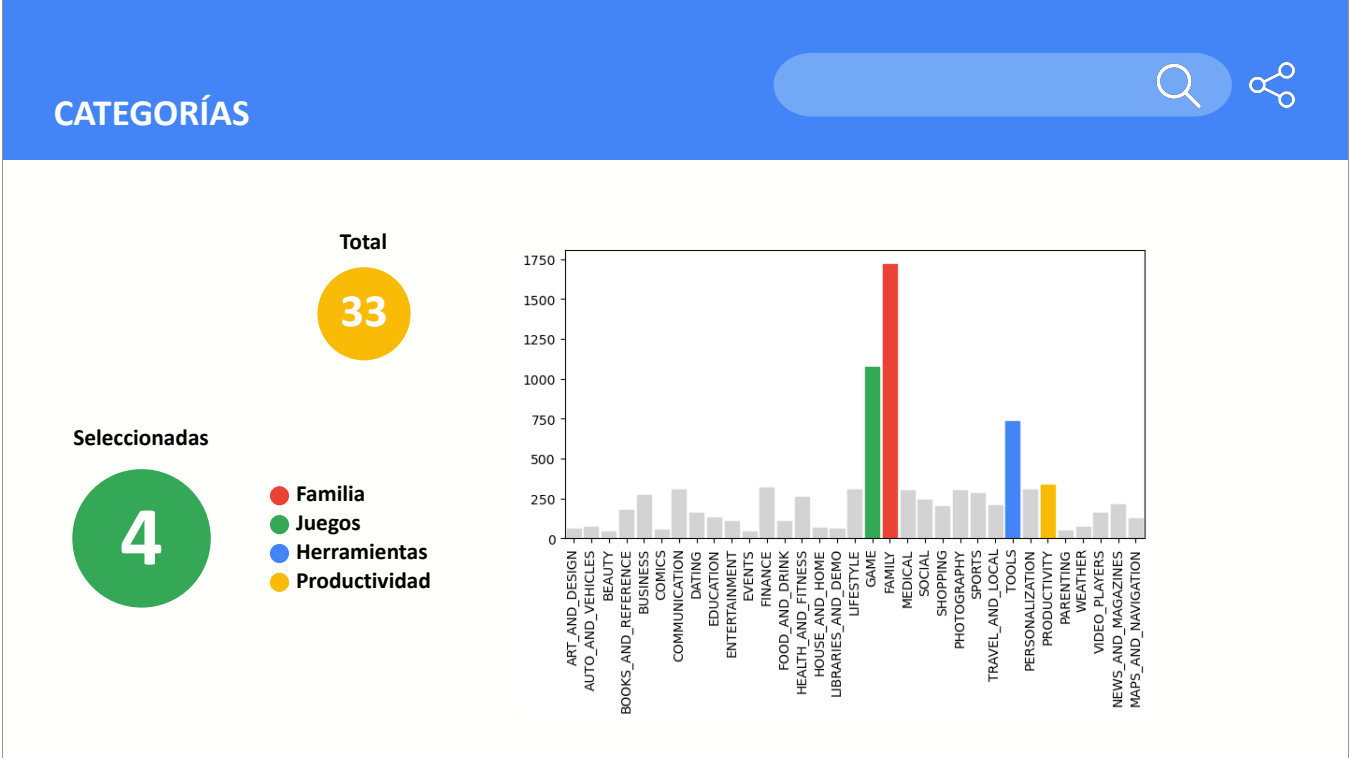
- El nombre de la aplicación (8190)
- La categoría a la cual pertenece (33)
- El rating promedio, las estrellas con las cuales los usuarios valoran dicha app
- La cantidad de comentarios que recibe aquella app
- Si es de pago o gratuita
- La calificación (6)
- El genero a la que pertenece (115)
- La versión de Android en la que se publicó



Análisis Exploratorio de Datos o simplemente EDA, es crucial para desarrollar modelos de aprendizaje

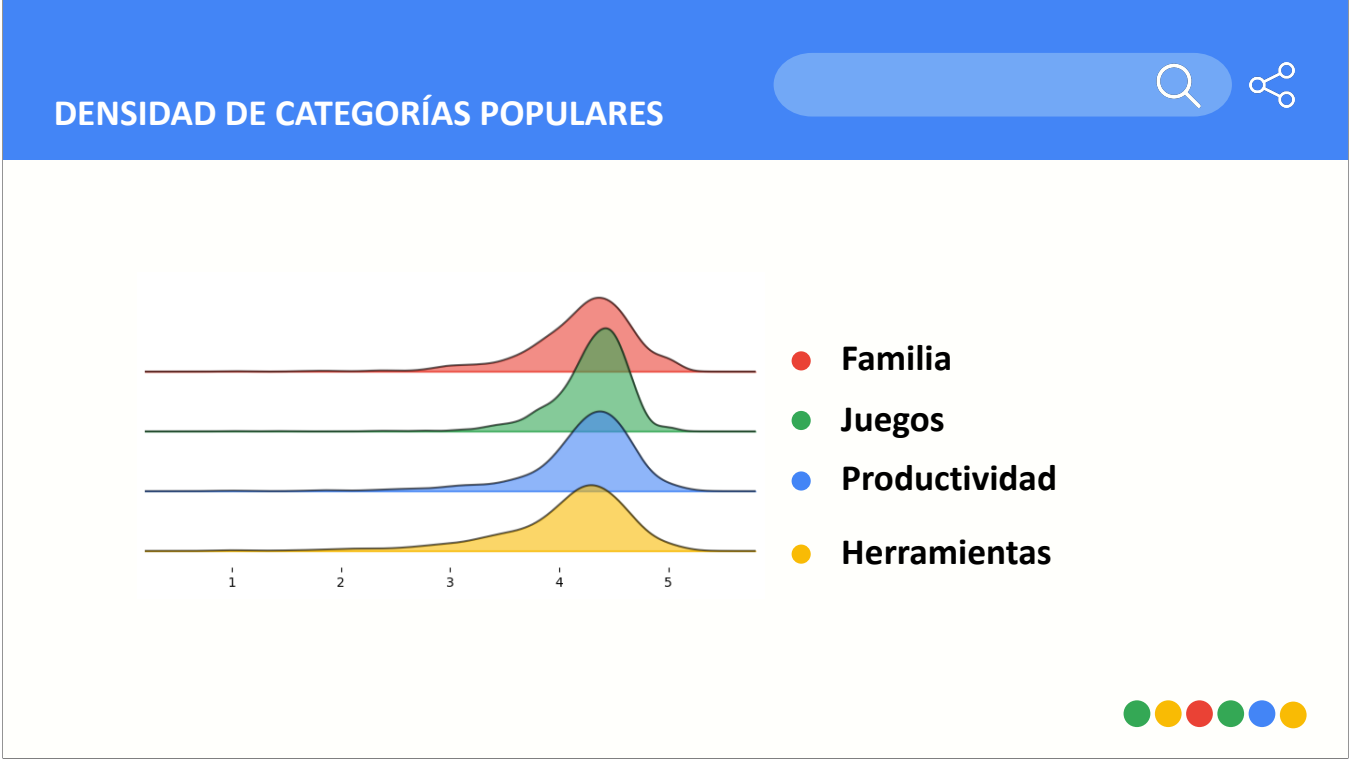
Conjunto de métodos y técnicas para analizar los datos de entrada. La elección depende de los datos y objetivos del análisis

- Visualización de datos, para identificar tendencias y patrones. Gráficos líneas, barras, dispersión, etc.
- Estadísticas descriptivas, como resumen importante sobre los datos, como media, median, moda, varianza, desviación estándar. Medidas de tendencia central.
- Tendencias y patrones, que pueden revelar información valiosa sobre los datos.
- Limpieza de datos,
- Tratamiento de datos numéricos (Standar Scaler, ) y categóricos (Label Encoder, One Hot Encoder, Dummies)



Visualización de datos





Visualización de datos

## APLICACIONES



1

### Mayor Valor

I'm Rich - Trump Edition  
LIFESTYLE  
\$400.0

251

### Versión Android

Jelly Bean  
4.1

49

### Más Descargadas

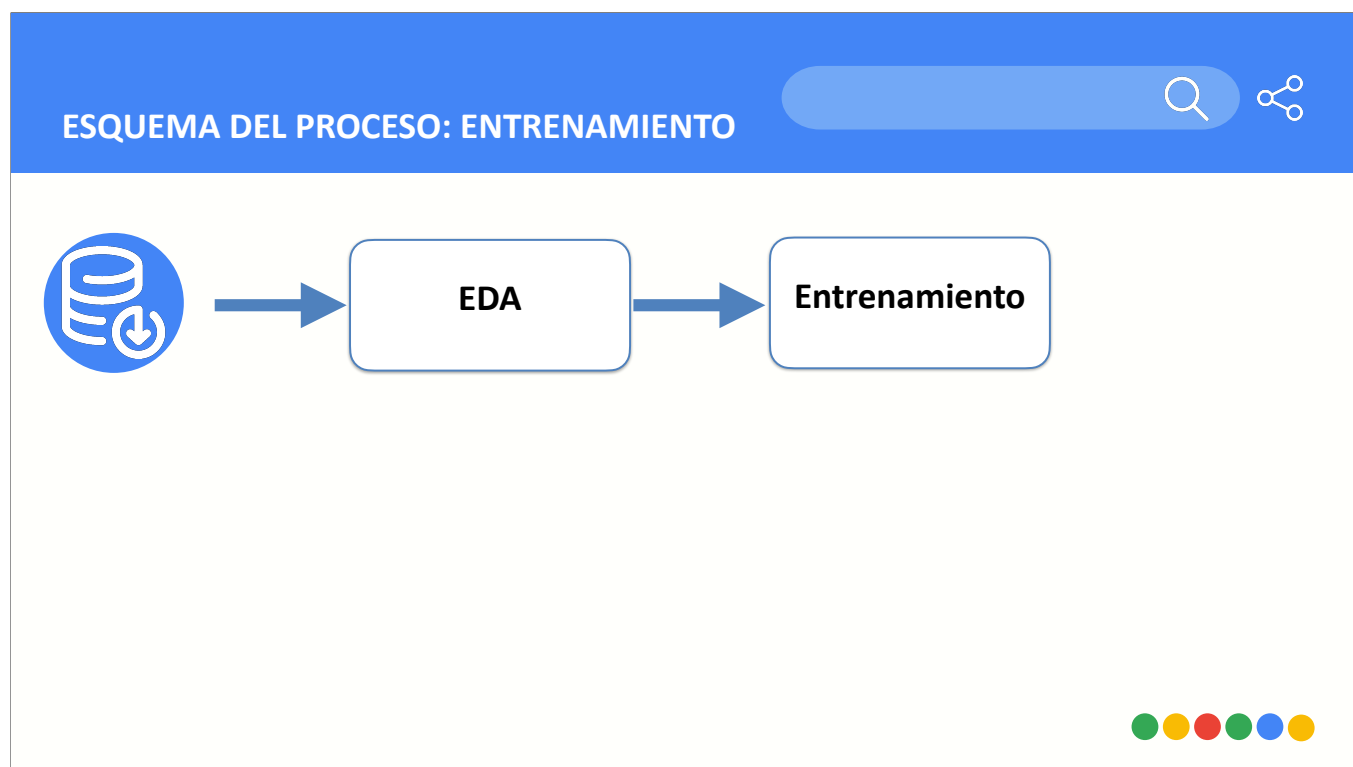
Más de mil millones

271

### Mayor Valoración

243 gratuitas  
28 de pago





En esta sección, se procede a usar otro set de datos que contiene:

- El nombre de la app
- El comentario de aquella app
- El sentimiento



Selección y optimización de modelos

Elegir el modelo más adecuado depende de la naturaleza de los datos y del problema a resolver.

Algunas opciones incluyen regresión lineal, árboles de decisión, Máquinas de Vector Soporte y redes neuronales.

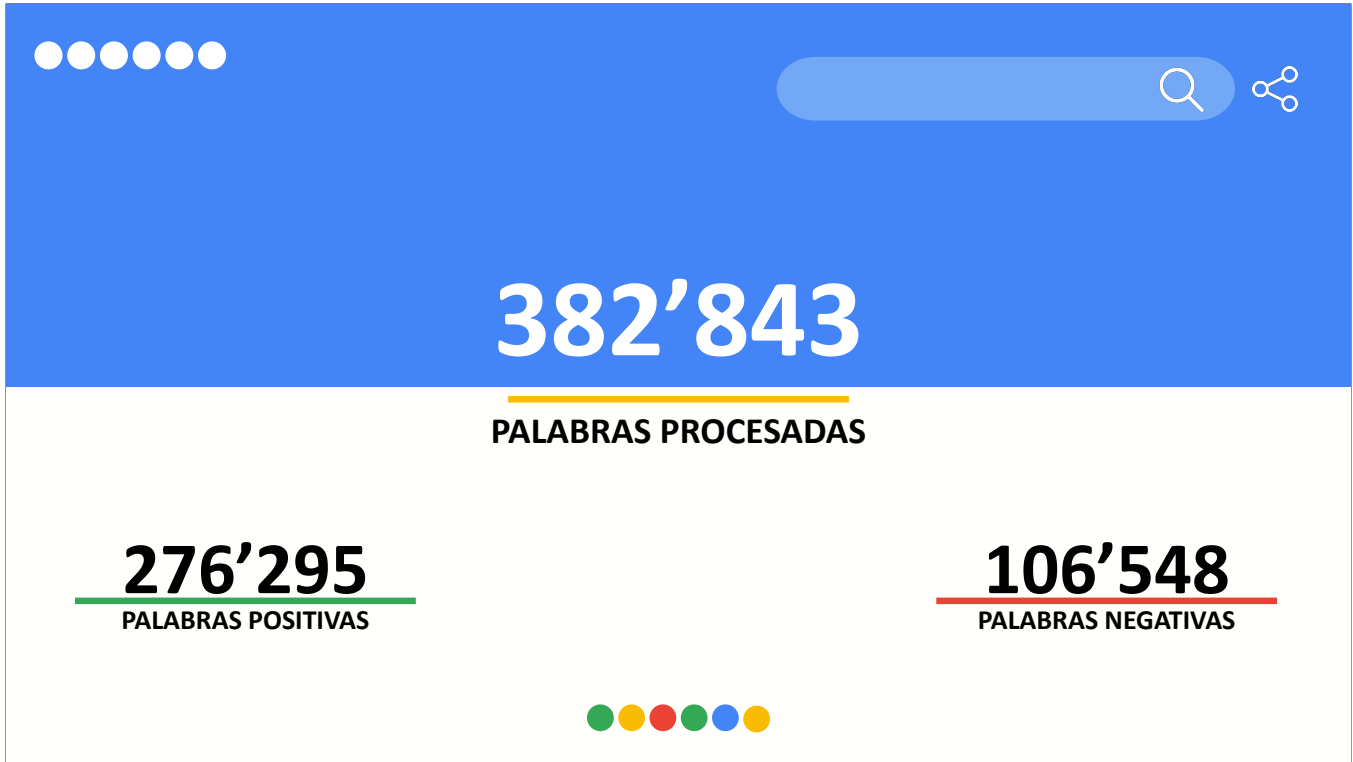
- Etiquetado de sentimientos
- Limpieza de texto (todo a minúsculas, eliminación basura, números, tokenización)
- Vectorización
- Entrenamiento del modelo
- Evaluación del modelo

Preparación de los datos, separación de los datos de entrenamiento y prueba

Elección del modelo

Evaluación y predicciones

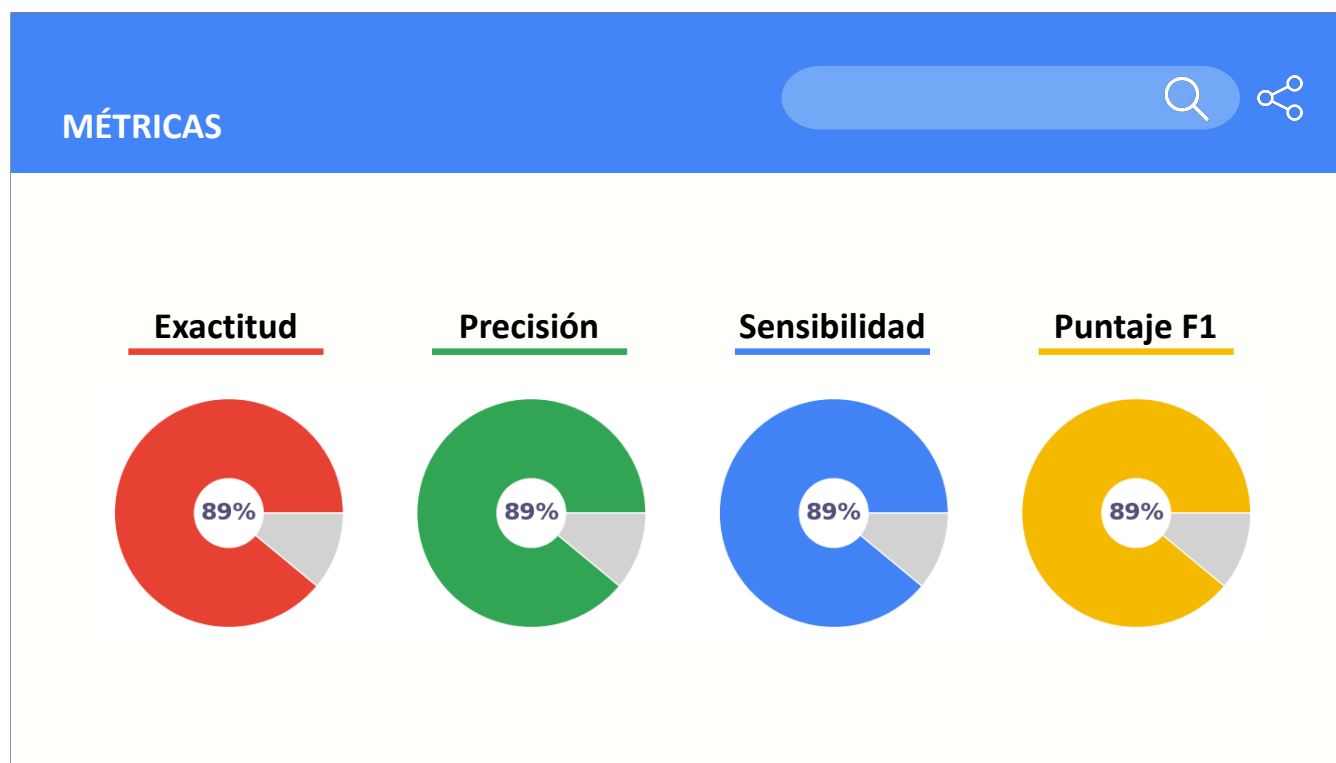






Nubes de palabras, es una herramienta útil para resumir los conceptos más importantes de un texto, una pagina web, un libro, o como en este caso las palabras de los comentarios que nos servirán para predecir los sentimientos de los comentarios.

Cuanto más presente esté una palabra en el texto, más grande aparecerá en la nube de palabras



Se utilizan diferentes métricas de rendimiento para evaluar modelos de clasificación

Exactitud: Es el número de predicciones correctas realizadas por el modelo por el número total de registros

Precisión: Con esta métrica estamos evaluando nuestros datos por su desempeño en predicciones **positivas**

Sensibilidad: También llamada Recall, es una medida que indica qué proporción de clase real fue predicha por el algoritmo

Puntaje F1: Es el promedio ponderado de la precisión y la sensibilidad, por lo tanto, esta puntuación representa las variables antes mencionadas

## MATRIZ DE CONFUSIÓN

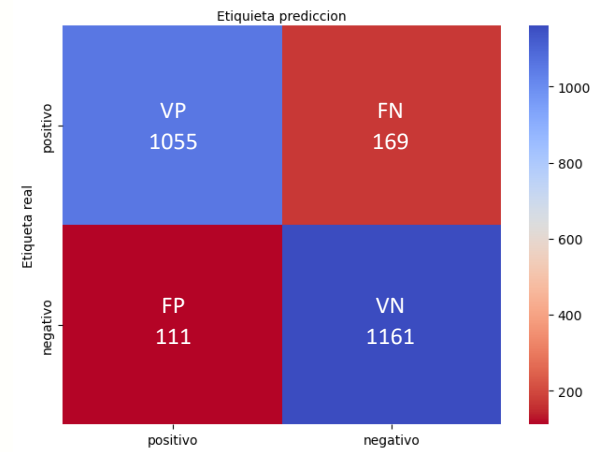


**Verdaderos positivos: 1055**

**Verdaderos negativos: 1161**

**Falsos positivos: 111**

**Falsos negativos: 169**



¿Qué es una matriz de confusión?

¿Por qué la necesitamos?

Es una herramienta que ayuda a entender si el modelo esta funcionando realmente.

Se basa en el hecho de que necesitamos comparar la **clase predicha** por el clasificador con la **clase real**

Verdaderos positivos: La clase predicha es la misma que la clase real

Falsos negativos: La clase predicha es negativa y no coincide con la clase real

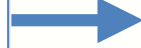
Falsos positivos: La clase predicha es positiva, pero la clase real es negativa

Verdaderos negativos: la clase predicha y la real son la misma, en este caso negativas

## ESQUEMA DEL PROCESO: PRODUCCIÓN



EDA



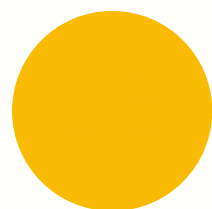
Entrenamiento



Producción







## Producción





CLASIFICADOR VECTOR DE SOPORTE

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.





1. Decisiones Informada:

Decisiones, como lanzamientos de nuevas versiones o funciones, pueden beneficiarse al considerar el sentimiento general de los usuarios.

2. Experiencia del Usuario - UX Mejorada:

Al abordar los problemas y las expectativas de los usuarios, se puede mejorar la experiencia general del usuario, lo que lleva a una retención más alta.

3. Priorización de Desarrollo:

Al comprender qué características o aspectos generan emociones positivas o negativas, los equipos de desarrollo pueden priorizar áreas específicas para mejorar o expandir.

4. Reputación de la Marca:

Monitoriar el sentimiento en las reseñas puede ayudar a gestionar la reputación de la marca. Identificar y abordar problemas de manera proactiva para mejorar la percepción general de la marca.

5. Mejora de Productos:

Se puede utilizar análisis de sentimientos para comprender cómo los usuarios perciben las aplicaciones y realizar mejoras específicas para abordar preocupaciones o críticas recurrentes.

6. Mejora en la Atención al Cliente:

Al integrar con sistemas de atención al cliente, la predicción de sentimientos ayudaría a priorizar automáticamente las consultas de los usuarios según la gravedad del sentimiento expresado.

Muchas  
**Gracias**

Referencias

**Kaggle** <https://www.kaggle.com>