



Proyecto 1

Nicolás Ortega - 201814515

Juan Diego Cardona - 201819447

María Camila Terán - 201822000



ÍNDICE

1

COMPRENSIÓN DEL NEGOCIO Y ENFOQUE ANALÍTICO

2

PERFILAMIENTO DE DATOS Y LIMPIEZA

3

MODELADO Y EVALUACIÓN

4

RESULTADOS

5

BIBLIOGRAFÍA

Comprensión del negocio



OPORTUNIDAD

Se busca clasificar aquellas emociones basándose en los diferentes textos que los usuarios de las redes sociales publican. También se busca conocer la reacción de las personas a ciertos eventos, productos o situaciones para facilitar a los negocios optimizar campañas de mercadeo.



DESCRIPCIÓN DESDE EL PUNTO DE VISTA DE APRENDIZAJE DE LA MÁQUINA

Textos con etiquetas para implementar algoritmos de aprendizaje supervisado.



Perfilamiento de datos



IDIOMA TEXTO



Inglés



INFORMACIÓN



mensaje y sentimiento



TENIENDO EN CUENTA ESTO, ¿QUÉ TÉCNICAS SE UTILIZARON?



TOKENIZACIÓN

Eliminar del texto todo aquello que no aporte información relevante para luego dividir las palabras de los mensajes.



STOP WORDS

Excluir las palabras más utilizadas del idioma para evaluar las más relevantes.



FRECUENCIA DE TÉRMINOS

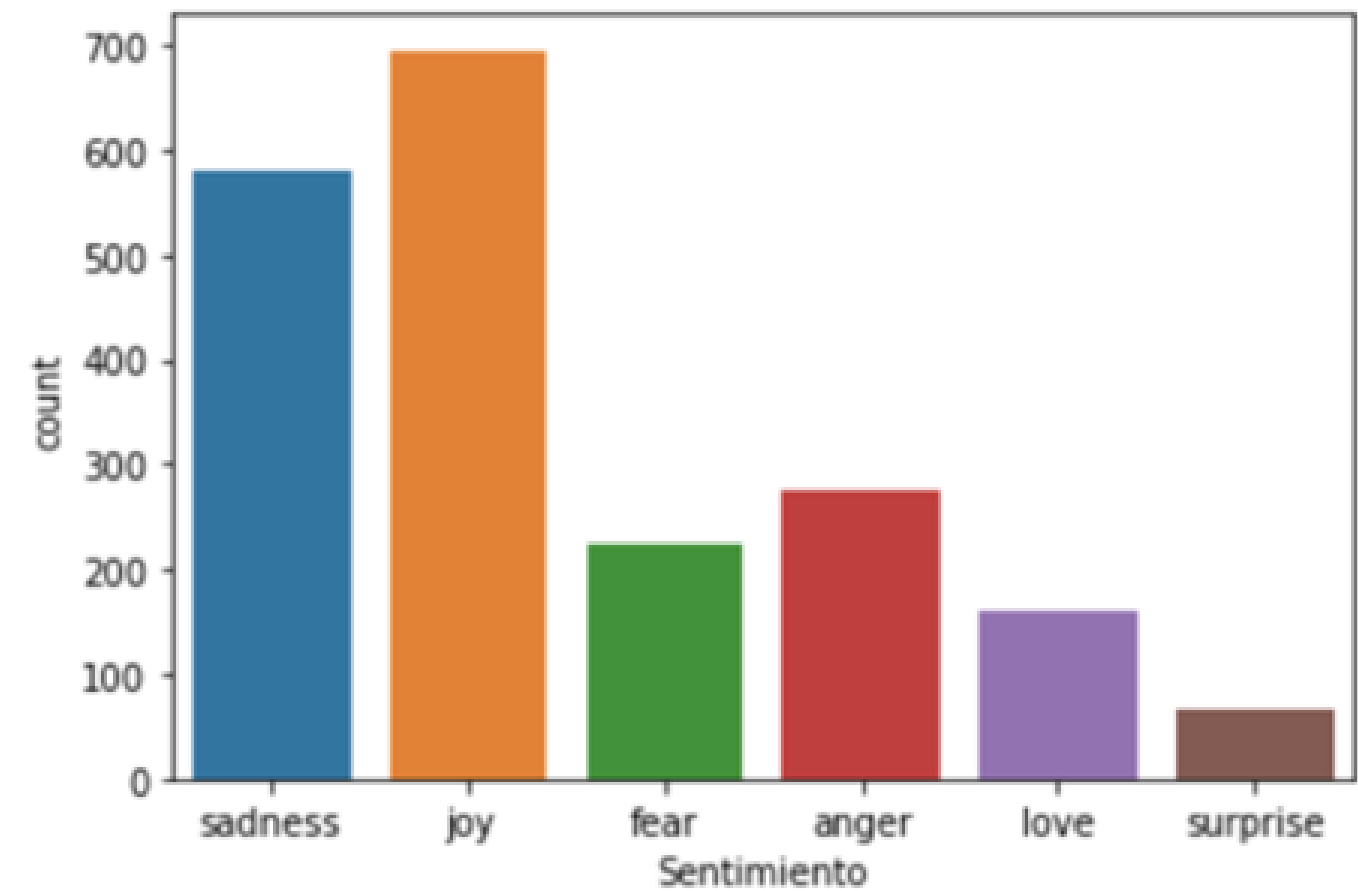
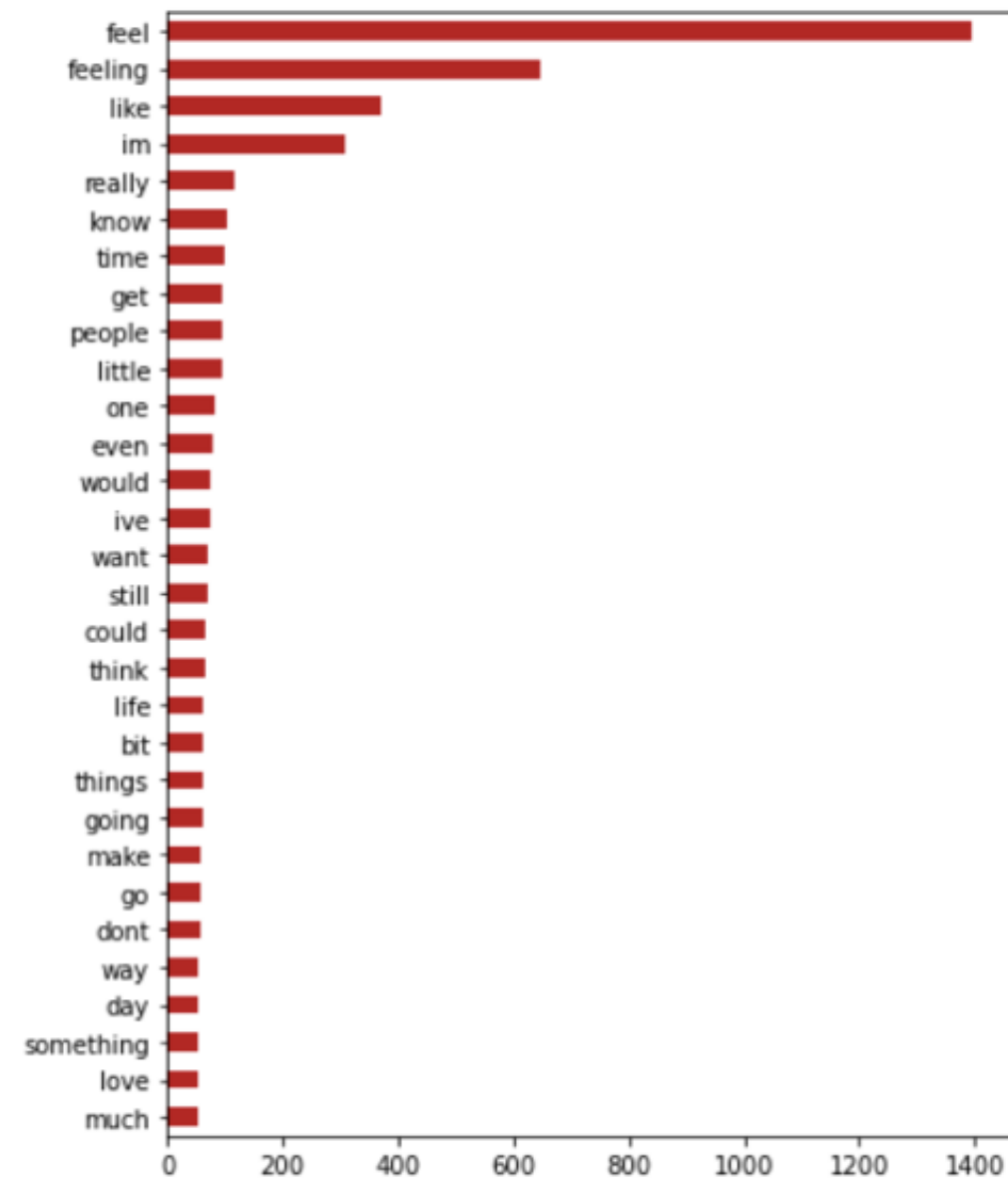
Medir la importancia de un término dentro del texto.



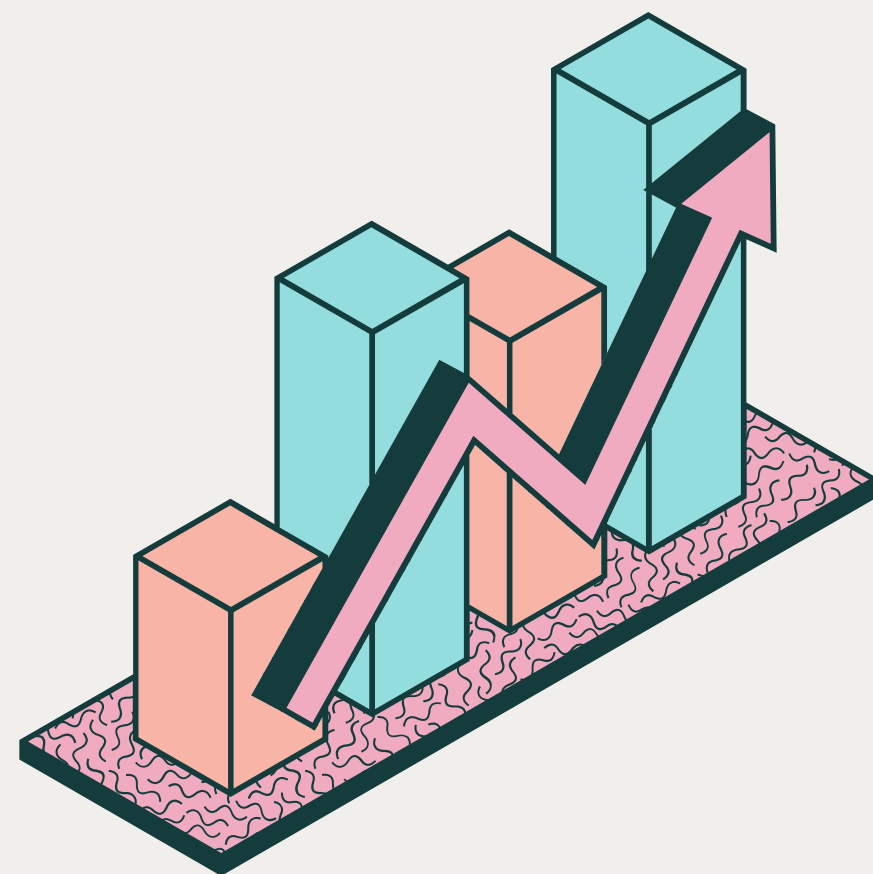
VECTORIZACIÓN

Crear una matriz de con las palabras como filas, el documento en columnas y el estadístico de la frecuencia en la intersección.

Perfilamiento de datos

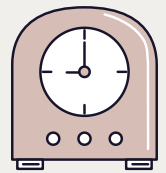


Modelado y evaluación



Modelo SVM lineal

Algoritmo de clasificación que correlaciona datos a un espacio de características de grandes dimensiones.



EXACTITUD

0.93



PRECISIÓN

0.9304



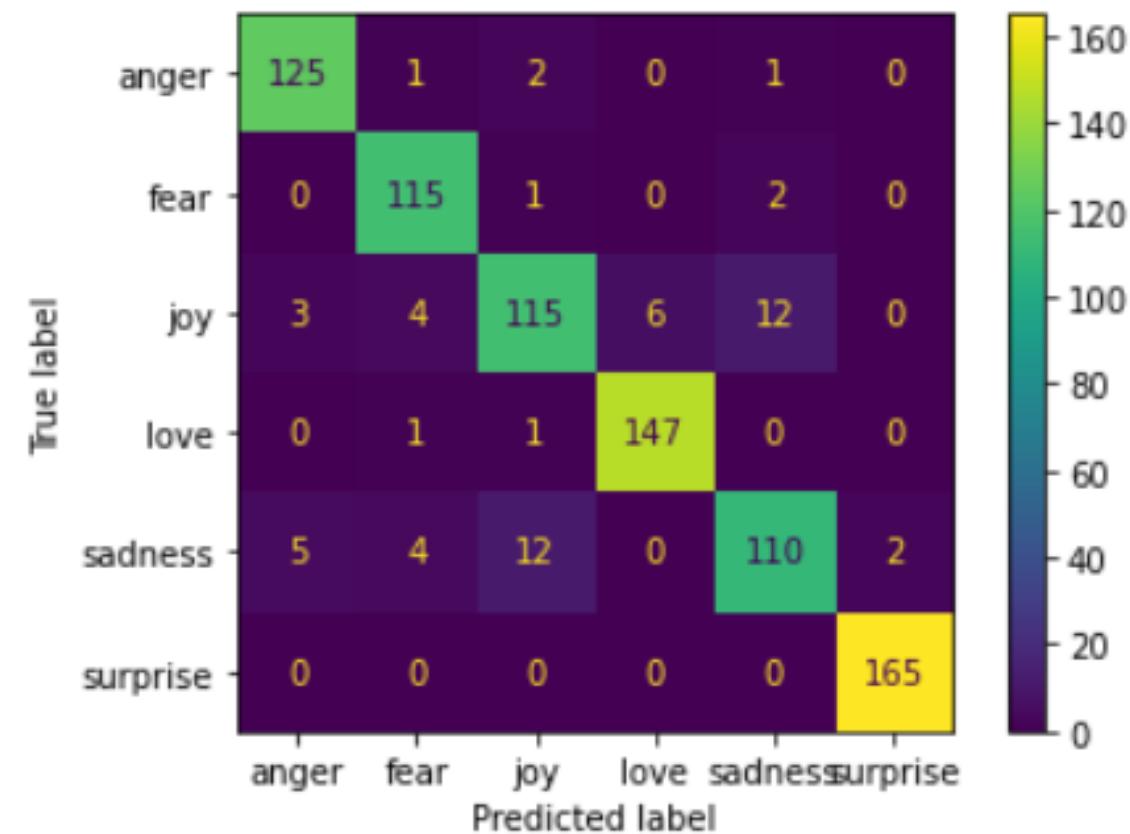
RECALL

0.9316



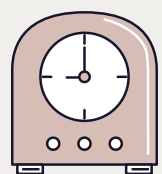
PUNTUACIÓN F1

0.9305



Modelo árboles de decisión

Funciona como un diagrama de flujo, separando los puntos de datos en dos categorías similares a la vez, desde el "tronco del árbol" hasta las "ramas" y las "hojas".



EXACTITUD

0.51



PRECISIÓN

0.5658



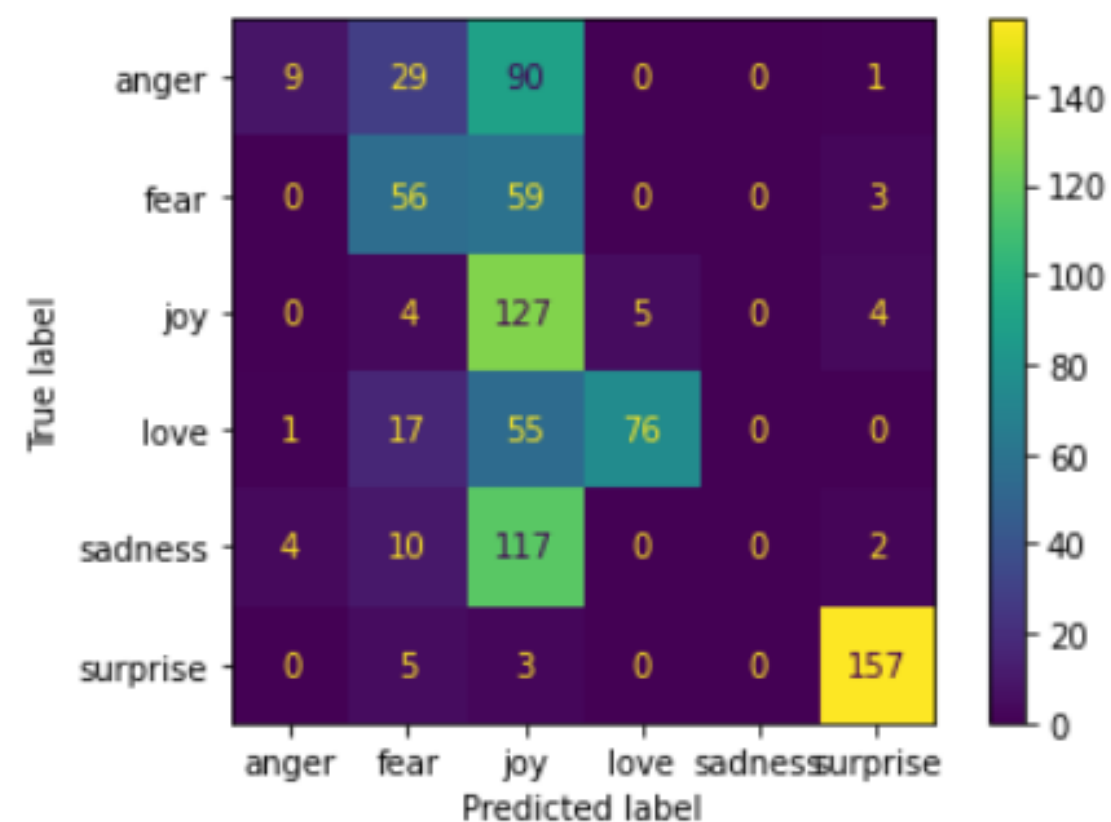
RECALL

0.5096



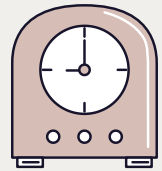
PUNTUACIÓN F1

0.4631



Modelo Random Forrest

Crea árboles de decisión sobre muestras de datos seleccionadas al azar, obtiene predicciones de cada árbol y selecciona la mejor solución mediante votación.



EXACTITUD

0.87



PRECISIÓN

0.8769



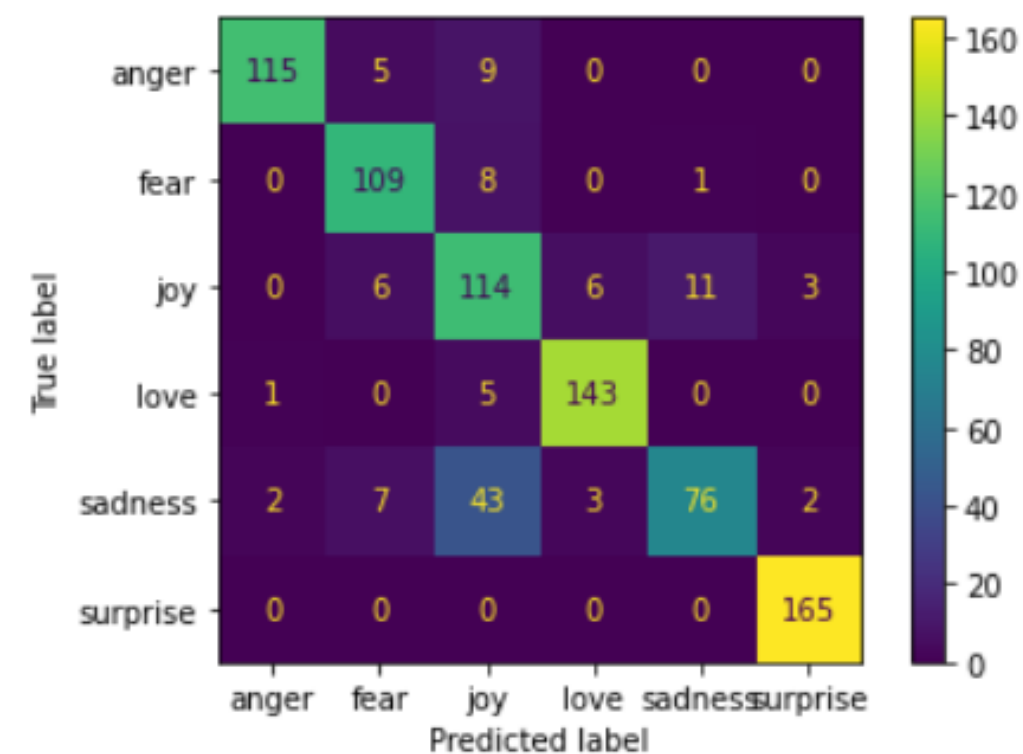
RECALL

0.8657



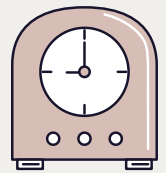
PUNTUACIÓN F1

0.8642



Modelo K-nearest neighbors

Se basa en el concepto de proximidad, es decir, clasifica los elementos similares de acuerdo a la distancia Euclidiana. El concepto se basa en que los elementos parecidos deben estar cerca.



EXACTITUD

0.88



PRECISIÓN

0.8761



RECALL

0.8777



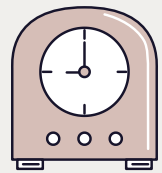
PUNTUACIÓN F1

0.8713



Modelo Naive-Baynes

Se basa en las probabilidades y es útil para obtener clasificaciones precisas. Se aplica la ecuación del teorema de Baynes para predecir una instancia de prueba x.



EXACTITUD

0.90



RECALL

0.8980



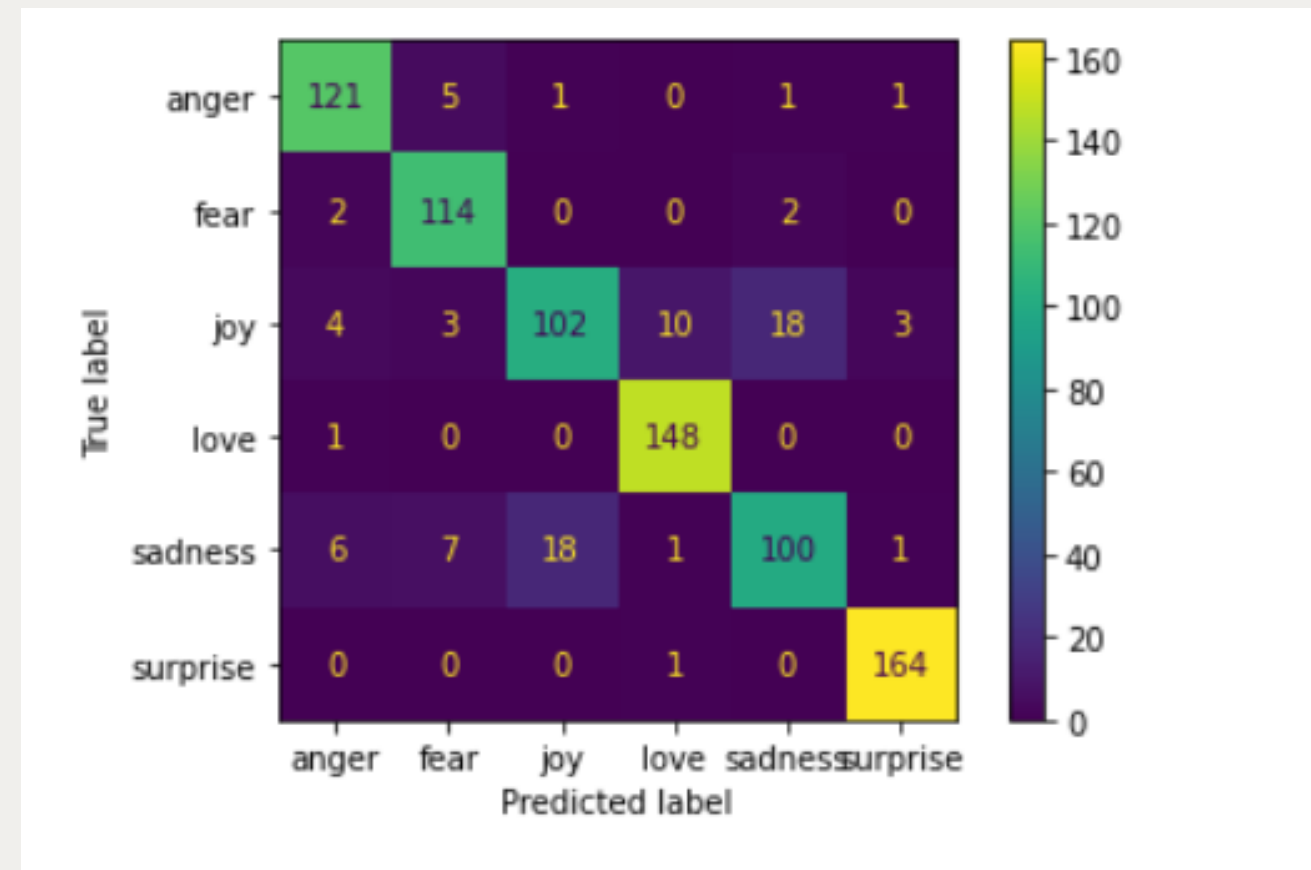
PRECISIÓN

0.8952



PUNTUACIÓN F1

0.8951



BI

Resultados

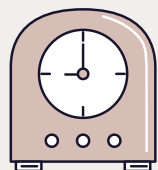


Resultados

Se obtuvo la siguiente matriz de resultados respecto a:



PARÁMETROS



EXACTITUD



PRECISIÓN



RECALL



PUNTUACIÓN F1

	Modelo	Params	Accuracy	Recall	Precission	F1 Score
0	Árbol de clasificación	{'criterion': 'entropy', 'max_depth': 20, 'min...	0.509592	0.509592	0.565810	0.463103
1	Random Forest	{'criterion': 'gini', 'max_depth': None, 'min_...	0.865707	0.865707	0.876914	0.864229
2	Modelo SVM	{'C': 2.154434690031882}	0.931655	0.931655	0.930362	0.930537
3	K Nearest Neighbors	{'n_neighbors': 1, 'p': 1}	0.877698	0.877698	0.876116	0.871313
4	Naive Baynes	{'alpha': 0, 'fit_prior': True}	0.898082	0.898082	0.895253	0.895132

Resulta bastante evidente concluir que el mejor modelo es el modelo SVM. Este modelo podría ser utilizado por un negocio para conocer la opinión que puedan tener los usuarios de las redes sociales sobre algunos productos o campañas de marketing que se han realizado. El conocer la perspectiva de las personas puede suponer una ventaja comercial para el negocio con respecto al mercado.

Bibliografía

- Amat, J. (diciembre 2020). Análisis de texto con Python. Recuperado de <https://www.cienciadedatos.net/documentos/py25-text-mining-python.html>
- AprendelA. (2021). Naive-Baynes Teoría [Blog virtual]. Recuperado de <https://aprendeia.com/naive-bayes-teoria-machine-learning/>
- Mosquera, R., Castrillón, O., Parra, L.(2018). Máquinas de Soporte Vectorial, Clasificador Naïve Bayes y Algoritmos Genéticos para la Predicción de Riesgos Psicosociales en Docentes de Colegios Públicos Colombianos. Recuperado de https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07642018000600153&lang=pt