

01

TRANSFORMANDO INFORMACIÓN EN OPORTUNIDADES

Investigación sobre Modelos de
Clasificación de tono de Noticias
para Bancolombia

Lorenzo Ramirez Calderon

Colegio Marymount

02

Introducción

“El tono que usan los medios de comunicación constituye una representación muy fidedigna de los ánimos de los inversionistas”(Darmien Puy, 2019)

- Poseedores confianza financiera publica
- Modelo de clasificación de noticias más eficaz para Bancolombia
- Automatización, Optimización, Reconociendo de patrones



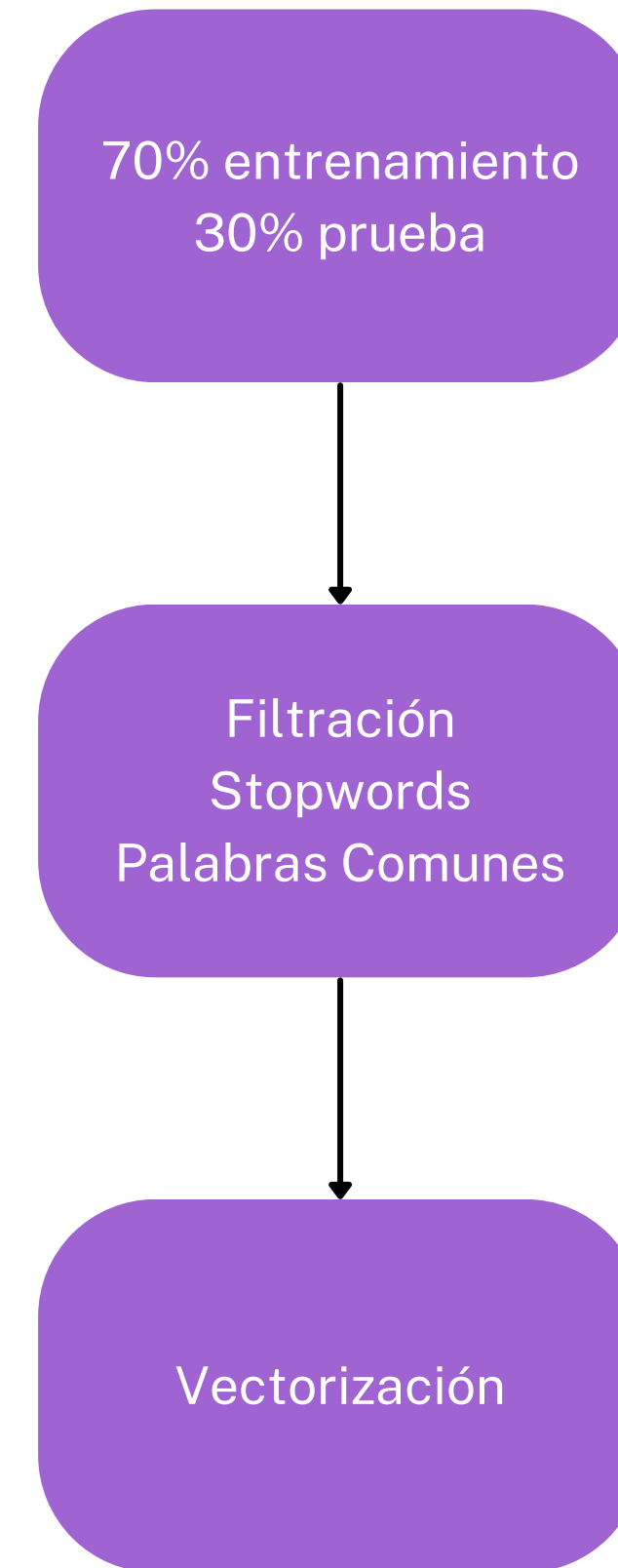
 **Bancolombia**



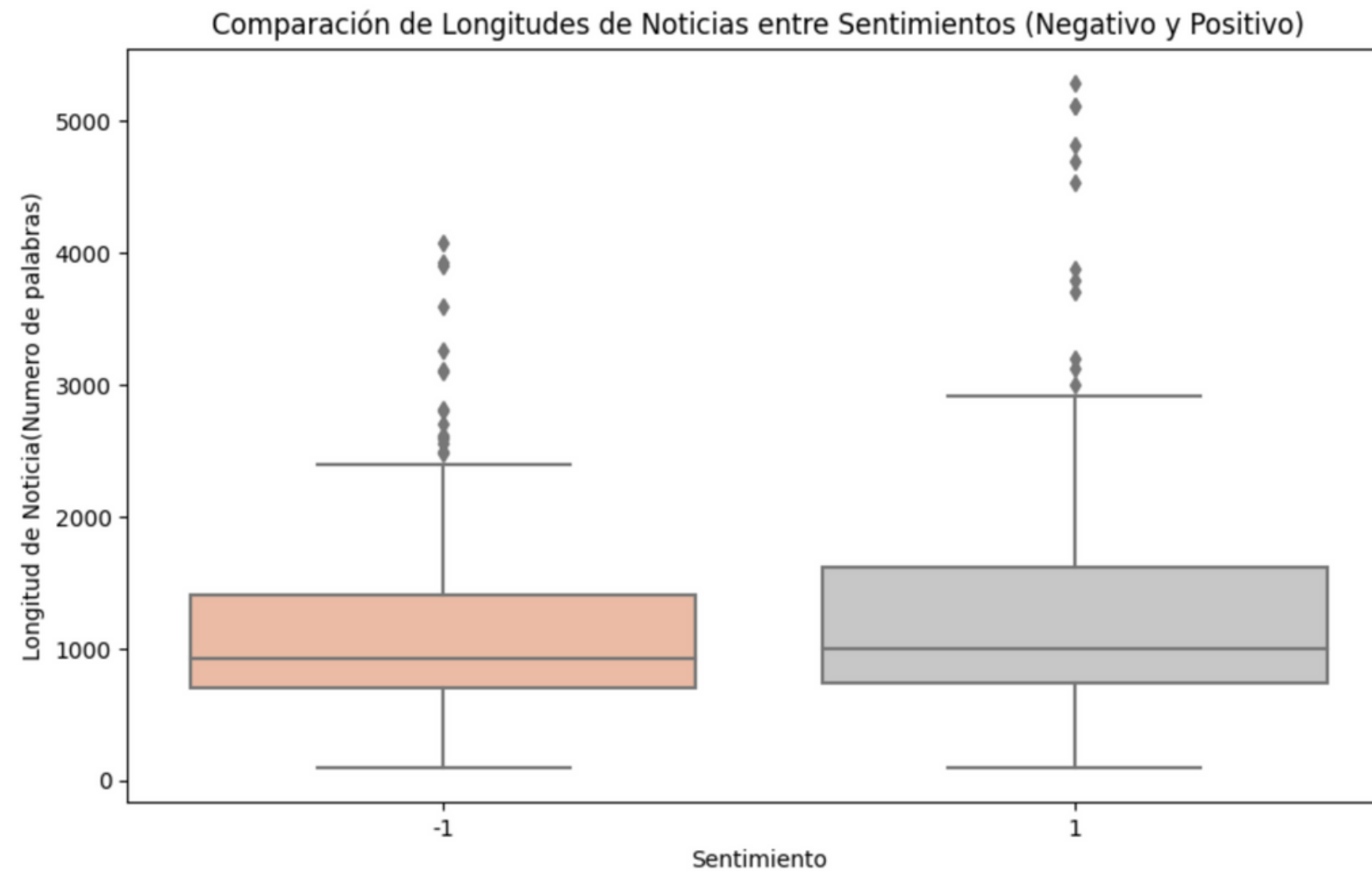
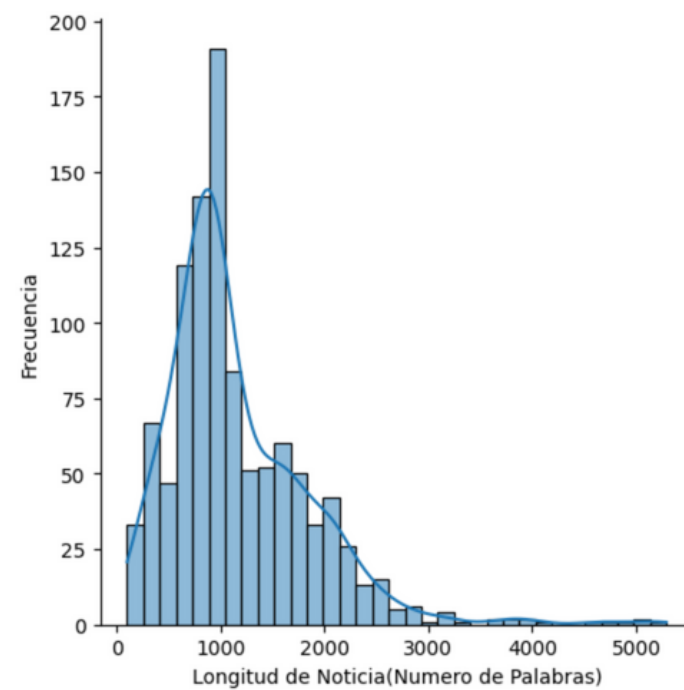
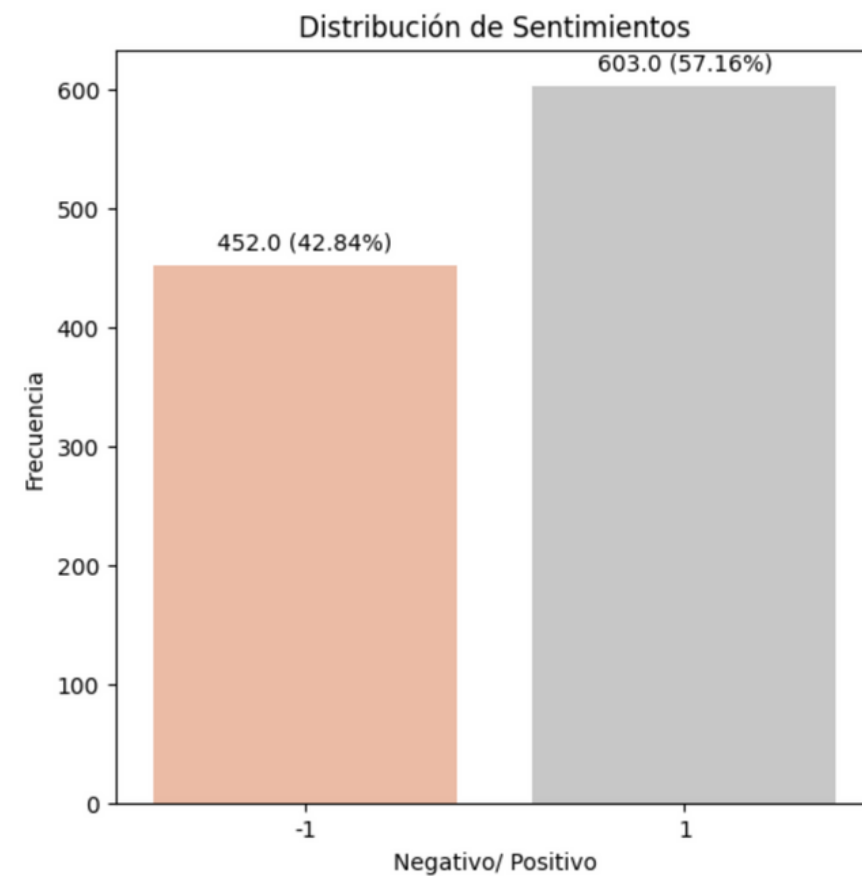
03

Metodología

- Aprendizaje Supervisado
- 1055 Noticias de diferentes ambitos
- Titulo, Descripción, Texto
- Análisis Exploratorio
- K Vecinos, Random Forest, Árbol de decisión, y Support Vector Machine
- Proceso de Hiperparametrización
- Precision, Recall, Exactitud, y F1-Score

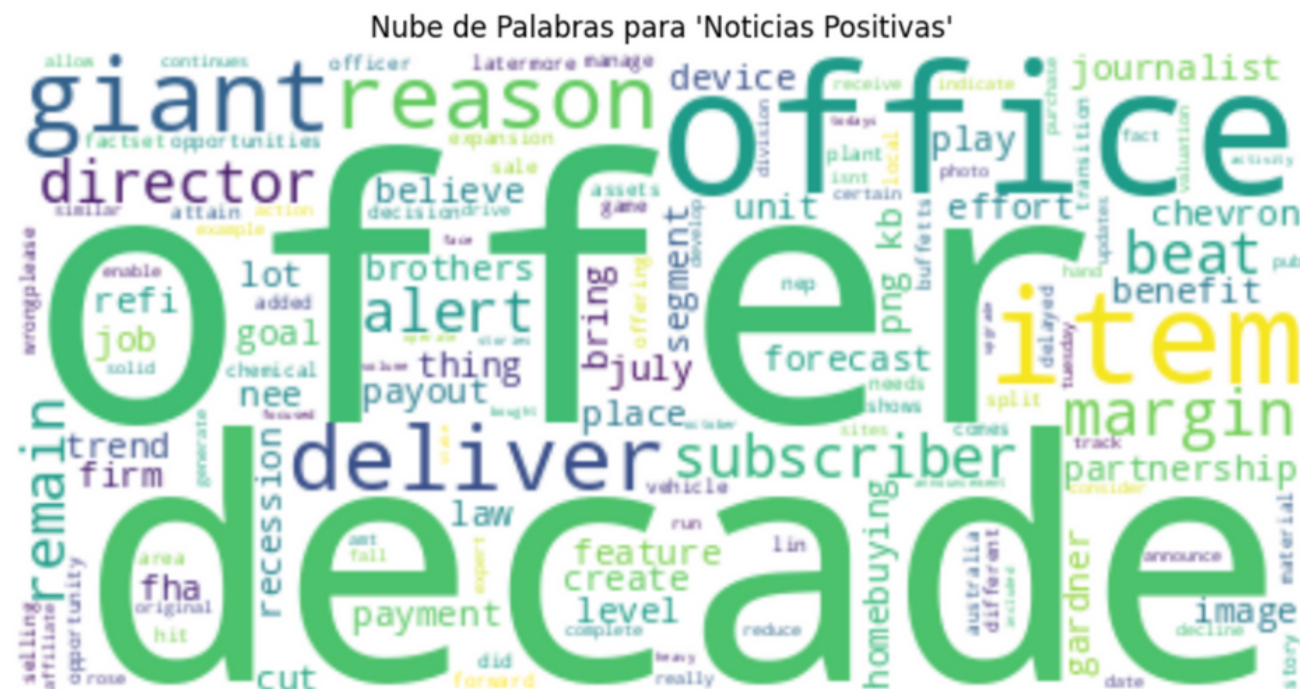
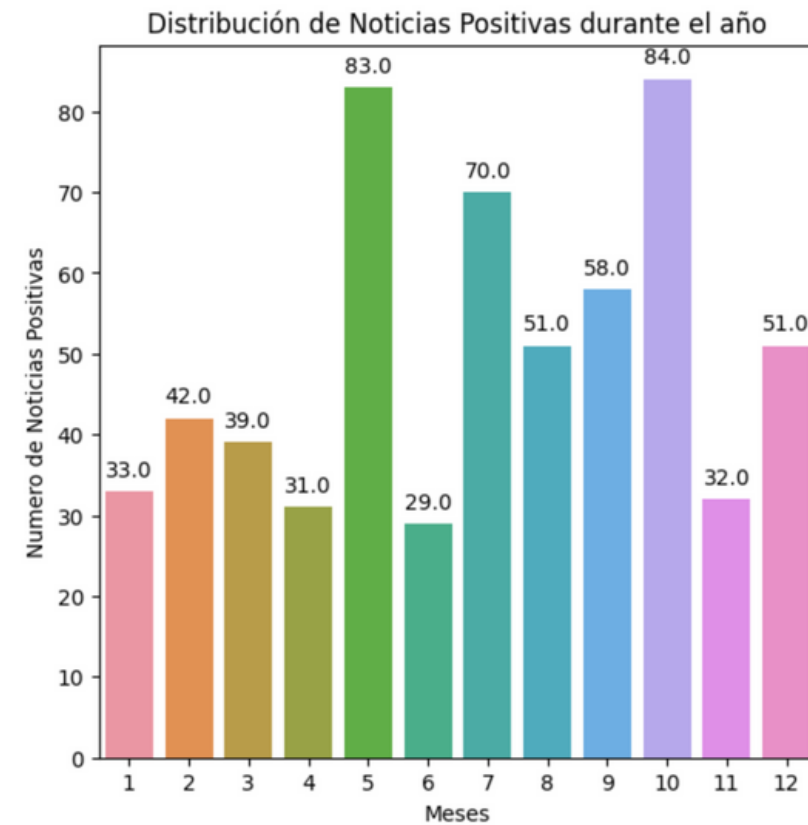


Análisis Exploratorio General

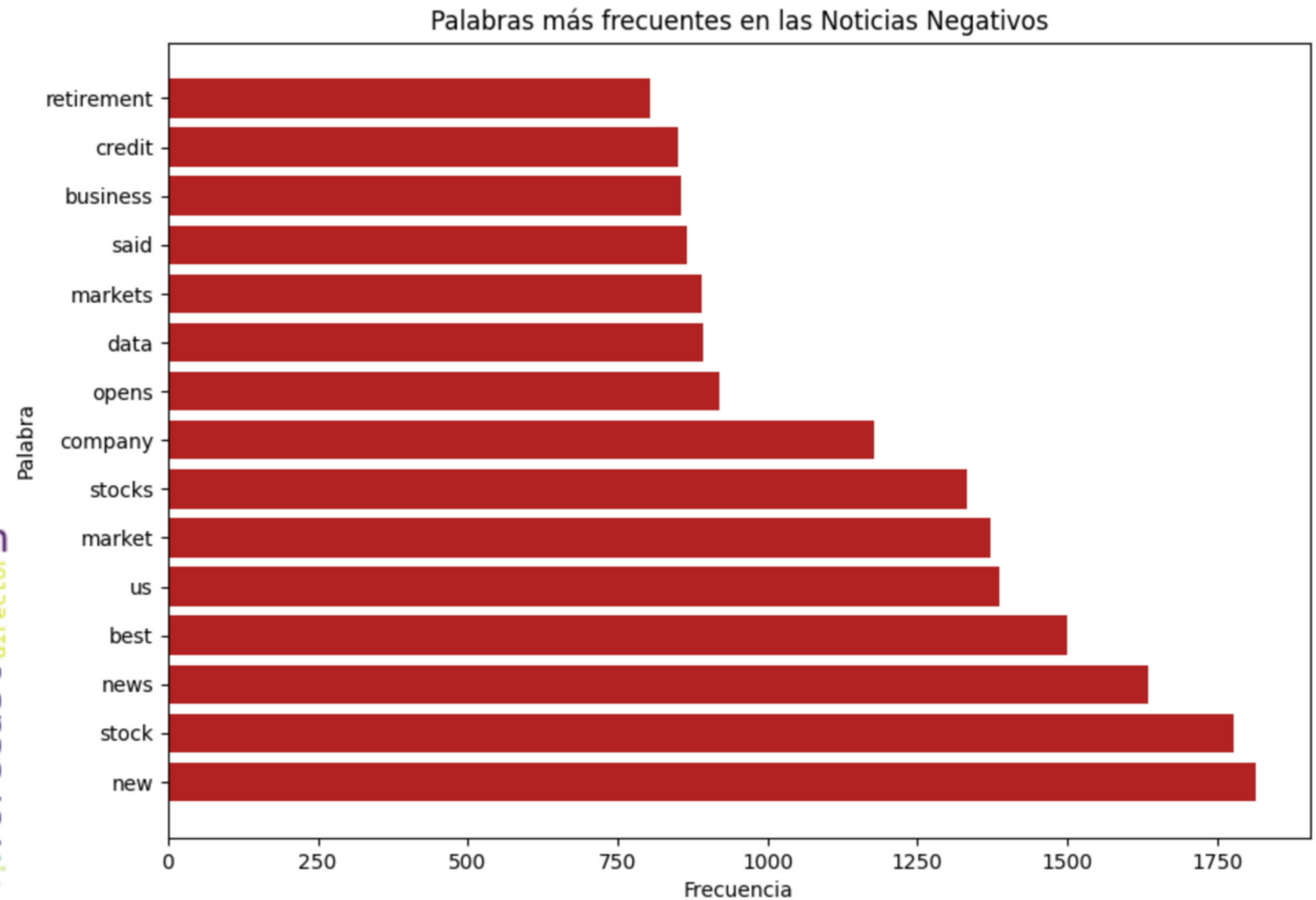
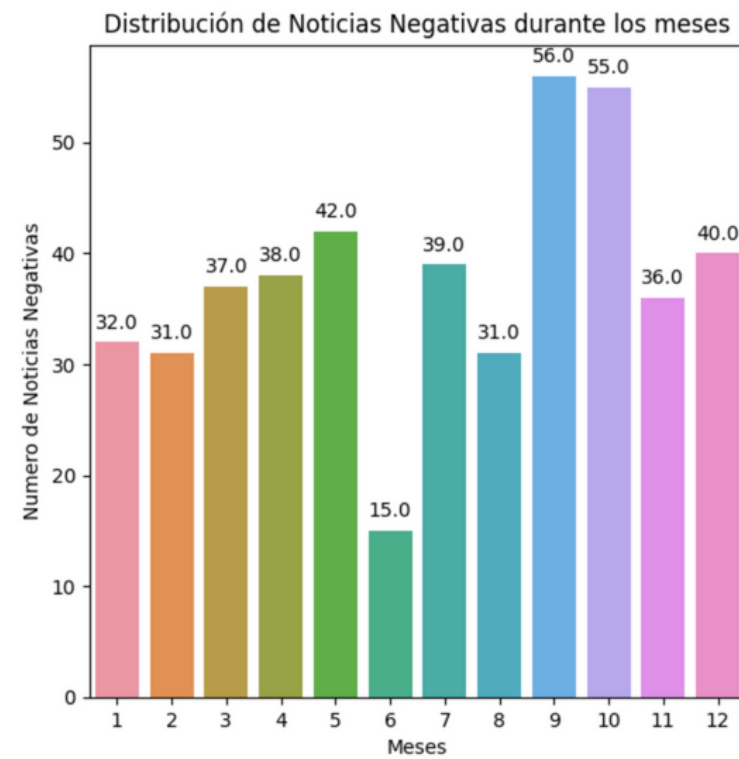


- Correlación entre Longitud de Noticias y Sentimiento: 0.08113459837230774
- Shapiro-wilk- Longitud no sigue distribución normal, asimetría a la derecha
- Valores atípicos desde 3241 palabras

Análisis Exploratorio Positivo

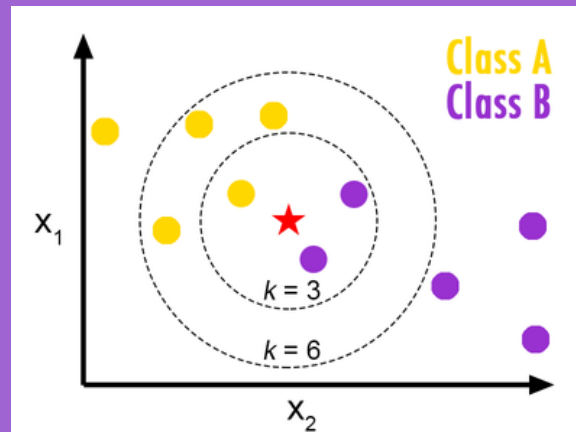


Análisis Exploratorio Negativo



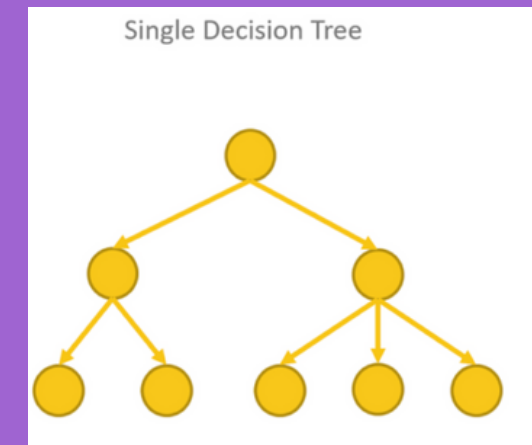
Modelos de Clasificación

KNN Vecinos



$$d(x,y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Arbol de Decision



Proceso de Hiperparametrización

KNN- 39 Vecinos

Arbol de Decision- Random State 42

Random Forest- 200 arboles, Depth 15

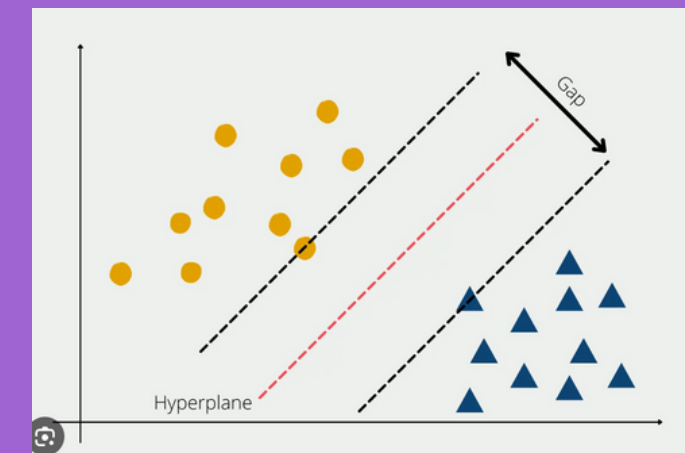
SVM- C=1, Kernel-RBF

Random Forest



$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

Support Vector Machine



08

Resultados

Notebook

Discussion

Precision>Exactitud> F1-Score

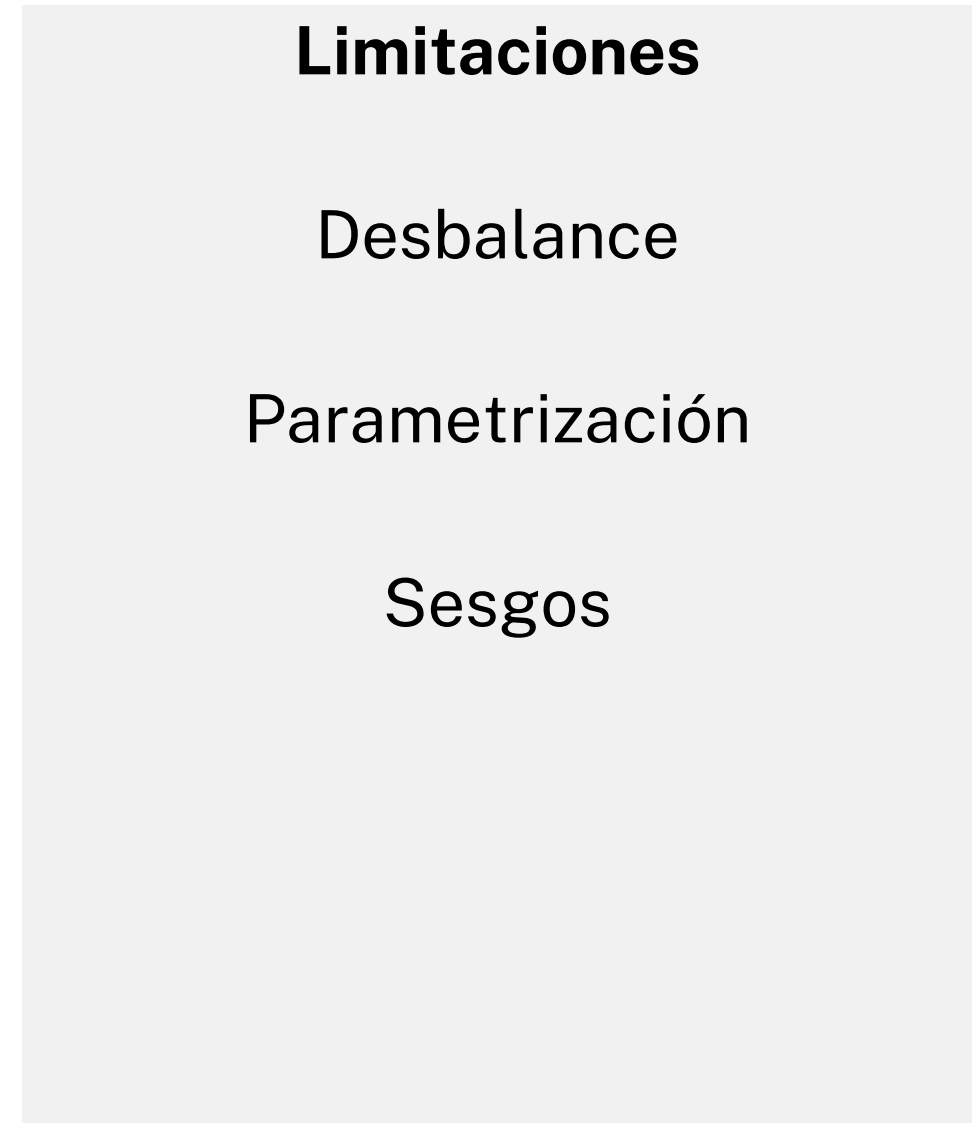


Sectorizacion, Diferentes Planos



Parametrización

Sesgos

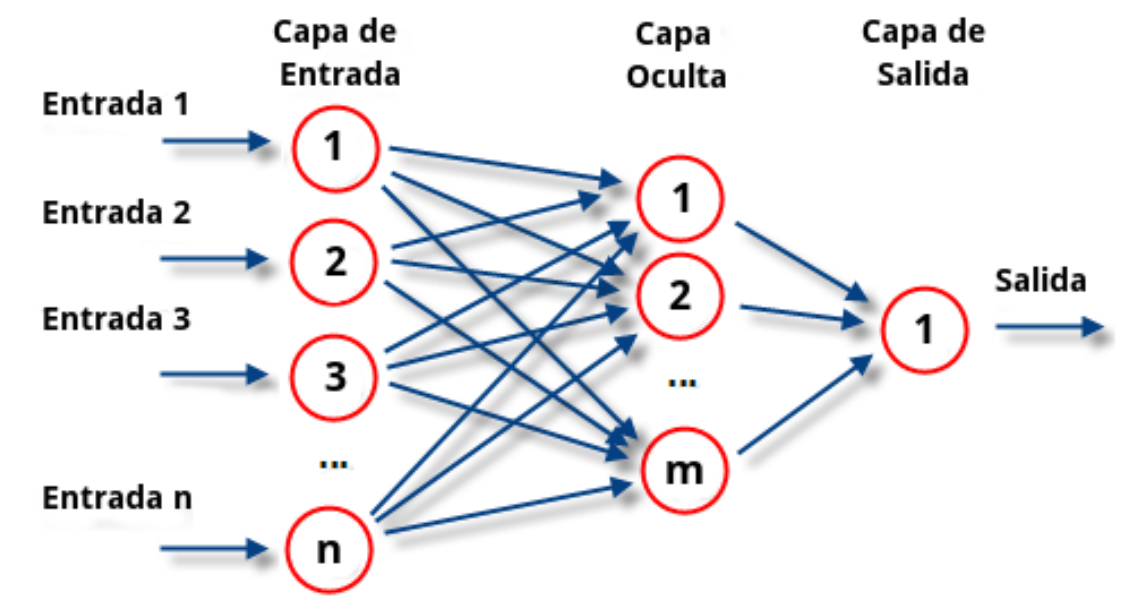


Conclusiones

- Pico de Noticias durante el año
- Definición del tono

Maquina de Soporte Vectorial la ideal para Bancolombia

- Mas características
- Filtración de noticias falsas
- Redes Neuronales





Recomendaciones

- Maquina de Soporte Vectorial y dedicación de recursos a futuras investigaciones(Transacciones Fraudulentas)
- Atención a la Geopolitica Internacional
- Precuacion, o riesgo dependiendo del año

Referencias

El poder de la palabra: La influencia de las noticias en los ánimos de los mercados financieros. (2019, December 16). IMF. [El poder de la palabra: La influencia de las noticias en los ánimos de los mercados financieros. \(2019, December 16\). IMF.](#)

Shung, K. P. (2020, April 10). Accuracy, precision, recall or F1? - towards data science. Medium. <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>
