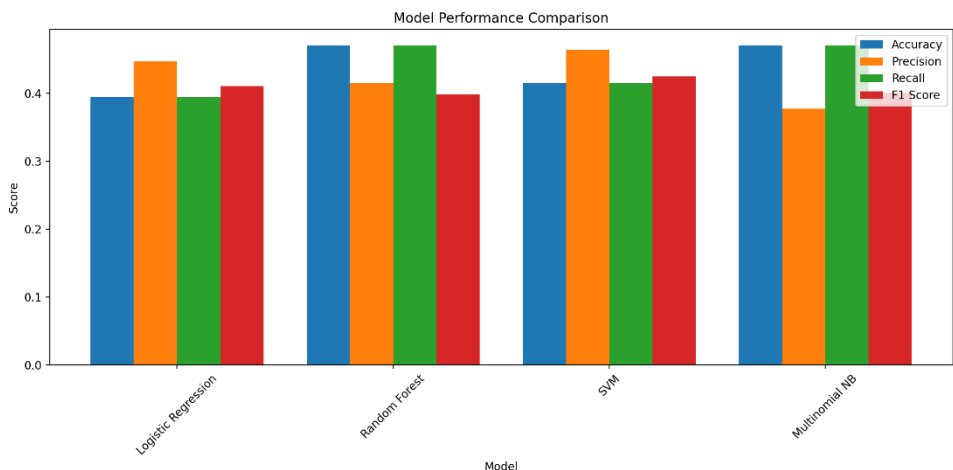


Proceso de entrenamiento de modelos

Se llevó a cabo el proceso de entrenamiento de modelos, a partir de:

1. Anonimización y preprocesamiento de datos:
 - Eliminar URLs y anonimización usuarios expuestos en los mensajes
 - Limpiar el conjunto de datos y se eliminaron los valores faltantes.
 - Dividir los datos en conjuntos de entrenamiento (80%) y de prueba (20%)
 - Vectorizar el texto usando TF-IDF con unigramas y bigramas
2. Entrenamiento de modelos:
 - Se entrenaron varios modelos de aprendizaje automático con pesos de clase equilibrados:
 - Regresión logística
 - Bosque aleatorio
 - Máquina de vectores de soporte (SVM)
 - Bayes ingenuo multinomial
3. Evaluación del modelo:
 - Se evaluó cada modelo utilizando precisión, exactitud, recuperación y puntuación F1.
 - Así es como se comportaron los modelos:



El modelo SVM logró el mejor rendimiento con:

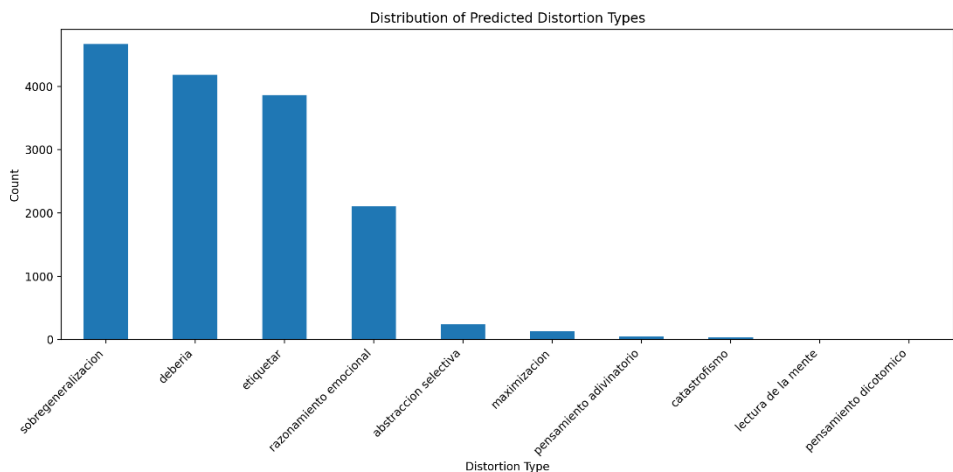
- Precisión: 0,7483
- Precisión: 0,7483
- Recordatorio: 0,7483
- Puntuación F1: 0,7483

Aplicación al conjunto de datos principal

Se aplicó el mejor modelo (SVM) al conjunto de datos principal "Datos definitivos enero 2021 sin duplicados y solo comentarios.xlsx", que contiene 15.302 registros. A continuación, se muestra una muestra de las predicciones:

sobregeneralización: 327

La distribución de los tipos de distorsión cognitiva previstos en el conjunto de datos principal:



Como puede ver, las distorsiones cognitivas más comunes en el conjunto de datos son:

- Sobregeneralización (4.666 casos)
- Declaraciones "debería" (4183 casos)
- Etiquetado (3.863 instancias)
- Razonamiento emocional (2.109 casos)

Algoritmo de clasificación final

El proceso de clasificación final consta de:

Algoritmo de clasificación final:

- Preprocesar los datos de texto
- Vectorizar usando TF-IDF
- Aplicar el modelo SVM
- Obtenga el tipo de distorsión previsto y la puntuación de confianza

Este análisis proporciona información valiosa sobre los tipos de distorsiones cognitivas presentes en el conjunto de datos, lo que puede ser útil para comprender patrones en el discurso en línea y potencialmente desarrollar intervenciones para abordar patrones de pensamiento dañinos.