

Cyberbullying

**Maestría en Ciencia de Datos
Procesamiento y Clasificación de Datos**

Profesora: Mayra Cristina Berrones Reyes

José Alberto López Álvarez	1553133
Irving Daniel Estrada López	1739907
América Victoria Ramírez Cámara	1458051

Agenda

- Introducción
 - ¿Qué es el cyberbullying?
 - ¿Por qué es importante detectar el cyberbullying?
 - Consecuencias del cyberbullying
- Planteamiento del Problema
 - Conjunto de Datos
 - Problema a resolver
- Desarrollo
- Resultados
- Conclusión
- Referencias

Introducción

¿Qué es el cyberbullying?

- Acoso con el uso de tecnologías digitales. Puede tener lugar en las redes sociales, plataformas de mensajería, plataformas de juegos y teléfonos móviles. Es un comportamiento repetido, dirigido a asustar, enojar o avergonzar a quienes son blanco.
- Deja una huella digital - un registro que puede resultar útil y proporcionar evidencia para ayudar a detener el abuso.



¿Por qué es importante detectar el cyberbullying?

Consecuencias del cyberbullying

Mentalmente

- Sentirse molesto, avergonzado, estúpido, incluso asustado o enojado.

Emocionalmente

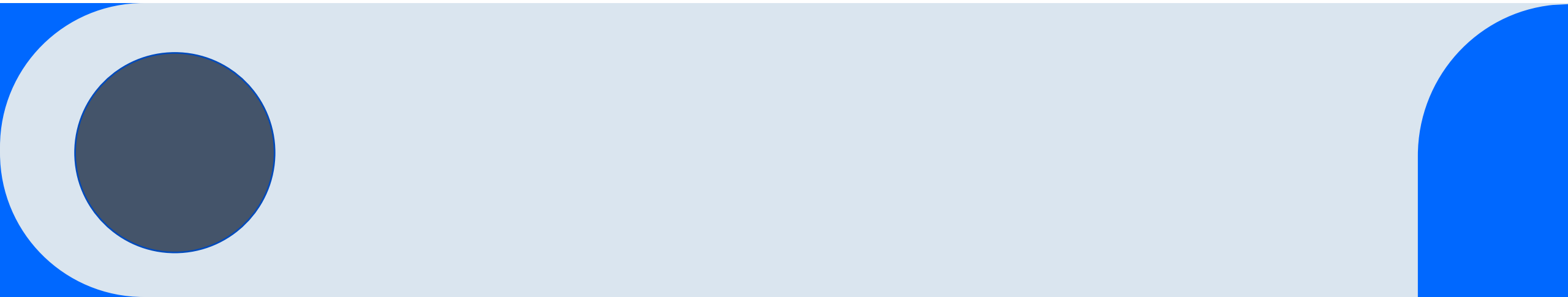
- Sentirse avergonzado o perder interés en las cosas que ama.

Físicamente

- Cansado (pérdida de sueño), o experimentar síntomas como dolores de estómago y dolores de cabeza.

Cuando sufres ciberacoso puedes sentirte avergonzado, nervioso, ansioso y tener dudas sobre lo que la gente dice o piensa de ti. Esto puede llevarte a aislarte de tus amigos y familiares, a tener pensamientos negativos y a sentirte culpable por las cosas que has hecho o dejado de hacer, y a creer que te están juzgando negativamente. También es habitual sentirse solo y abrumado, y sufrir dolores de cabeza, náuseas o dolores de estómago frecuentes.

Planteamiento del Problema



Conjunto de Datos

Nuestro dataset “cyberbullying_tweets.csv” se obtuvo de Kaggle y cuenta con 48,000 registros y 2 variables:

Variables	Tipo de Variable
tweet_text	texto
cyberbullying_type	texto

Problema a resolver

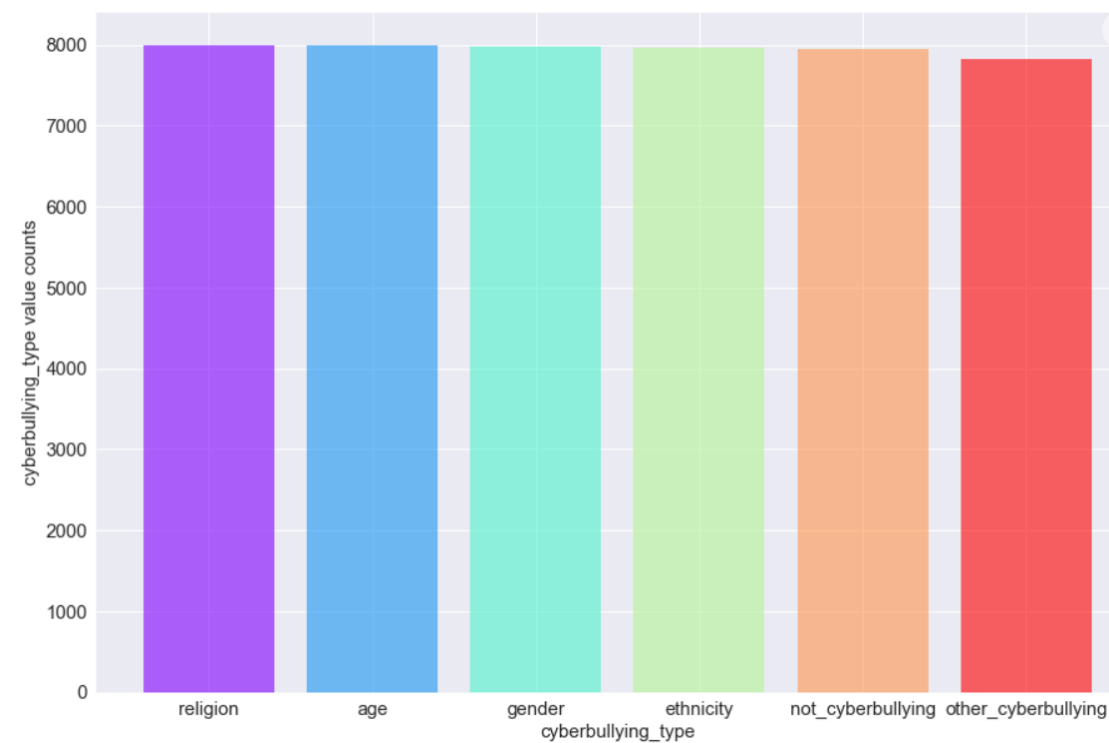
- Obtener palabras relevantes de cada uno de los tipos de cyberbullying.
- Predecir si un tweet es cyberbullying.
- Predecir el tipo de cyberbullying.



Desarrollo



	tweet_text	cyberbullying_type
0	In other words #katandandre, your food was cra...	not_cyberbullying
1	Why is #aussietv so white? #MKR #theblock #ImA...	not_cyberbullying
2	@XochitlSuckkks a classy whore? Or more red ve...	not_cyberbullying
3	@Jason_Gio meh. :P thanks for the heads up, b...	not_cyberbullying
4	@RudhoeEnglish This is an ISIS account pretend...	not_cyberbullying



Preprocesado de texto

1. Remover usuario
2. Remover emojis
3. Remover URL
4. Remover signos de puntuación y números
5. Convertir a minúsculas
6. Obtener el lemma de cada una de las palabras
7. Tokenization
8. Remover Stop Words

TF - IDF

TF: se refiere al cálculo de la frecuencia para cada palabra en un determinado texto.

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

IDF: es el cálculo de la frecuencia inversa y se utiliza para calcular el peso de las palabras "raras".

$$IDF(w) = \log \frac{N}{df_t}$$

Combinando estos 2 parámetros obtenemos el puntaje TF-IDF.

$$W_{i,j} = TF_{i,j} \cdot \log \frac{N}{df_i}$$

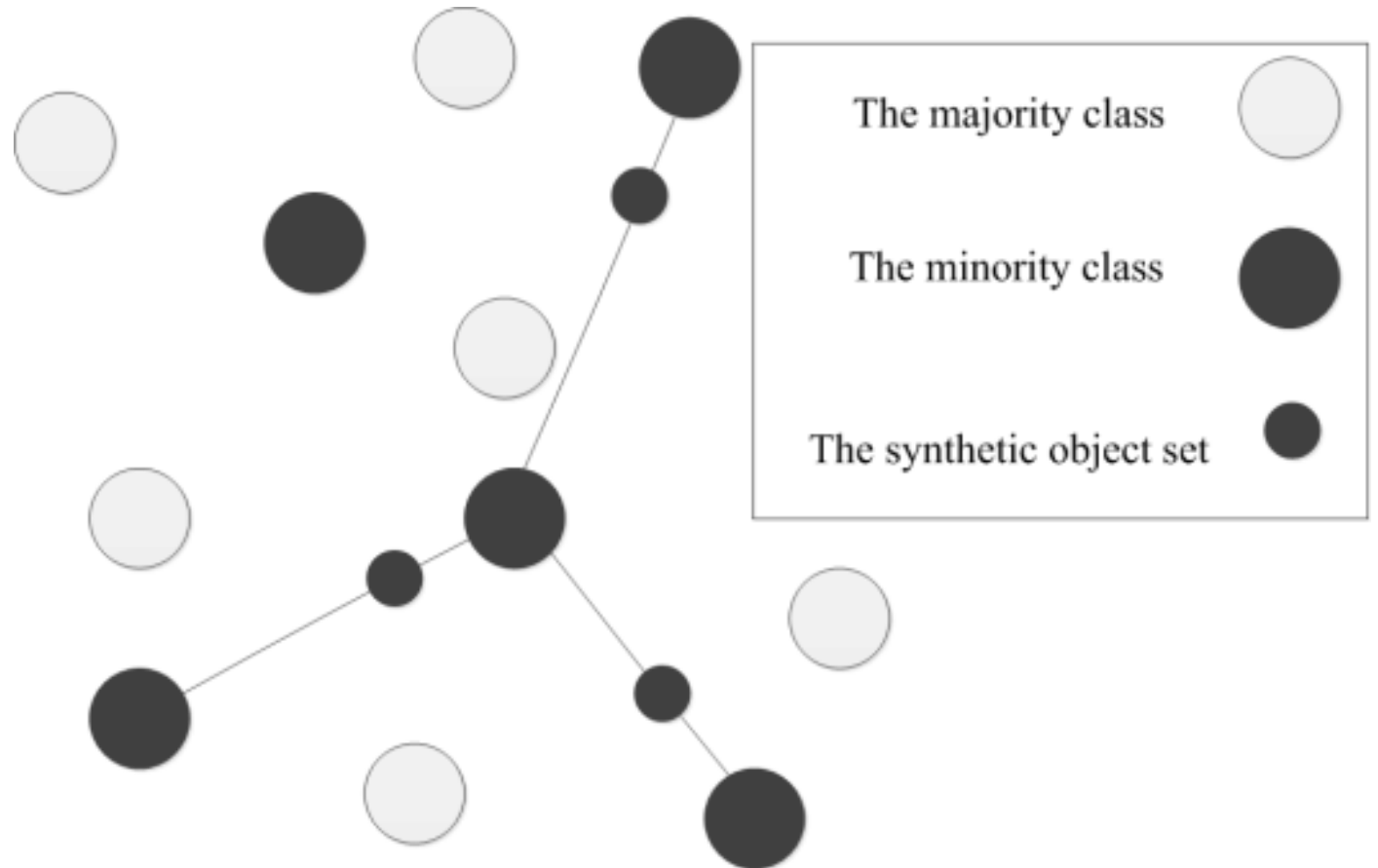
En donde:

$TF_{i,j}$ = Número de ocurrencias de i en j

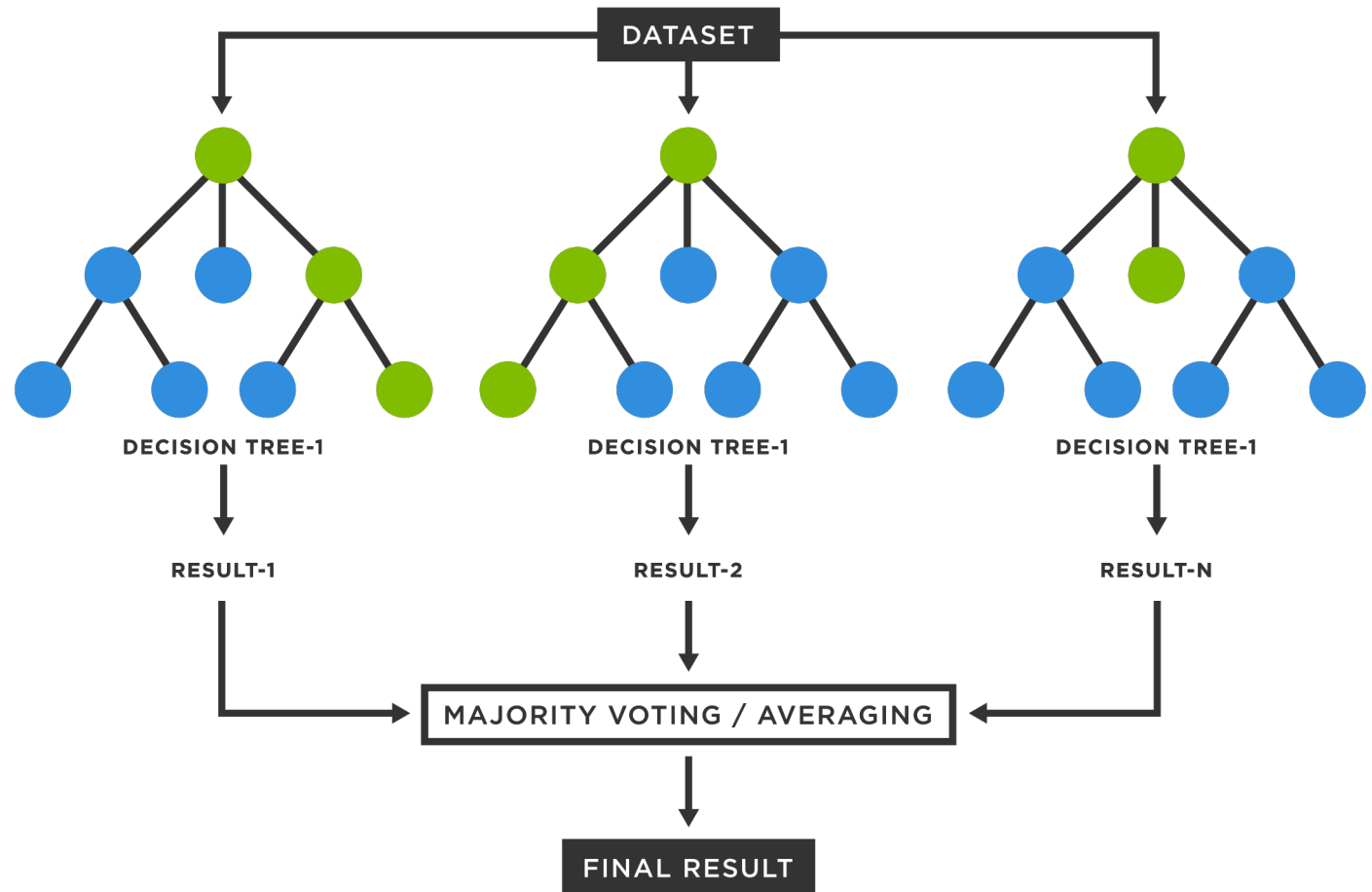
df_i = Número de archivos que contienen i

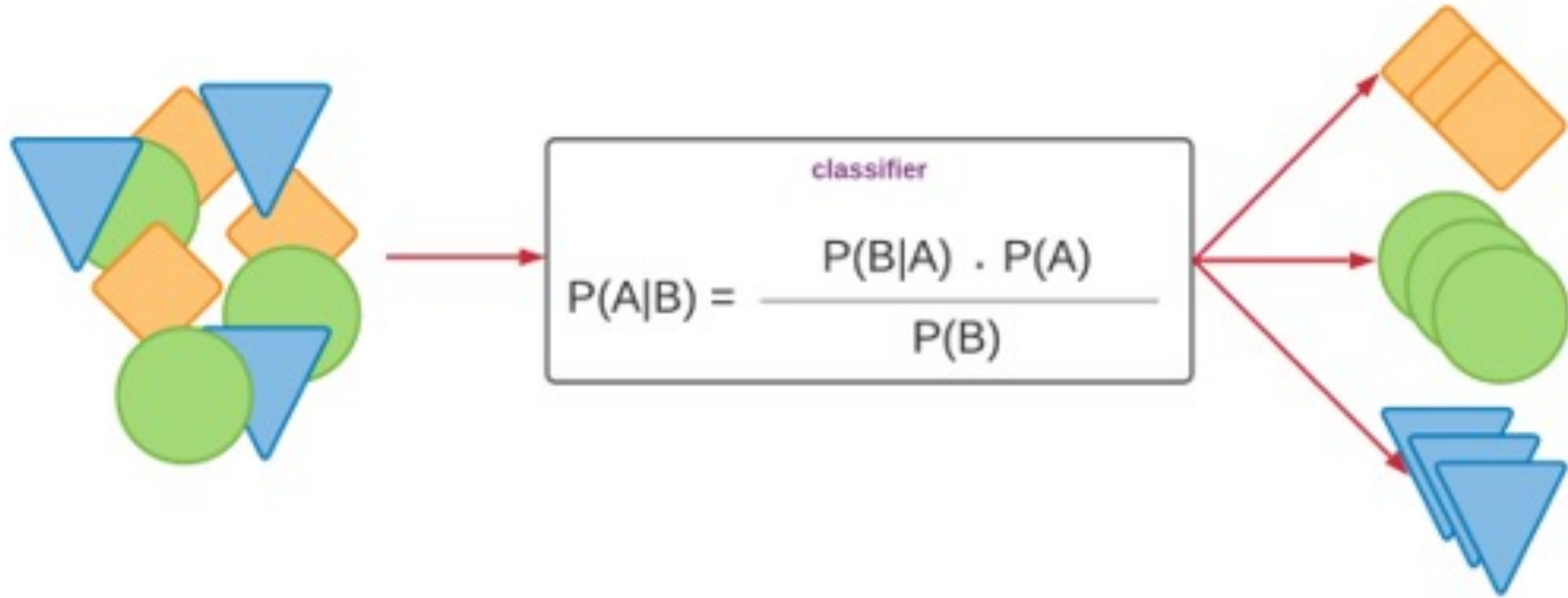
N = Número total de archivos

SMOTE



Random Forest

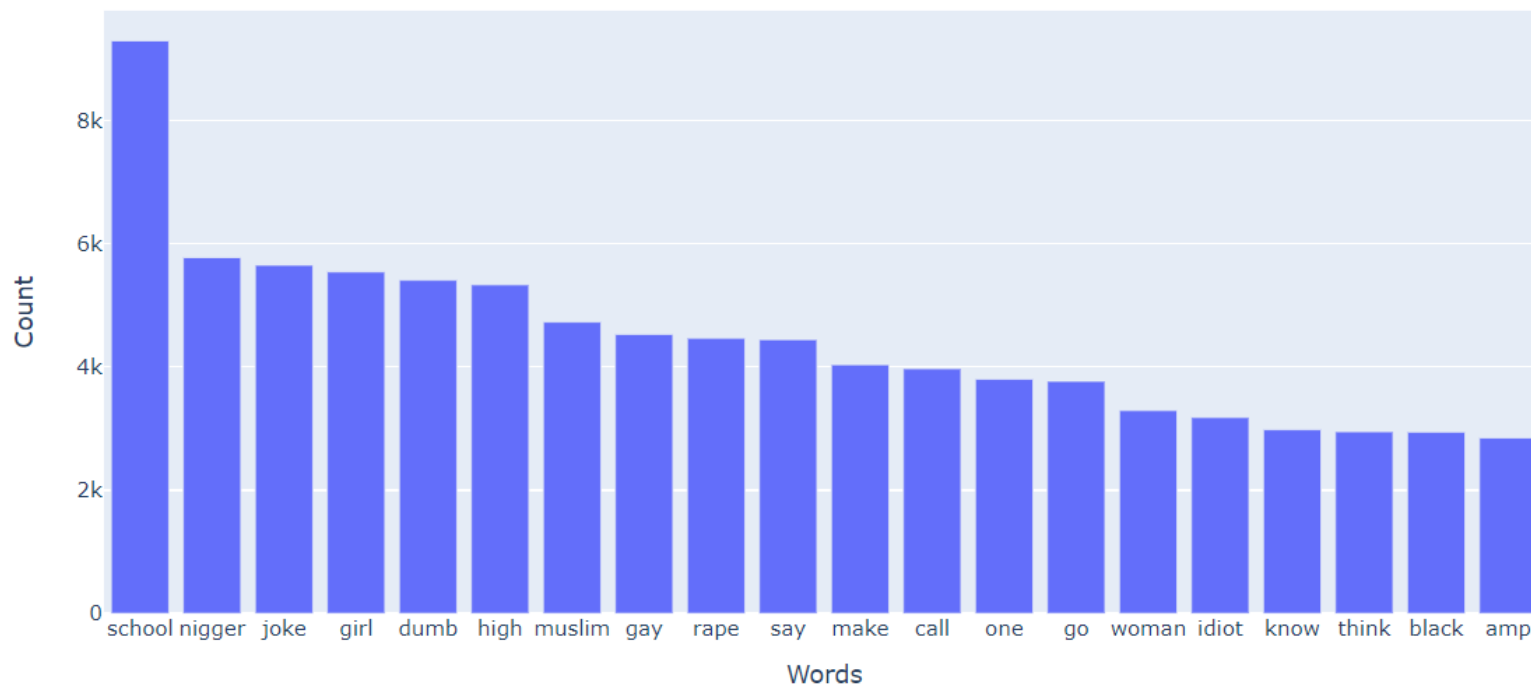




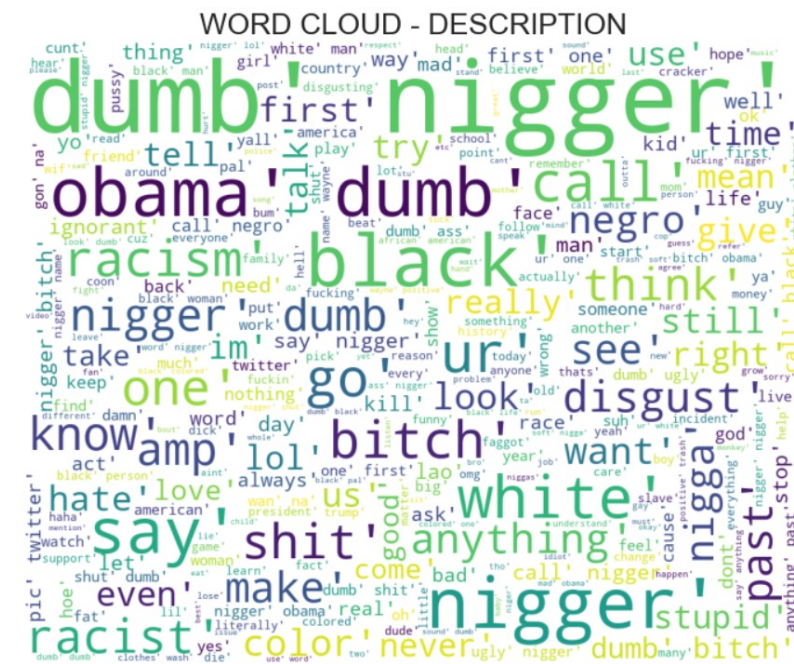
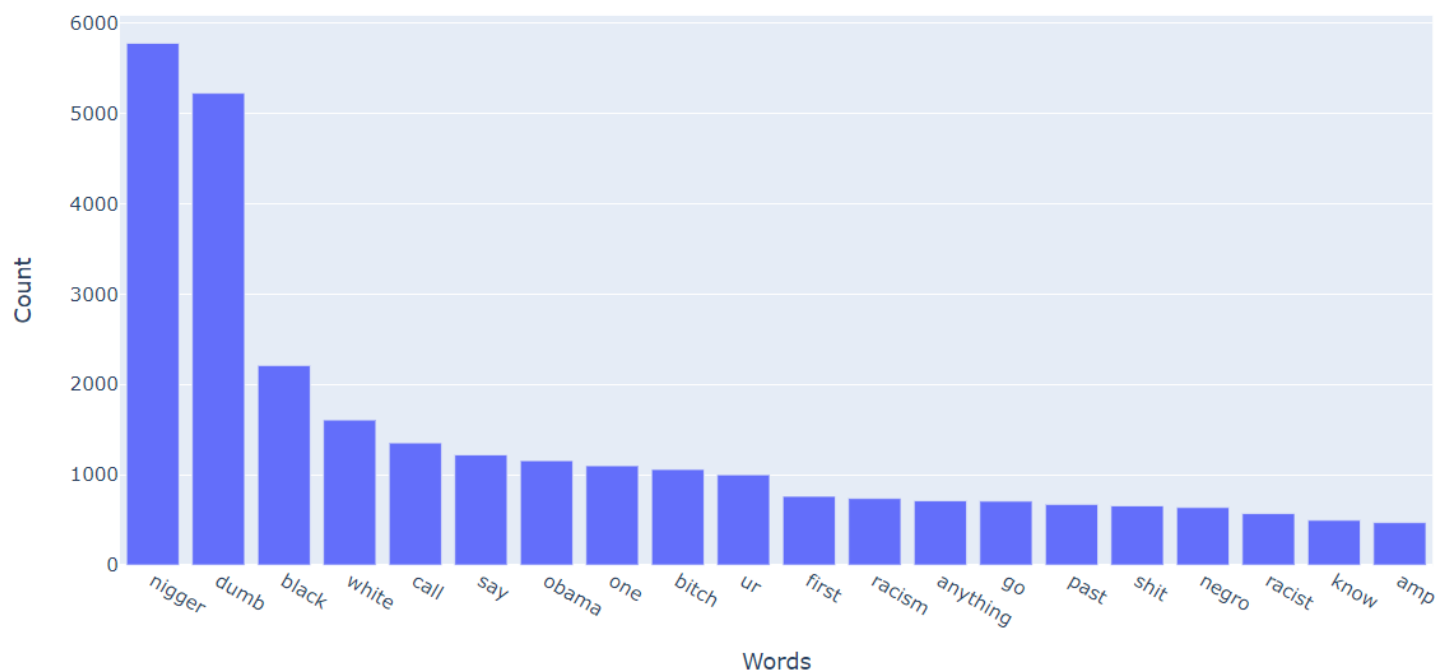
Multinomial Naïve Bayes

Resultados

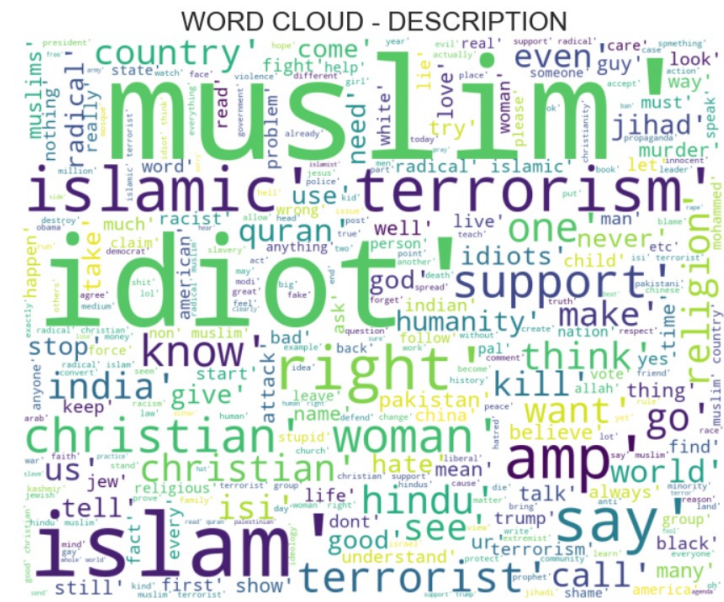
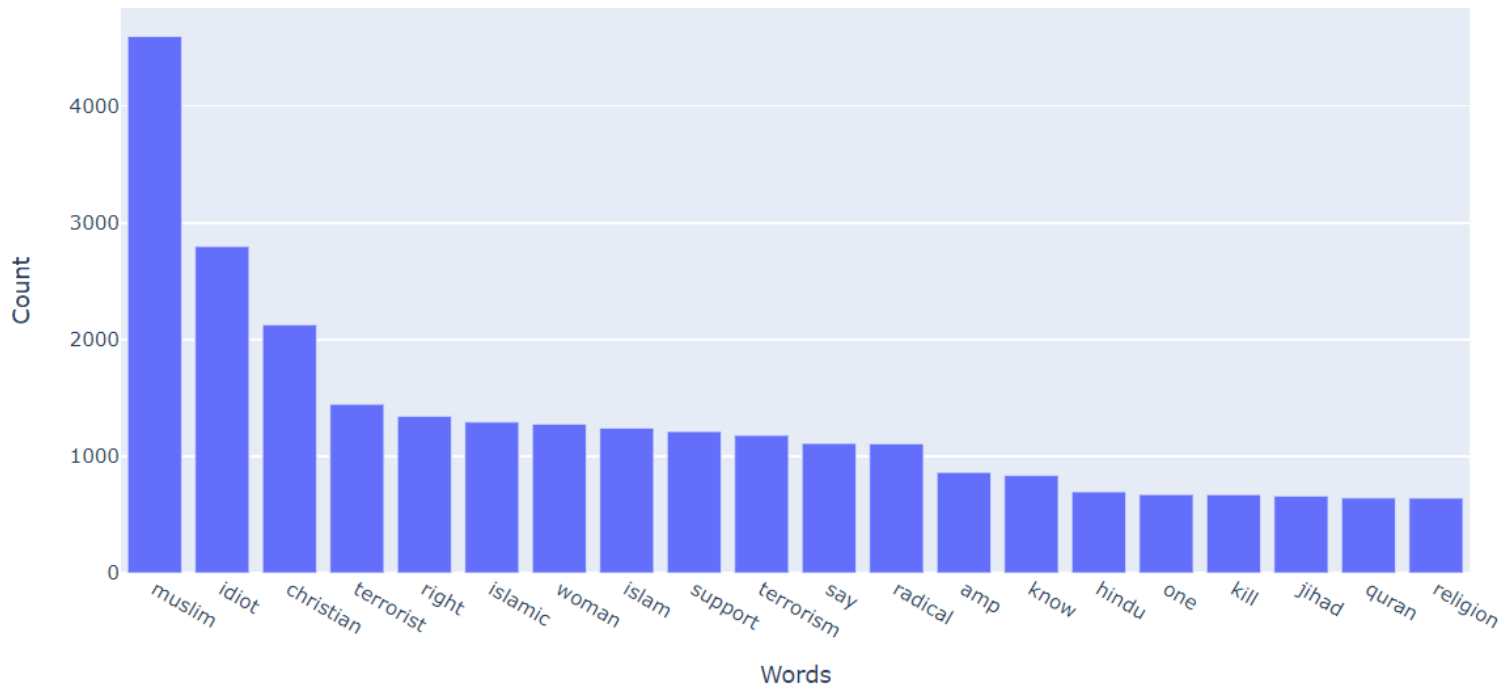
En General



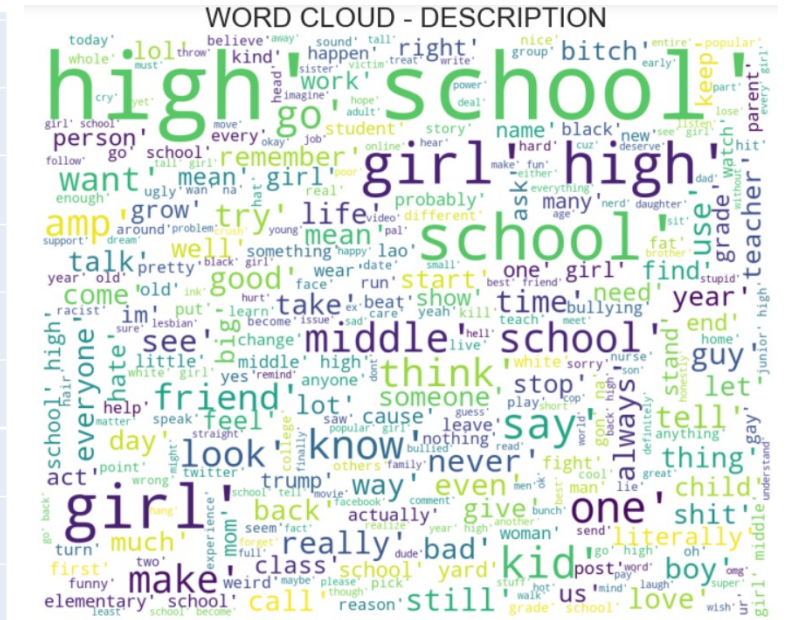
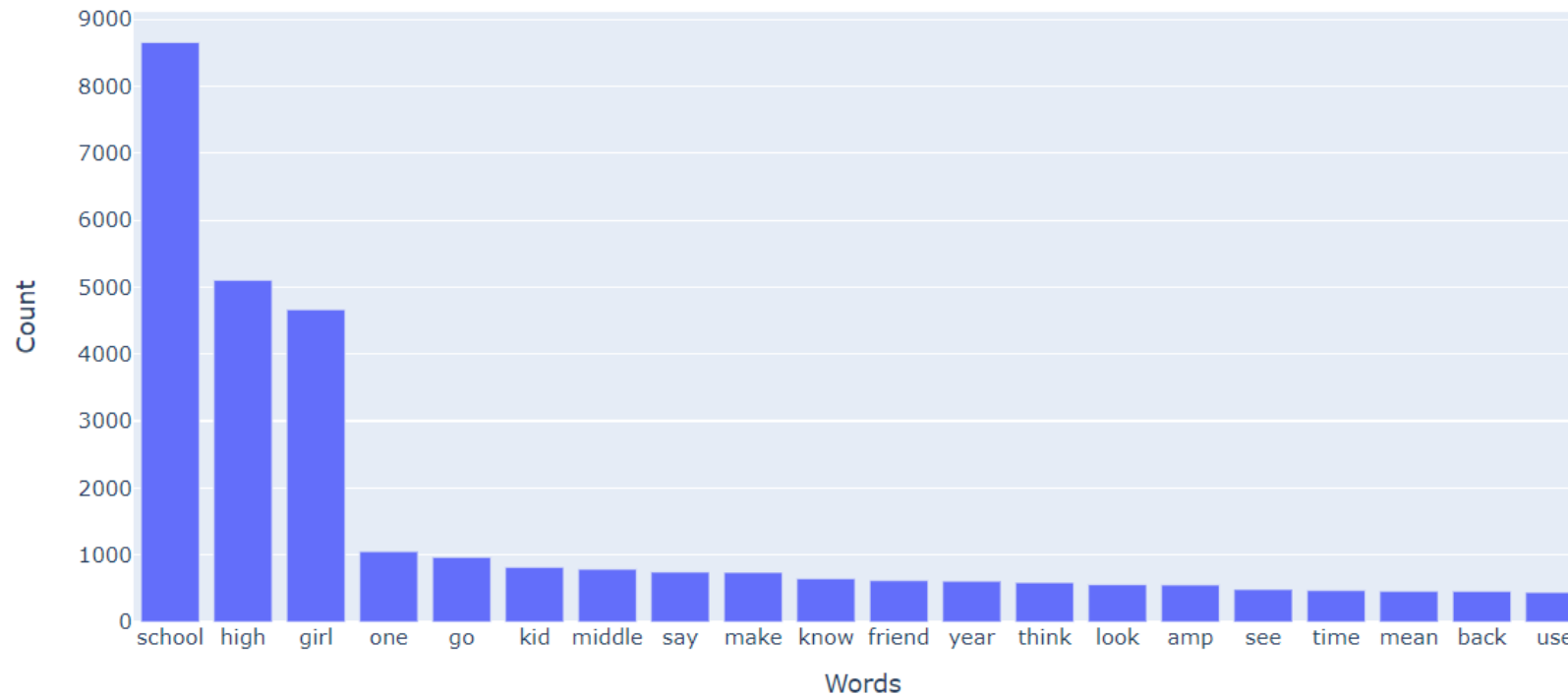
Ethnicity



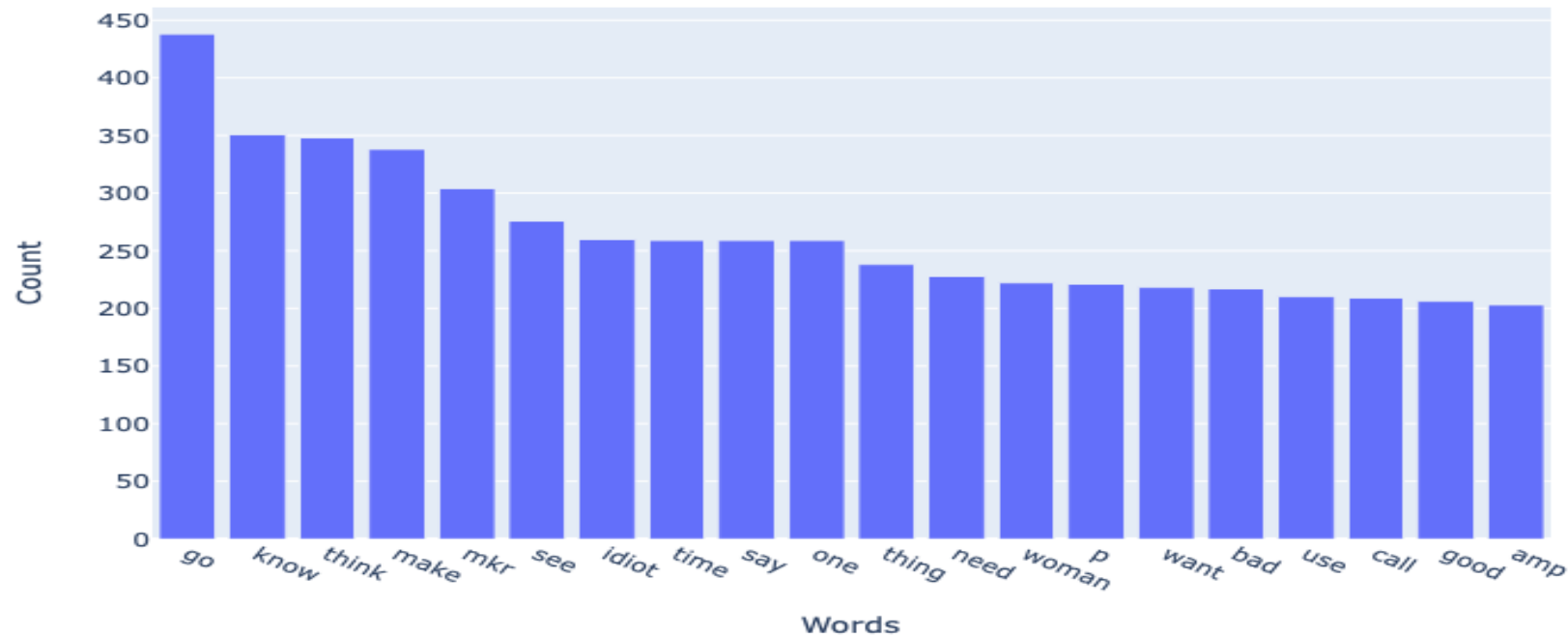
Religion



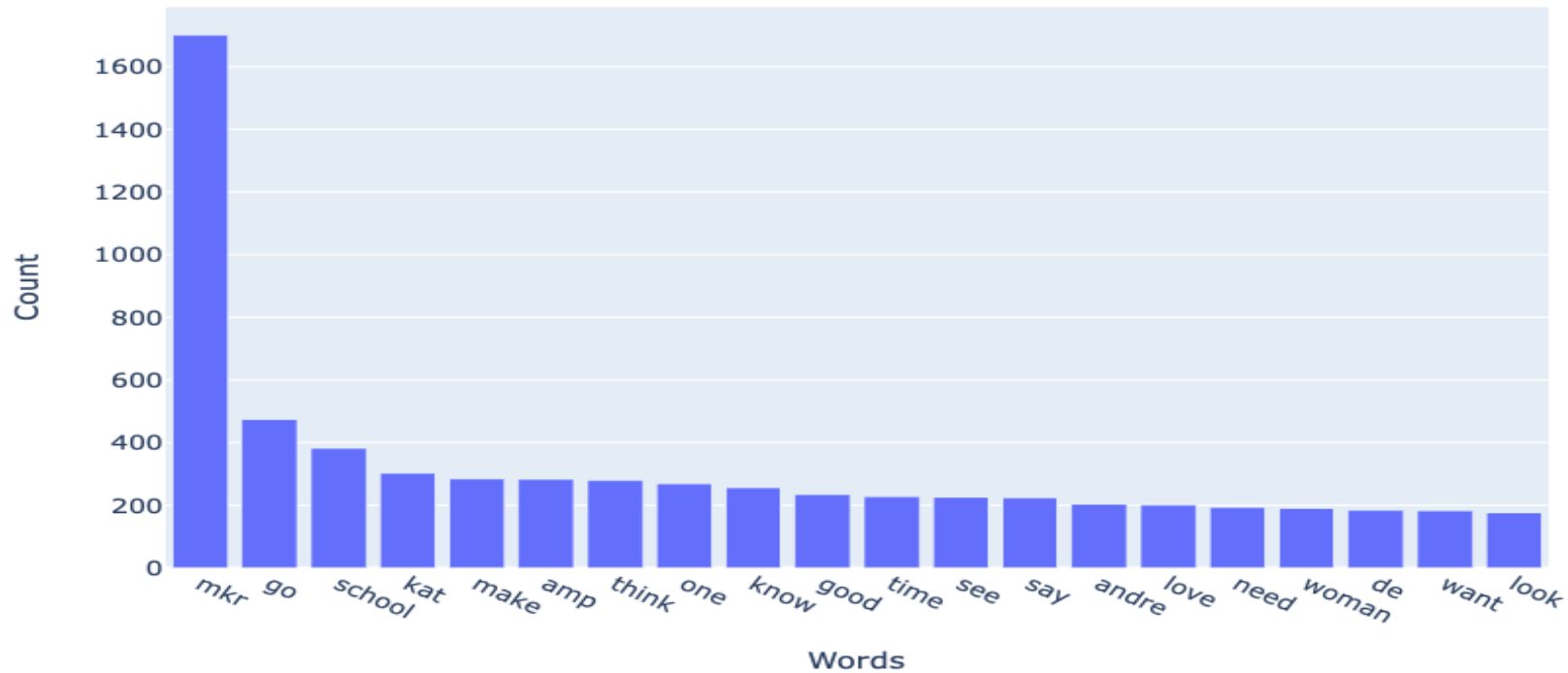
Age



Otros

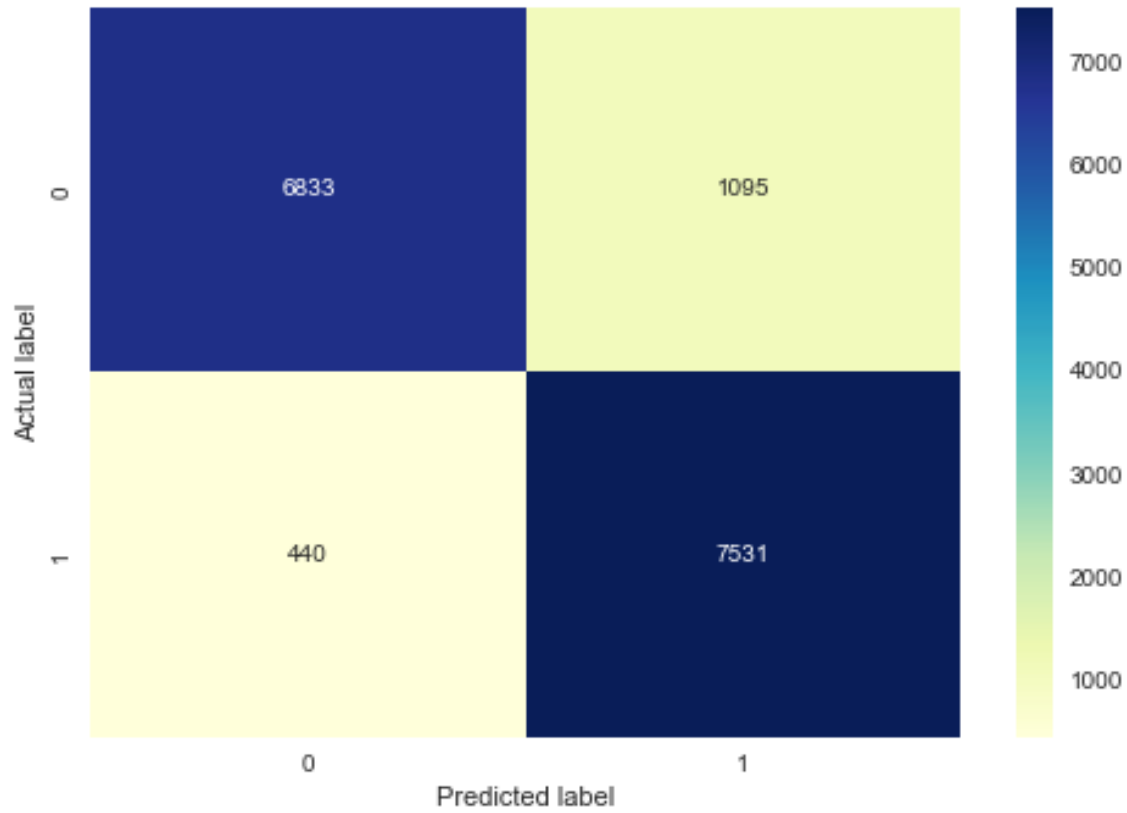


Not Cyberbullying



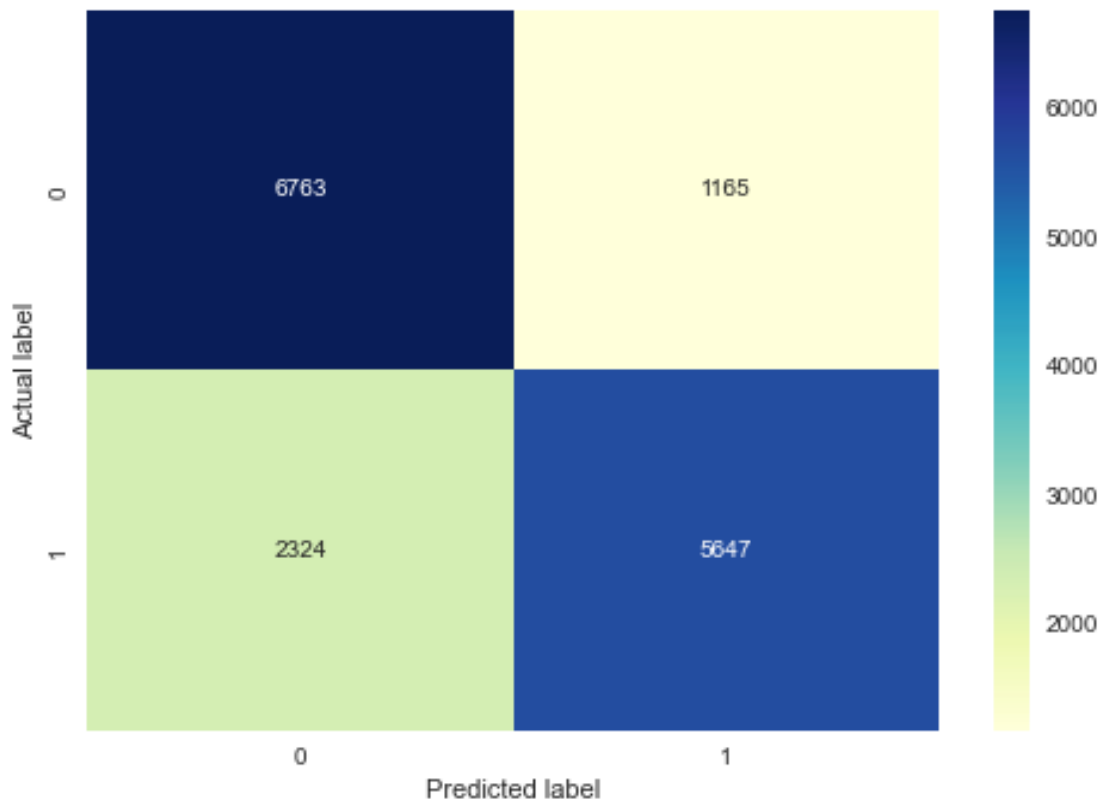
Cyberbullying

Random Forest



	precision	recall	f1-score	support
cyberbullying	0.94	0.86	0.90	7928
not_cyberbullying	0.87	0.94	0.91	7971
accuracy			0.90	15899
macro avg	0.91	0.90	0.90	15899
weighted avg	0.91	0.90	0.90	15899
0.9034530473614693				

Multinomial Naive-Bayes



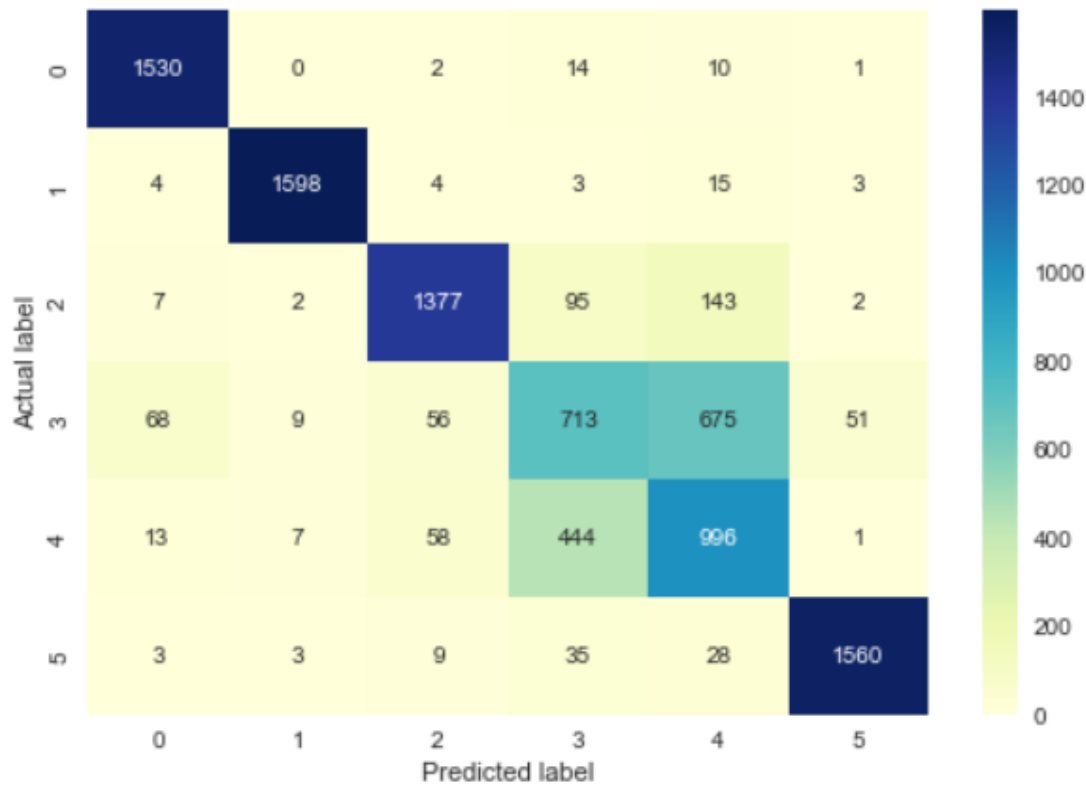
	precision	recall	f1-score	support
cyberbullying	0.74	0.85	0.79	7928
not_cyberbullying	0.83	0.71	0.76	7971
accuracy			0.78	15899
macro avg	0.79	0.78	0.78	15899
weighted avg	0.79	0.78	0.78	15899
0.7805522359896849				

Comparativa

Modelo	F1-Score		Accuracy
	Si	No	
Naive Bayes	0.79	0.76	0.78
Random Forest	0.9	0.91	0.9

Tipo de Cyberbullying

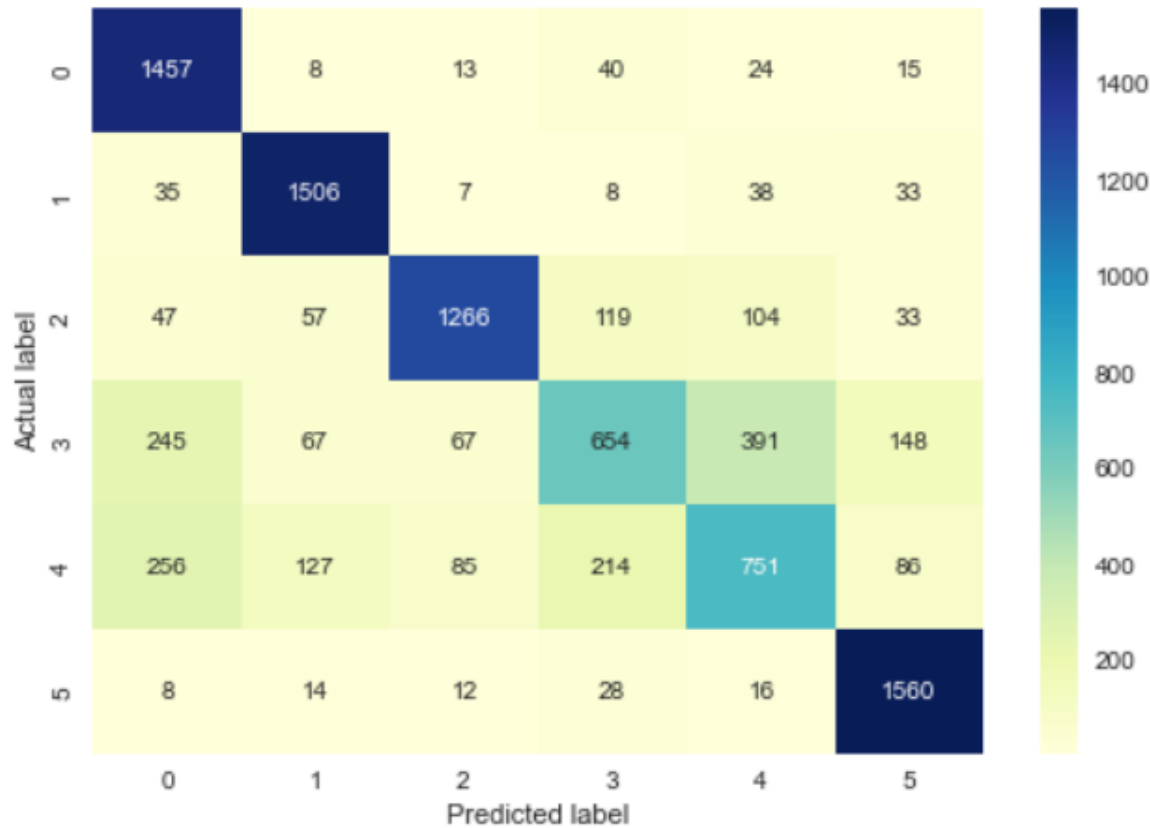
Random Forest



	precision	recall	f1-score	support
age	0.94	0.98	0.96	1557
ethnicity	0.99	0.98	0.98	1627
gender	0.91	0.85	0.88	1626
not_cyberbullying	0.55	0.45	0.50	1572
other_cyberbullying	0.53	0.66	0.59	1519
religion	0.96	0.95	0.96	1638
accuracy			0.81	9539
macro avg	0.81	0.81	0.81	9539
weighted avg	0.82	0.81	0.81	9539

0.8149701226543663

Multinomial Naive-Bayes



	precision	recall	f1-score	support
age	0.71	0.94	0.81	1557
ethnicity	0.85	0.93	0.88	1627
gender	0.87	0.78	0.82	1626
not_cyberbullying	0.62	0.42	0.50	1572
other_cyberbullying	0.57	0.49	0.53	1519
religion	0.83	0.95	0.89	1638
accuracy			0.75	9539
macro avg	0.74	0.75	0.74	9539
weighted avg	0.74	0.75	0.74	9539

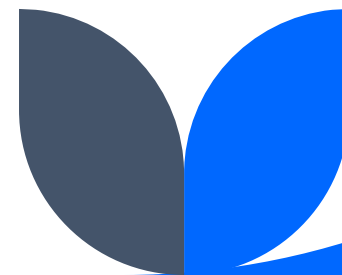
0.7541671034699654

Comparativa

Modelo	F1-Score						Accuracy
	1	2	3	4	5	6	
Naive Bayes	0.81	0.88	0.82	0.5	0.53	0.89	0.75
Random Forest	0.96	0.98	0.88	0.5	0.59	0.96	0.81

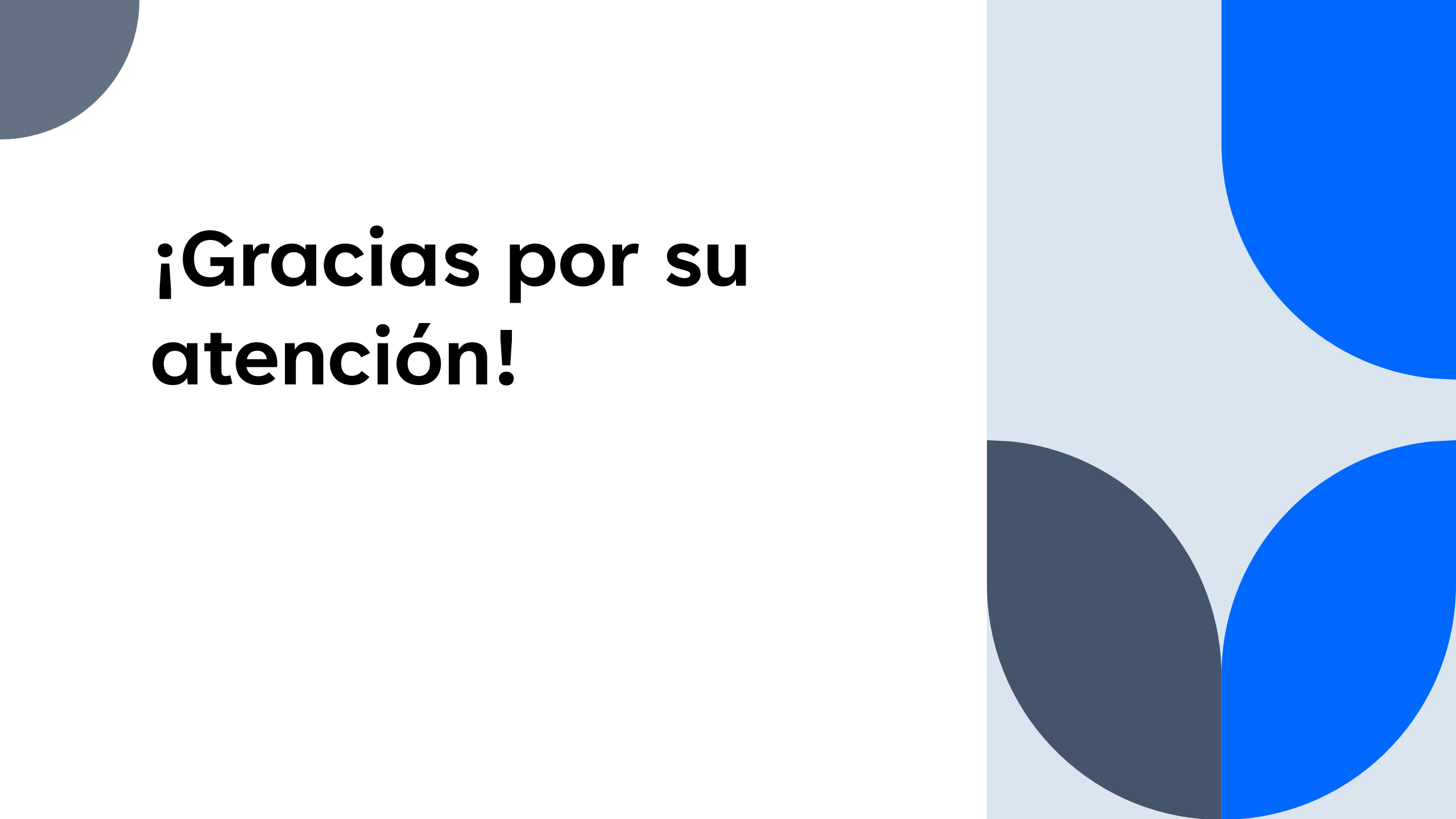
Conclusión

- Al observar los resultados del preprocesamiento de texto cabe destacar mucho las palabras que sobresalen en cada tipo de cyberbullying entre ellos, por ejemplo, en el tipo de ethnicity: nigger, black y dumb son las más frecuentes; en el tipo de religión: muslim, idiot y christian; en el tipo de age: school, high y girl; y en el tipo de gender: joke, rape y gay. Podemos deducir que hay mucho acoso hacia la gente de color, las personas que practican la religión musulmana y cristiana, las mujeres en edad de preparatoria y los gays.
- Analizando esto mismo de manera general, las palabras más frecuentes son school, nigger y joke, por lo que podemos concluir que existe más acoso hacia la gente de color y en las escuelas.
- En las predicciones realizadas se encontraron mejores resultados en el random forest comparado con el multinomial naive-bayes, arrojando una precisión más alta entre ambos.



Referencias

- LARXEL. (2020). Cyberbullying Classification. Junio del 2022, de Kaggle Sitio web: <https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification?datasetId=1869236&sortBy=voteCount>
- Irving Estrada. Github. 2022, Sitio web: <https://github.com/Irving-Estrada/Procesamiento>
- Thair Nu Phyu. (March 2009). Survey of Classification Techniques in Data Mining. Proceedings of the International MultiConference of Engineers and Computer Scientists , I.



**¡Gracias por su
atención!**