

Detección de amenazas cercanas en la ciudad mediante el uso de redes sociales

Daniel Arnedo Nieto

Máster Universitario en
Ingeniería de Telecomunicación

Enero 2023

SMART CITIES

Equipo docente

David Crespo García

Carlos Monzo Sánchez



Contenido

1. Introducción, contexto y motivación del trabajo
2. Estado del arte
3. Diseño del prototipo
4. Análisis, resultados y validación del diseño
5. Conclusiones y trabajos futuros



1. Introducción, contexto y motivación del trabajo

- Información en RRSS desaprovechada
- Posibilidad de alertar a ciudadanos de emergencias locales (dispositivos conectados en ciudades)
- Viabilidad de un sistema escalable que detecte incidencias
- Objetivos:
 - *Extracción de datos reales directamente de la plataforma Twitter*
 - *Aplicar distintos algoritmos de Machine Learning para detectar eventos en redes sociales*



2. Estado del arte

Aplicaciones similares

- *ES-Alert*
- *Cruz Roja Americana* 

Visualización y análisis de los datos

- *Jupyter Notebook* 
- *Google Colab* 
- *Apache Zeppelin* 

Ingesta

- *RapidMiner* 
- *NiFi* 
- *Script Python* 

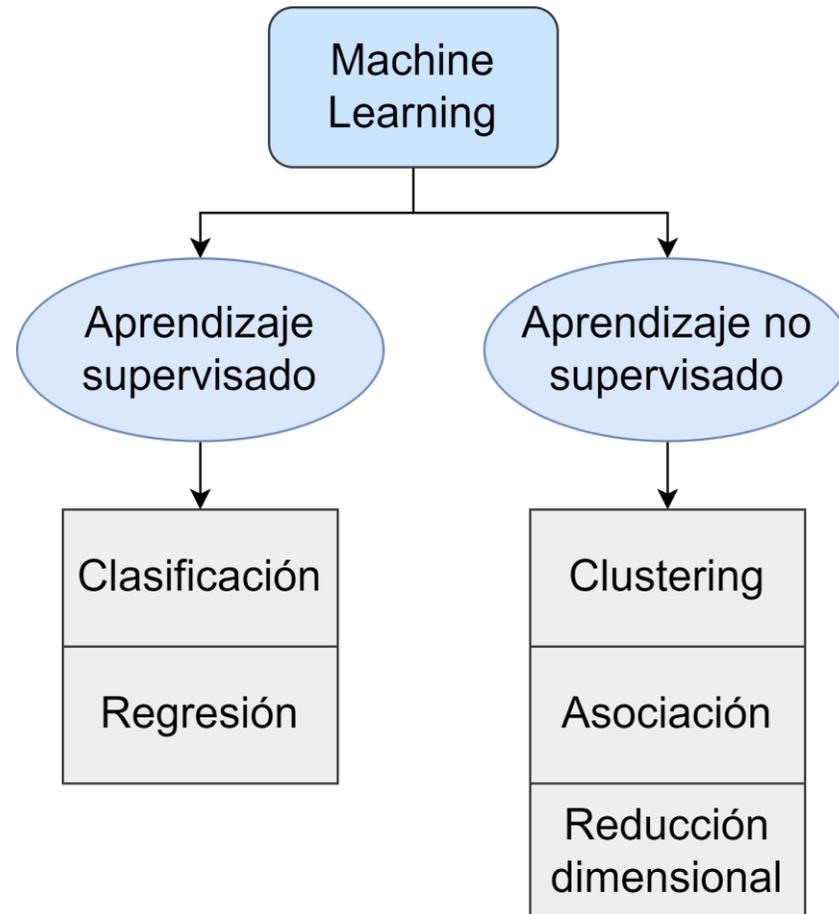
Arquitecturas

- *Cloudera Data Platform* 
- *Amazon Web Services* 
- *Google Cloud* 
- *Microsoft Azure* 



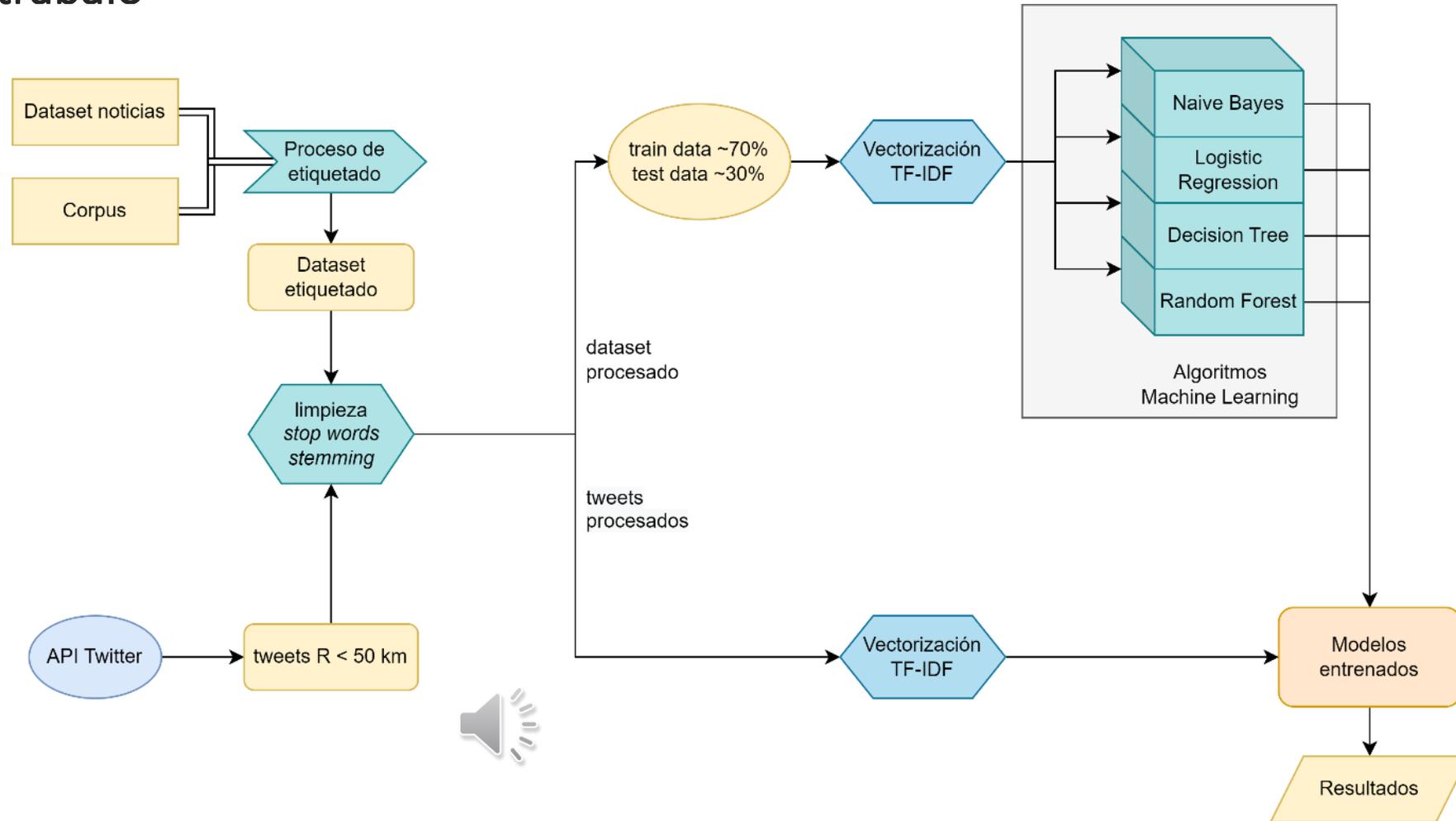
2. Estado del arte

Aplicación de ML en publicaciones de Twitter



3. Diseño del prototipo

Flujo de trabajo



3. Diseño del prototipo

Creación de modelos de ML. Algoritmos.

Naive Bayes

$$P(Y_i|X_1, X_2, \dots, X_N) = \frac{P(Y_i) \cdot P(X_1, X_2, \dots, X_N|Y_i)}{P(X_1, X_2, \dots, X_N)}, \quad i = 1, 2; N = [1, n]$$

Logistic Regression

$$Y = \sigma(Z) = \sigma(X \cdot W) = \sigma\left(\sum (x_i \cdot w_i)\right)$$

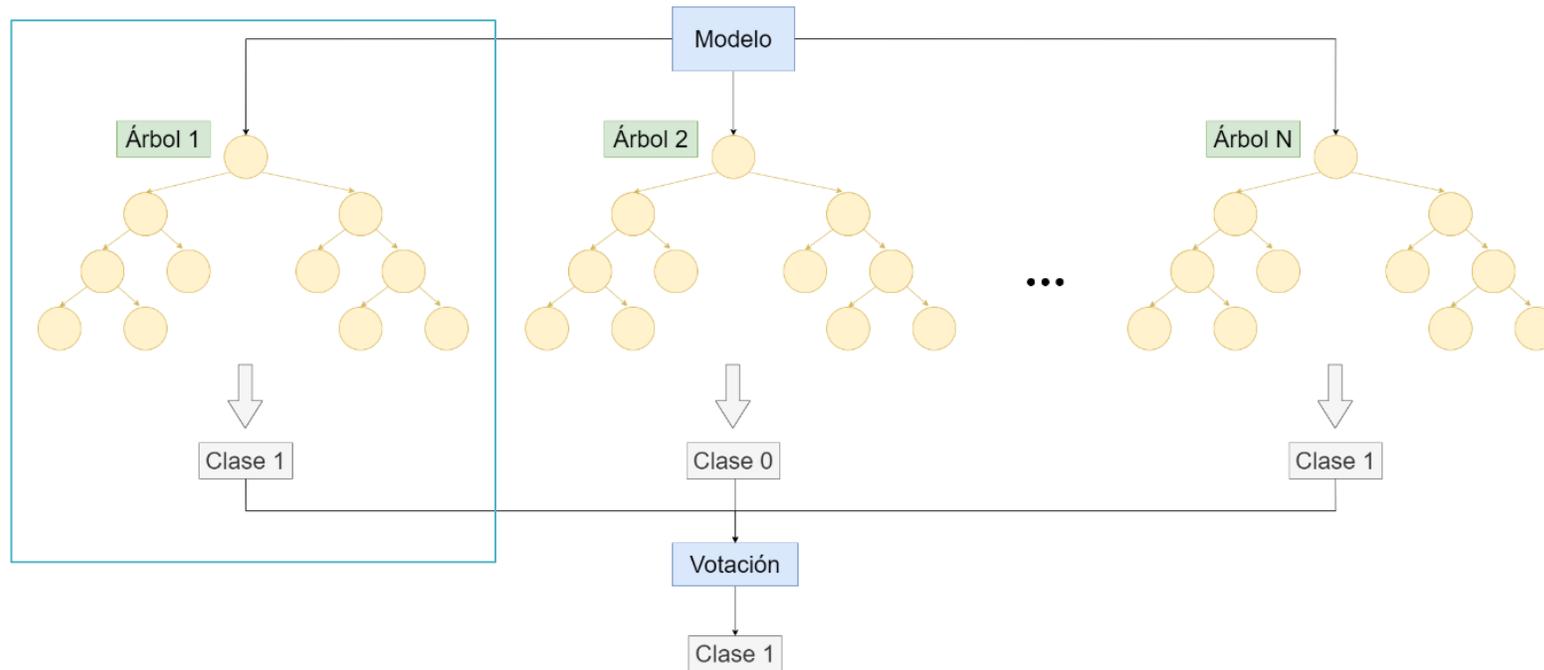


3. Diseño del prototipo

Creación de modelos de ML. Algoritmos.

Decision Tree

Random Forest



3. Diseño del prototipo

Creación de modelos de ML. Resultados.

Algoritmo	Clase	Precisión	Recall	F1-score	Exactitud
Naive Bayes	0	93,08%	85,73%	89,26%	89,73%
	1	86,89%	93,69%	90,16%	
Logistic Regression	0	91,93%	95,77%	93,81%	93,71%
	1	95,63%	91,68%	93,61%	
Decision Tree	0	62,07%	98,55%	76,16%	69,30%
	1	96,55%	40,33%	56,89%	
Random Forest	0	77,46%	96,24%	85,83%	84,19%
	1	95,09%	72,25%	82,11%	

Corpus 1

Corpus 2

Algoritmo	Clase	Precisión	Recall	F1-score	Exactitud
Naive Bayes	0	92,78%	87,59%	90,11%	90,45%
	1	88,40%	93,28%	90,77%	
Logistic Regression	0	99,82%	97,98%	98,89%	98,91%
	1	98,04%	99,82%	98,93%	
Decision Tree	0	98,25%	98,30%	98,27%	98,28%
	1	98,32%	98,27%	98,30%	
Random Forest	0	91,32%	94,20%	92,74%	92,67%
	1	94,09%	91,16%	92,60%	



3. Diseño del prototipo

Creación de modelos de ML. Resultados.

	Clase detectada		Algoritmo
Clase verdadera	25601 (TN)	4260 (FP)	Naive Bayes
	1903 (FN)	28236 (TP)	
Clase verdadera	28598 (TN)	1263 (FP)	Logistic Regression
	2509 (FN)	27630 (TP)	
Clase verdadera	29427 (TN)	434 (FP)	Decision Tree
	17984 (FN)	12155 (TP)	
Clase verdadera	28737 (TN)	1124 (FP)	Random Forest
	8364 (FN)	21775 (TP)	

Corpus 1

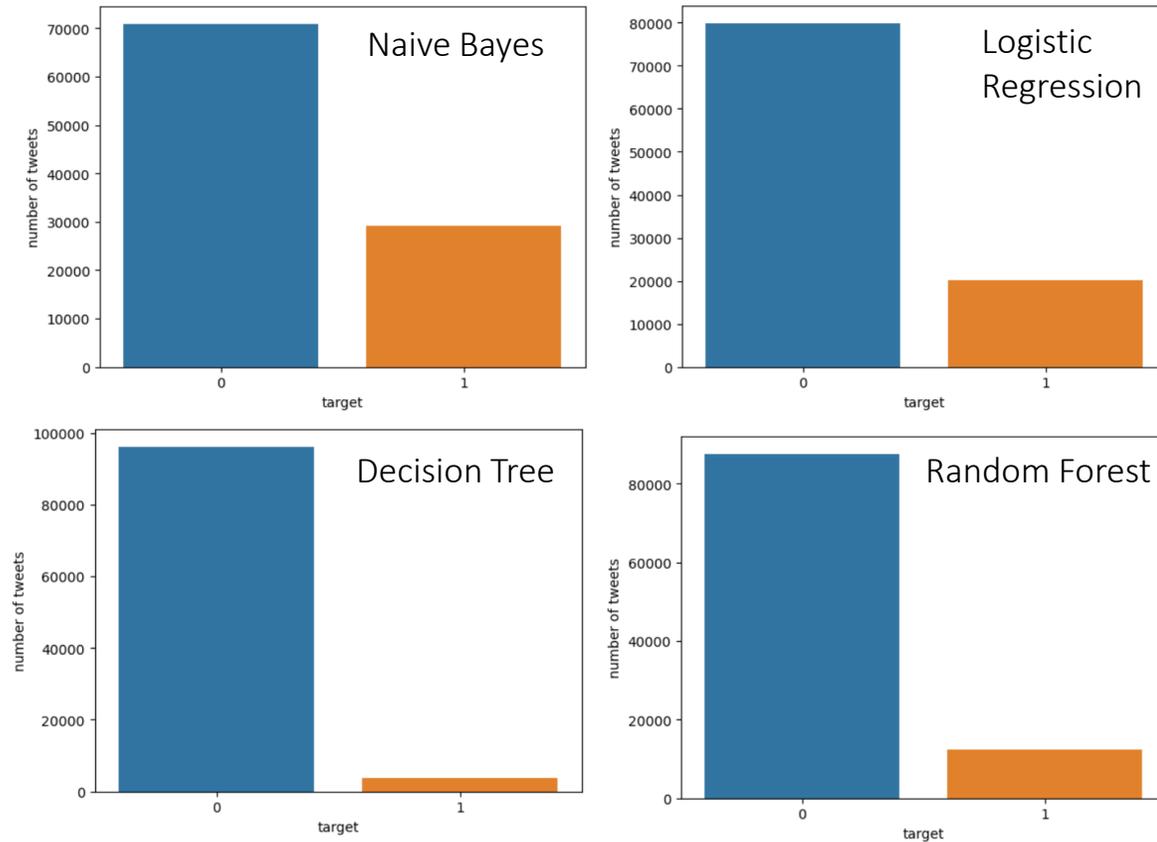
Corpus 2

	Clase detectada		Algoritmo
Clase verdadera	19579 (TN)	2773 (FP)	Naive Bayes
	1523 (FN)	21125 (TP)	
Clase verdadera	21901 (TN)	451 (FP)	Logistic Regression
	40 (FN)	22608 (TP)	
Clase verdadera	21971 (TN)	381 (FP)	Decision Tree
	391 (FN)	22257 (TP)	
Clase verdadera	21055 (TN)	1297 (FP)	Random Forest
	2001 (FN)	20647 (TP)	

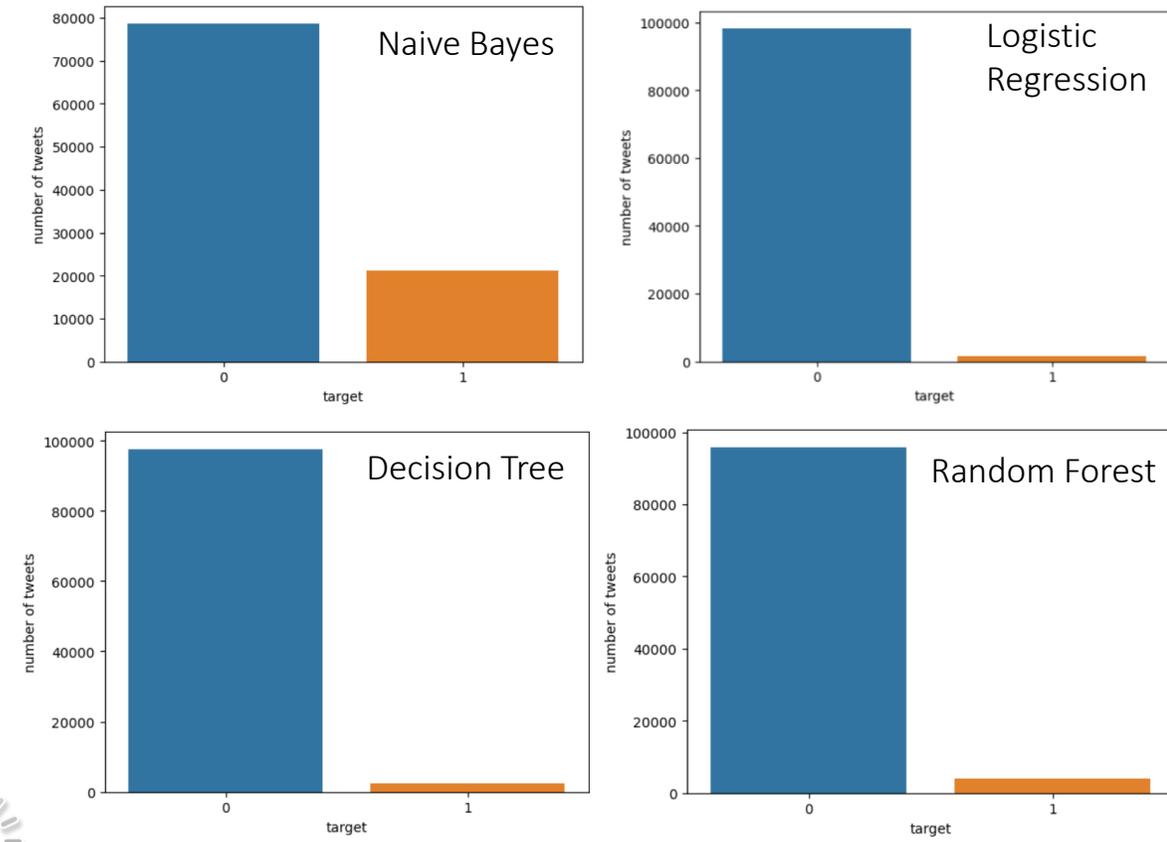


4. Análisis, resultados y validación del diseño

Corpus 1

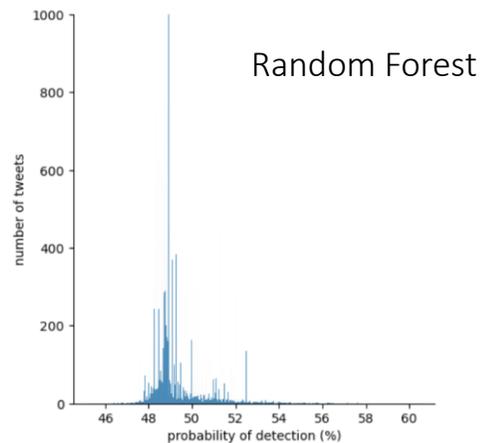
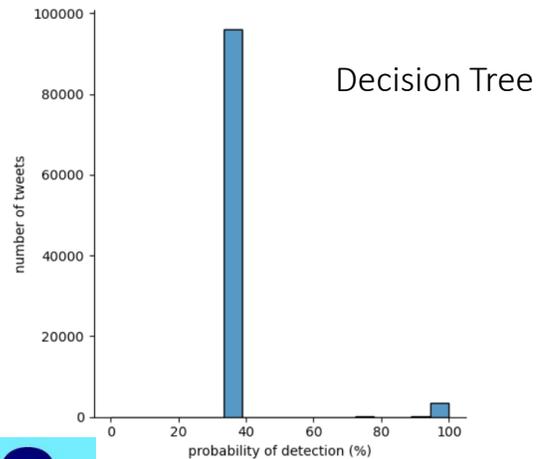
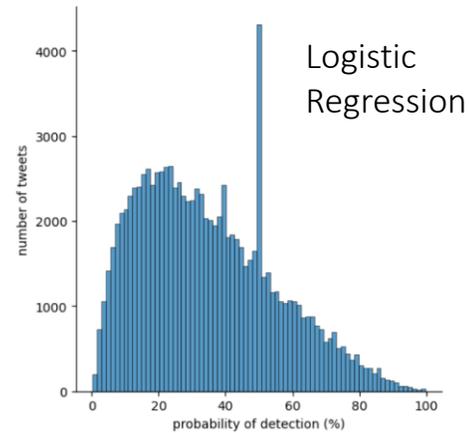
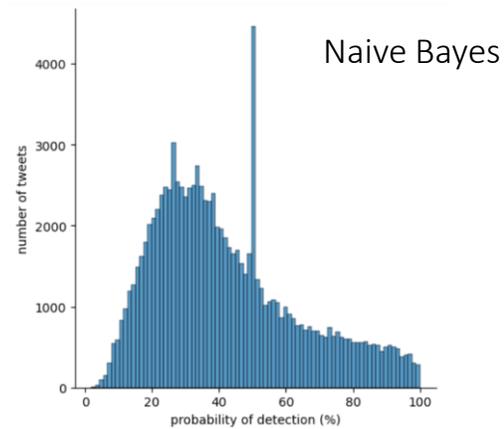


Corpus 2

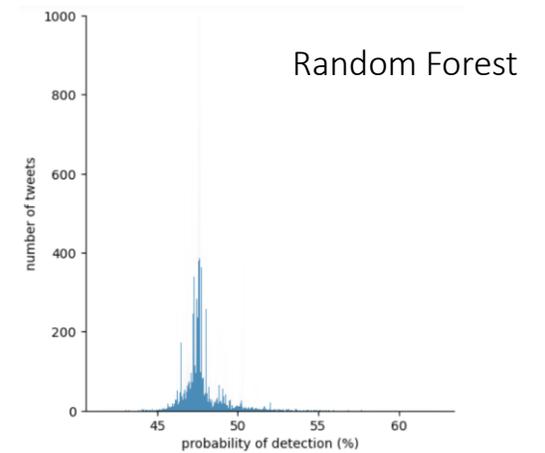
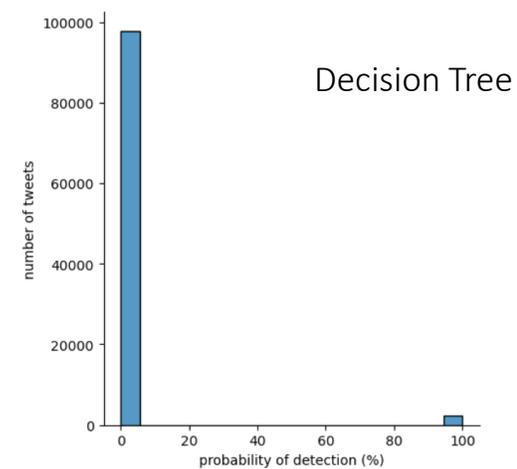
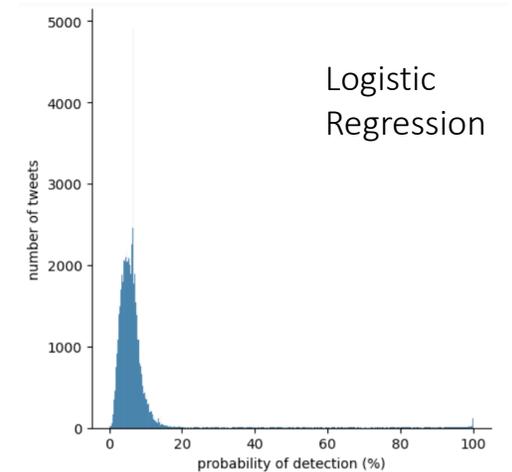
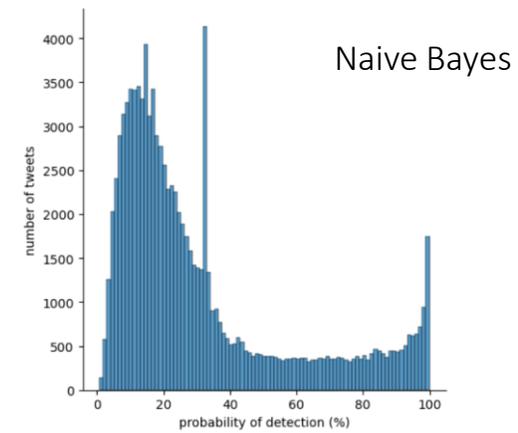


4. Análisis, resultados y validación del diseño

Corpus 1



Corpus 2



4. Análisis, resultados y validación del diseño



Naive Bayes

Jail for the rapist
death sentence
for killers

~99,4%

"Prisión para los violadores, sentencia de muerte para los asesinos"

Tomatoes are
on their way to
murdering me



~99,99%

Logistic Regression

"Los tomates van camino de asesinarme"

District line severe delays between tower hill and ealing
broadway minor delays between turnham green and richmond
while we fix a signal failure at gloucester road good service on
the rest of the line tickets are being accepted on london buses

"Retrasos graves en la línea entre Tower Hill y Ealing Broadway y retrasos menores entre Turnham Green y Richmond mientras que arreglamos un fallo de señal en Gloucester Road. En el resto de la línea el servicio es correcto y los tickets de la línea son aceptados en los buses de Londres"

~100%



Decision Tree



I was just listening to murder case where the
defense attorney argued that a video of a man
the accused choking a woman the murder victim
was no big deal because stuff like that is all over
movies and television

"Estaba escuchando un caso de asesinato donde la defensa alegaba que el vídeo de un hombre acusado por ahogar a una mujer, víctima de asesinato, no era suficiente porque este tipo de cosas se ve a menudo en películas y televisión"

~61,36%

Random Forest



5. Conclusiones y trabajos futuros

- El proyecto es viable pero necesita de más trabajo y especialización
 - Los resultados en ocasiones son aleatorios y están fuertemente influenciados por el set de datos de entrenamiento
 - El dataset utilizado perjudica mucho los resultados y desarrollo del proyecto
 - La ubicación de eventos es compleja pues muchos usuarios deshabilitan la ubicación por privacidad
-
- Trabajo futuro:
 - *Elaboración de algoritmos desde 0*
 - *Estudio de más técnicas de PLN*
 - *Infraestructura orientada a Big Data*
 - *Incorporar algoritmos y técnicas de Deep Learning*



¡MUCHAS GRACIAS
POR SU ATENCIÓN!

