

UNIVERSIDAD DEL VALLE DE GUATEMALA



Predicciones de desastres mediante tweets.

Daniela Villamar, 19086
Diego Crespo, 19541

GUATEMALA, 13 de octubre de 2022

Paleta de colores

Elegimos la paleta de colores de esa manera, celeste y azul, ya que, aparte de ser unos colores llamativos y que agregan color y vida al documento. La aplicación Twitter las cuales están basadas nuestras predicciones, tienen el color celeste como color primario en la aplicación. Las predicciones son en base a desastres mediante tweets, en la aplicación Twitter. Otra razón por la cual consideramos esta paleta como la mas optima, fue debido a los colores de nuestra bandera, el celeste no solamente representa Twitter, también a nuestra patria.

Algoritmos

Naïve Bayes:

Elegimos este algoritmo como herramienta en las predicciones debido a que es un clasificador probabilístico con un fuerte supuesto en la independencia condicional. Este algoritmo fue el mas efectivo de los seleccionados, con un acercado 80% de asertividad, la matriz de confusión de esta herramienta fue la óptima.

LSTM:

La segunda herramienta utilizada fue, LSTM de redes neuronales, Long short-term memory. El algoritmo tiene buen rendimiento en aprendizaje de secuencias, lo cual fue buena opción para este contexto. Esta herramienta nos fue muy asertiva, el rendimiento fue el optimo para el contexto. La matriz de confusión de esta herramienta fue ideal.

Regresión Lineal:

La tercera herramienta seleccionada fue, regresión lineal, este modelo hace un conteo de los vectores con los datos a trabajar. El modelo es discriminatorio, siendo probabilidad condicional. Consideramos que pudiese brindar buenos resultados debido al peso de cada palabra o "token". Desafortunadamente, este modelo no resulto tan efectivo como se esperaba, al observar la matriz de confusión logramos detectar que esta herramienta es la menos optima de las tres en este contexto.

Orden de las predicciones

Elegimos este orden ya que, inicialmente se coloca el título de las predicciones, para que el lector se de una idea muy básica respecto a lo que leerá. Continuando con la cantidad de tweets que se utilizaron para realizar las predicciones, junto con, las

palabras mas repetidas en el dataset. Colocamos esto continuo al titulo ya que consideramos que es la información más básica que se puede brindar previo a los algoritmos y resultados como tal.

Después colocamos los modelos de predicción que fueron utilizados para realizar las predicciones, esto es necesario para entender que herramientas fueron utilizadas para obtener los resultados, el orden fue descendiente respecto a su efectividad y precisión en los resultados, para tener al final el modelo mas asertivo con el mayor tamaño posible.