Google ha anunciado un innovador vectorizador de texto multilingüe llamado RETVec (acrónimo de Resilient and Efficient Text Vectorizer) diseñado para identificar contenido potencialmente perjudicial, como correos no deseados y mensajes electrónicos maliciosos en Gmail.

Según la descripción del proyecto en GitHub, *«RETVec está capacitado para resistir manipulaciones a nivel de caracteres, tales como inserciones, eliminaciones, errores tipográficos, homoglifos, sustituciones LEET y más».*

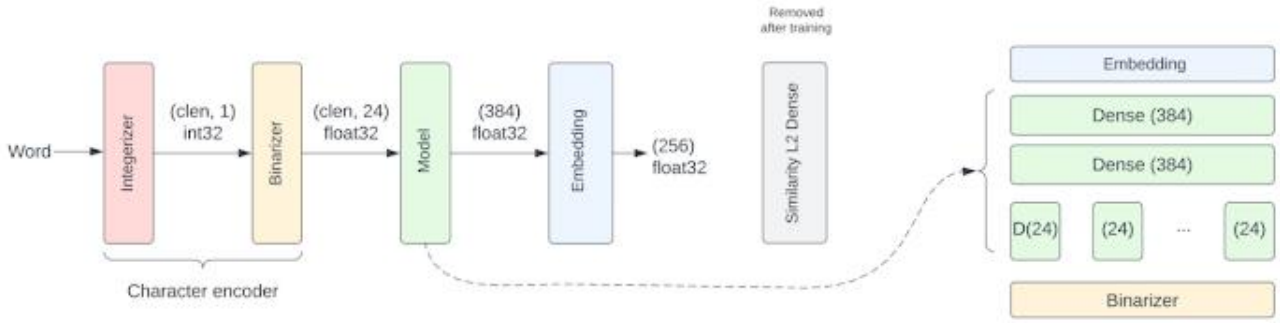
> *«El modelo RETVec se ha entrenado sobre un codificador de caracteres novedoso capaz de codificar eficientemente todos los caracteres y palabras UTF-8».*

A pesar de que plataformas de gran envergadura como Gmail y YouTube confían en modelos de clasificación de texto para detectar ataques de phishing, comentarios inapropiados y estafas, se sabe que los actores de amenazas idean estrategias para eludir estas medidas de defensa.

Estos actores a menudo recurren a manipulaciones adversas de texto, que van desde el uso de homoglifos hasta el relleno de palabras clave e incluso la inserción de caracteres invisibles.

RETVec, que es compatible con más de 100 idiomas de manera predeterminada, tiene como objetivo contribuir a la creación de clasificadores de texto más resilientes y eficientes, tanto en servidores como en dispositivos, al tiempo que se vuelven más robustos y eficientes.

La vectorización es una metodología en el procesamiento del lenguaje natural (NLP) que asigna palabras o frases de vocabulario a una representación numérica correspondiente, permitiendo así análisis adicionales, como análisis de sentimientos, clasificación de texto y reconocimiento de entidades nombradas.



> *«Gracias a su arquitectura innovadora, RETVec funciona de manera eficiente en todos los idiomas y caracteres UTF-8 sin necesidad de procesamiento de texto previo, convirtiéndolo en la opción ideal para implementaciones de clasificación de texto en dispositivos, web y a gran escala»*, destacaron Elie Bursztein y Marina Zhang de Google.

La integración de este vectorizador en Gmail, según el gigante tecnológico, mejoró la tasa de detección de spam en un 38% en comparación con el valor base y redujo la tasa de falsos positivos en un 19,4%. Además, disminuyó el uso de la Unidad de Procesamiento Tensorial (TPU) del modelo en un 83%.

> Bursztein y Zhang añadieron: *«Los modelos entrenados con RETVec muestran una velocidad de inferencia más rápida gracias a su representación compacta. La reducción del tamaño de los modelos no solo disminuye los costos computacionales, sino que también reduce la latencia, aspecto crítico para aplicaciones a gran escala y modelos en dispositivos».*