

# Clasificación de texto de clase múltiple con aprendizaje profundo utilizando BERT

Daniel Oswaldo Mora Malaver



## Resumen

La empresa HOLCREST S.A.S. tiene un problema y es que cada correo que llega a su poder llega sin clasificar y la tarea de clasificar estos correos es llevado manualmente. Los países de los que Holcim recibe correos son Ecuador, Argentina, El salvador, Nicaragua, Colombia, Costa rica y México. Las siete categorías a las que son clasificados los correos son: Aplicación de pago y pedido, No cliente, Datos maestro, Pedido, Otros, Aplicación de pago, y Logística. Para dar solución a este problema se diseñaron e implementaron varios modelos de clasificación de texto.

## Marco Teórico.

### Modelo Bert

BERT es una técnica basada en redes neuronales para el preentrenamiento del procesamiento del lenguaje natural (PLN). BERT es una representación de lenguaje bidireccional (es decir que se juzgan las palabras ubicadas a la izquierda y a la derecha de cada uno de los términos),

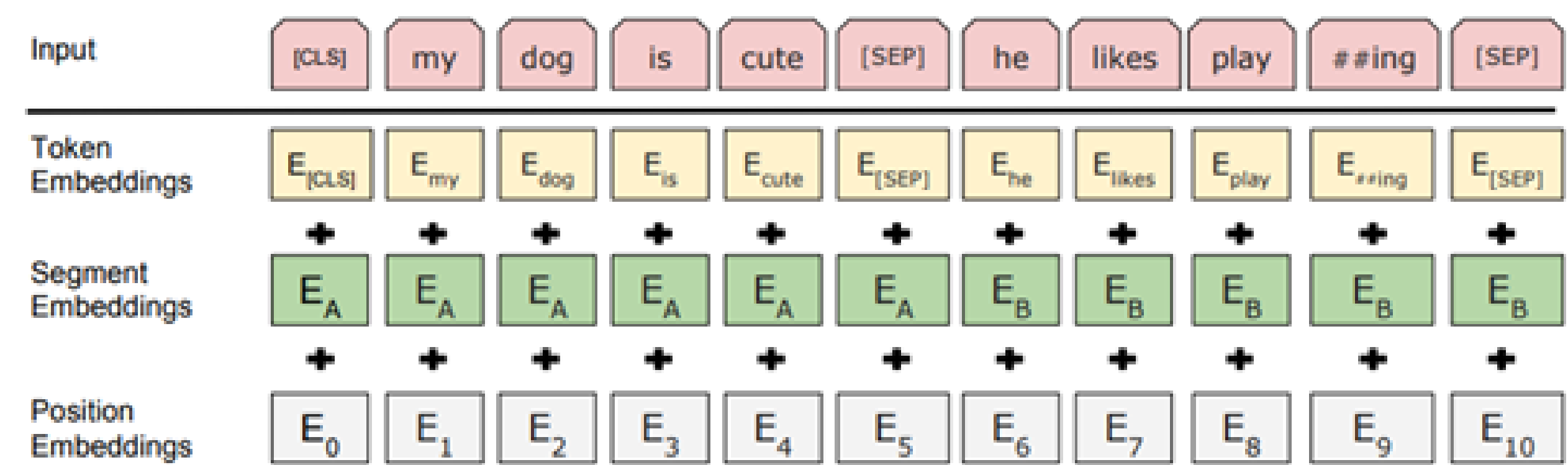


Figura: BERT-Input representación

## Resultados

Se entreno un modelo Bert para cada país y se obtuvieron los siguientes resultados.

País	F1 Score	Precisión validación
Ecuador	0.92	0.93
México	0.88	0.87
Nicaragua	0.88	0.89
El salvador	0.81	0.81
Argentina	0.89	0.90
Colombia	0.89	0.89
Costa rica	0.92	0.90

Además identificamos que cuando un correo queda mal clasificado, la segunda probabilidad mas alta, es un valor muy elevado en comparación cuando el correo queda bien clasificado. Por eso establecimos un umbral para cada país, todo correo cuya segunda probabilidad mas alta, supere este valor será considerado clasificado en una nueva categoría llamada “indefinido”.

## Conclusiones

- Bert es una herramienta muy útil para las tareas relacionadas con la clasificación de texto, permite relacionar las palabras con el contexto que las rodea, gracias a esto, es capaz de generar una representación consistente del texto que esta analizando.
- Casi todos los modelos tuvieron un muy buen rendimiento, a excepción del modelo de el salvador.
- AdamW es una variante del optimizador Adam que tiene una implementación mejorada de disminución de peso. El uso de la disminución del peso es una forma de regularización para reducir la posibilidad de sobreajuste.

## Transformers

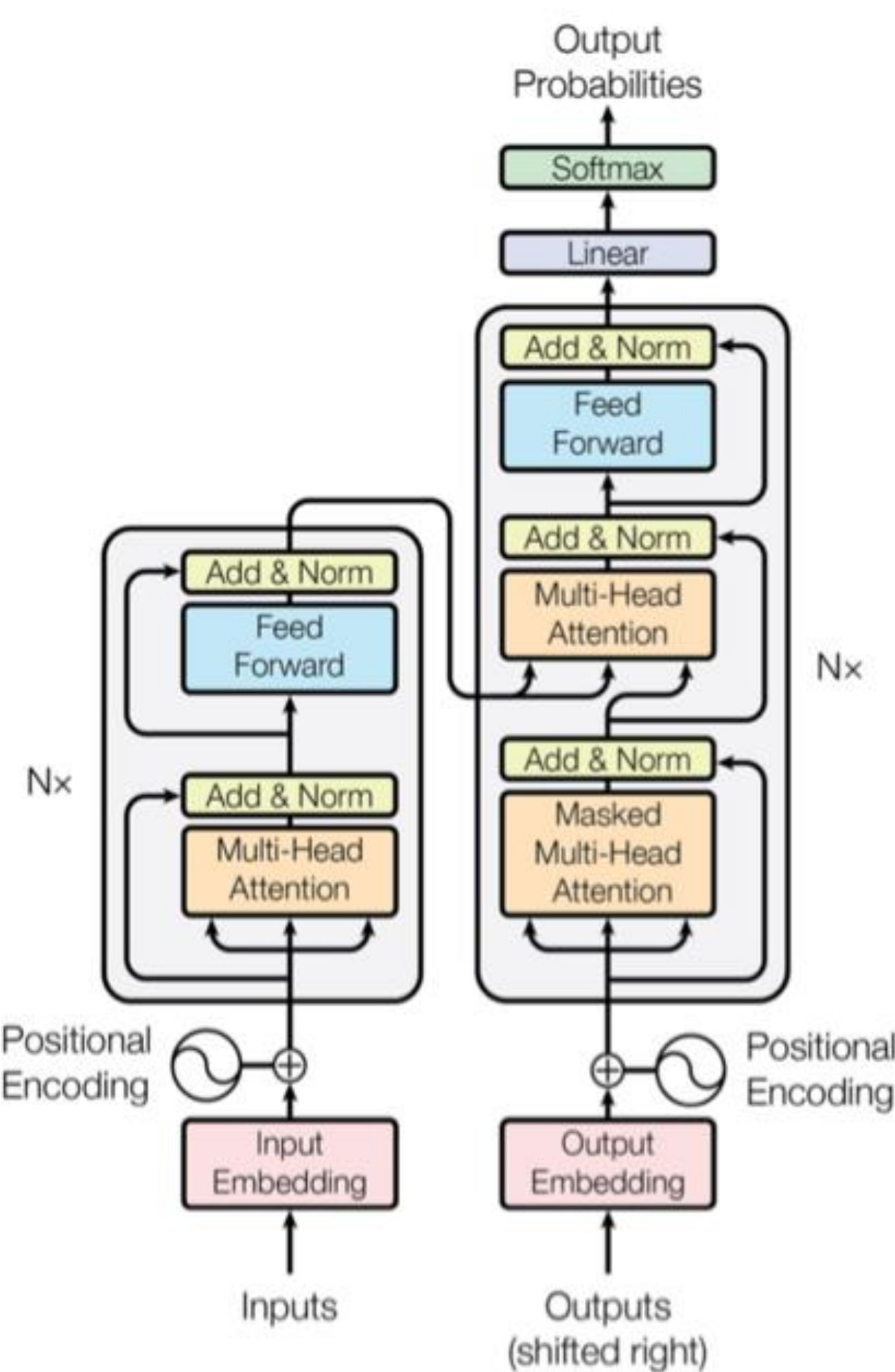


Figura. El Transformer modelo de arquitectura

Los Transformers son modelos de aprendizaje profundo, una arquitectura de red neuronal, diseñados para manipular datos secuenciales principalmente en el campo del procesamiento de lenguaje natural (NLP); traducción, predicción de texto, summarización, o bien en el campo de las series de tiempo, forecasting y/o trading.

## Referencias

[1] Dr. Dataman, 2018, Looking into Natural Language Processing (NLP). <https://medium.com/datamann-ai/natural-language-processing-nlp-forelectronic-health-record-ehr-part-i-4cb1d4c2f24b>

[2] J. Howard and S. Ruder, Universal language model fine-tuning for text classification, arXivpreprint arXiv:1801.06146, (2018).