

# Reporte Text Classification Using Transformer Networks (BERT)

## Clasificacion Multiclase

Adrian Pineda Sanchez A00834710

### Comparación entre BERT y Regresión Logística en Clasificación de Textos

#### Introducción

En esta comparación, se evaluaron dos enfoques diferentes para una tarea de clasificación de textos en cuatro categorías: "World", "Sports", "Business" y "Sci/Tech". El primer enfoque utiliza BERT, un modelo basado en transformadores preentrenado para tareas de procesamiento de lenguaje natural. El segundo enfoque utiliza un modelo de Regresión Logística.

#### Métricas de Evaluación

Las métricas clave consideradas para comparar ambos modelos fueron:

- Precision: Proporción de instancias correctamente clasificadas entre las clasificadas como positivas.
- Recall: Proporción de instancias correctamente clasificadas entre todas las verdaderas instancias positivas.
- F1-score: Promedio armónico entre precision y recall.
- Accuracy: Proporción de instancias correctamente clasificadas sobre el total.

---

#### Resultados

##### Modelo BERT

Categoría	Precision	Recall	F1-Score	Support
World	0.96	0.95	0.96	1900
Sports	0.99	0.99	0.99	1900
Business	0.93	0.92	0.92	1900
Sci/Tech	0.92	0.93	0.92	1900
Accuracy	0.95			7600

### Modelo de Regresión Logística

Categoría	Precision	Recall	F1-Score	Support
World	0.86	0.91	0.89	1900
Sports	0.95	0.96	0.95	1900
Business	0.85	0.85	0.85	1900
Sci/Tech	0.88	0.83	0.85	1900
Accuracy	0.89			7600

---

### Análisis de Resultados

#### Precisión y Recall

- El modelo BERT mostró una superioridad significativa en todas las métricas de evaluación. En particular, para las categorías "World" y "Business", BERT supera por un margen considerable al modelo de Regresión Logística, con un F1-score de 0.96 y 0.92 respectivamente, frente a 0.89 y 0.85 en el modelo de Regresión Logística.
- El desempeño más notable se da en la categoría "Sports", donde BERT logra un F1-score casi perfecto de 0.99, comparado con 0.95 en Regresión Logística.

#### Promedio Macro y Ponderado

- En el promedio ponderado y macro, que consideran el rendimiento global del modelo, BERT obtiene consistentemente un puntaje de 0.95, mientras que Regresión Logística queda en 0.89.

#### Accuracy

- En cuanto a exactitud global, BERT alcanza un 95%, frente al 89% de Regresión Logística, una mejora del 6%.

---

### Conclusión

La superioridad de BERT frente a Regresión Logística se debe a su capacidad para capturar mejor las relaciones semánticas complejas y el contexto en los datos de texto, gracias a su arquitectura de transformadores preentrenados. Esto lo convierte en una elección ideal

para tareas de clasificación de textos que requieren un entendimiento profundo del lenguaje.

### **Ventajas de Cada Enfoque**

- **BERT:**
  - Excelente en tareas que requieren una comprensión profunda del contexto.
  - Manejo de relaciones semánticas complejas en textos largos.
- **Regresión Logística:**
  - Menor costo computacional y tiempos de entrenamiento más cortos.
  - Adecuado para tareas con textos más simples o cuando se necesita rapidez.

Recomendación: Para tareas de clasificación de texto que exigen alta precisión y comprensión contextual, BERT es una opción superior a la Regresión Logística, aunque con un mayor costo computacional. Si la prioridad es la eficiencia en tiempo y recursos, los modelos tradicionales siguen siendo una alternativa válida.