

**Instituto Tecnológico y de Estudios  
Superiores de Monterrey**  
Campus Monterrey



**Materia**

Inteligencia artificial avanzada para la ciencia de datos II

**Tarea**

Cloud Computing | Actividad - BERT notebooks

**Estudiante**

Cleber Gerardo Pérez Galicia - A01236390

**Profesor**

Alfredo Esquivel Jaramillo

## **Pipeline Act 1**

### **- Configuración inicial**

Se definen las librerías necesarias, se configura el gpu para que el procesamiento de los datos sea mas rapido y se modifica la semilla inicial a 1122.

### **- Preprocesamiento de datos**

Se leen los datos de los archivos. Se realiza una division del conjunto de entrenamiento y validación del 90-10.

### **- Tokenización**

Se utiliza un tokenizer preentrenado de BERT para convertir el texto en tokens manejables por el modelo.

### **- Definición del modelo**

Se implementa la clase de BertForSequenceClassification y se incorpora una capa de clasificacion con una función de perdida CrossEntropyLoss para calcular el error de las predicciones.

### **- Configuración del entrenamiento**

Se definen parametros de entrenamiento con epochs, batch y weights. Se configura la funcion de Trainer de HuggingFace para manejar el proceso de entrenamiento, evaluación y predicción.

### **- Entrenamiento y evaluación**

Se entrena y se evalua, se define la funcion de compute\_metrics para calcular métricas de rendimiento, como la exactitud.

### **- Predicción y análisis**

El conjunto de pruebas se tokeniza y se procesa igual que los datos de entrenamiento y por ultimo se realizan predicciones con el conjunto de prueba y se calcula un reporte con las metricas de precision, recuperación y F1-Score.

Tabla Comparativa (Act 2)

