AI-Based ✅ TCFD Disclosure Classification

This project focuses on automating the classification of corporate disclosure texts into five TCFD categories:
Irrelevant, Metrics, Strategy, Risk, and Governance

AI SOLUTIONS USED:

✅ **Classical Machine Learning**

Model: TF-IDF + Logistic Regression
- Preprocessed via stop word removal, punctuation, lemmatization
- Vectorized text into numerical features
- Trained a Logistic Regression model

✅ **Transformer-Based Model (BERT)**

Model: bert-base-uncased from Hugging Face
- Tokenized text using BERT tokenizer
- Fine-tuned the BERT model for classification
- Adjusted learning rate, batch size, and epochs

HOW DO WE MEASURE MODEL PERFORMANCE?

Used to measure performance:
- **Accuracy**: Percentage of correct predictions
- **Weighted F1 Score**: Accounts for class imbalance
- **Macro F1 Score**: Treats all categories equally, regardless of size

WHICH MODEL SHOULD YOU USE?

- **Classical Machine Learning:** Fast, simple, and effective for general classification tasks using basic text features.
- **Transformer-Based Model (BERT)** – Captures deeper meaning and context, ideal for complex or domain-specific language like finance and climate disclosures.

Our board oversees climate policies and reports progress quarterly.

Predicted Category:
Governance — Oversight and sustainability leadership