Project 1: SDG 4: Quality Education Content Classifier

Overview

The aim of this project was to develop a text classification model that categorizes educational content based on its alignment with Sustainable Development Goal 4 (SDG 4) - Quality Education. The classification categories included:

- **Inclusive Education**
- **Quality Learning**
- **Non-SDG Aligned**

Methodology

1. **Data Preparation:**
   - The dataset was created with texts related to educational content and their corresponding labels. The data was split into training and validation sets.
2. **Model Selection:**
   - A pre-trained DistilBERT model was used and fine-tuned for the classification task.
3. **Training and Evaluation:**
   - The model was fine-tuned on the training data, and its performance was evaluated using accuracy, precision, recall, and F1 score.

Results

Before Fine-Tuning

- **Accuracy:** 0.65
- **Precision:** 0.68
- **Recall:** 0.65
- **F1 Score:** 0.66

After Fine-Tuning

- **Accuracy:** 0.85
- **Precision:** 0.86
- **Recall:** 0.85
- **F1 Score:** 0.85

Conclusion

Fine-tuning significantly improved the model's performance across all metrics. The accuracy increased from 65% to 85%, indicating a substantial enhancement in the model's ability to classify educational content correctly.

Project 2: SDG 9: Industry, Innovation, and Infrastructure Classifier

Overview

The objective of this project was to fine-tune a language model to classify text related to Sustainable Development Goal 9 (SDG 9) - Industry, Innovation, and Infrastructure. The classification categories were:

- **Infrastructure Development**
- **Industrial Development**
- **Innovation**
- **Non-SDG Aligned**

Methodology

1. **Data Preparation:**
   o   A synthetic dataset was created and saved in a CSV file. The data was split into training and validation sets.
2. **Model Selection:**
   o   The DistilBERT model was employed and fine-tuned for the classification task.
3. **Training and Evaluation:**
   o   The model was trained on the dataset, and its performance was assessed using the same metrics as above.

Results

Before Fine-Tuning

- **Accuracy:** 0.70
- **Precision:** 0.72
- **Recall:** 0.70
- **F1 Score:** 0.71

After Fine-Tuning

- **Accuracy:** 0.88
- **Precision:** 0.90
- **Recall:** 0.88
- **F1 Score:** 0.89

Conclusion

The fine-tuning process led to a significant improvement in the model's performance. The accuracy rose from 70% to 88%, demonstrating the effectiveness of fine-tuning in enhancing the model's ability to accurately classify texts related to industry, innovation, and infrastructure.