Performance Evaluation Report

1. Introduction

This report presents the evaluation of different machine learning models trained for text-pair classification.

The models are assessed based on accuracy, precision, recall, and F1-score. The results highlight the strengths

and weaknesses of each model in handling classification tasks.

2. Evaluation Metrics

The models were evaluated using the following metrics:

- Accuracy: Measures the percentage of correctly classified instances.
- Precision: Measures how many predicted positive labels were actually correct.
- Recall: Measures how many actual positive labels were correctly identified.
- F1-Score: Harmonic mean of precision and recall, balancing both metrics.

3. Model Evaluation Results

The models achieved the following accuracy results:

Model	Accuracy (%)
Model 1	36.64
Model 2	36.52
Model 3	38.13
Model 4	32.95

4. Confusion Matrix

A confusion matrix was used to analyze the performance of the models.

It helps in visualizing the distribution of correct and incorrect predictions.

The ideal model should have high values along the diagonal, indicating correct classifications.

5. Future Improvements

To enhance the model's performance, the following steps can be considered:

- Increasing the dataset size and diversity.

- Fine-tuning hyperparameters for better generalization.

- Exploring advanced NLP models such as transformer-based architectures.

- Reducing class imbalance using resampling techniques.