**DEEP LEARNING - ASSIGNMENT 2**

**20EG106116 (AI A)**

**SATHVIK KANDADI**

**Title: Sentiment Analysis for Social Media Posts Using RoBERTa**

**Summary:**

Sentiment analysis on social media posts is essential for understanding public opinion and trends. In this project, we harnessed the power of RoBERTa, a pre-trained language model, to develop an accurate sentiment analysis tool using a vast dataset of social media posts. The fine-tuned RoBERTa model exhibited exceptional performance in classifying social media posts into categories like positive, negative, and neutral sentiments. Evaluation on a test set with metrics like accuracy, precision, recall, and F1-score underscored its effectiveness in gauging public sentiment. This project showcases the potential of pre-trained language models like RoBERTa to provide valuable insights into online discourse and public sentiment.

**Introduction:**

**Problem Statement:**

Understanding sentiment in social media posts is crucial for businesses, policymakers, and researchers. In this project, we leverage RoBERTa, a pre-trained language model, on a diverse dataset of social media posts to build an advanced sentiment analysis tool. The goal is to create a robust model capable of classifying social media content into categories like positive, negative, and neutral sentiments, offering valuable insights into public opinion.

**Dataset:**

The dataset consists of social media posts from various platforms, each labeled with its corresponding sentiment category: positive, negative, or neutral. The dataset encompasses a wide range of topics and user-generated content.

**Methodology:**

**Data Pre-processing:**

- Text Pre-processing: Standardize text data by removing special characters, converting to lowercase, and tokenizing.

- Label Encoding: Encode sentiment categories (positive, negative, neutral) into numerical values.

**Model Selection:**

- Select RoBERTa as the foundational model for text classification tasks, given its exceptional performance in natural language understanding. Fine-tune RoBERTa on the labeled social media dataset.

**Fine-Tuning:**

- Fine-tuning the RoBERTa model for sentiment analysis involves adapting the model's output to predict one of the three sentiment categories (positive, negative, neutral). The model leverages its pre-trained language understanding capabilities to accurately classify social media content.

**Results:**

**Training:**

- Monitor model performance on historical data, considering metrics like accuracy, precision, recall, and F1-score. Ensure the model effectively classifies social media posts based on their sentiment.

**Testing:**

- Evaluate the fine-tuned RoBERTa model on a held-out test dataset to measure its effectiveness in sentiment analysis. Employ metrics such as accuracy, precision, recall, and F1-score.

**Comparison with State-of-the-Art Models:**

- Compare the performance of the RoBERTa-based sentiment analysis with other state-of-the-art sentiment analysis models. Consider metrics, computational resources, and practicality in real-world scenarios.

**Conclusion:**

This project illustrates the potential of leveraging pre-trained language models like RoBERTa for sentiment analysis on social media data. The model generates accurate and contextually relevant sentiment classifications for diverse social media posts, showcasing its capability to provide valuable insights into online discourse and public sentiment. This approach opens up possibilities for applications like understanding public opinion, social listening, and brand reputation management.