Project Report: Sentiment Analysis with Deep Learning using BERT

1. Project Overview

The objective of this project was to build a **Sentiment Analysis model** using a **pre-trained BERT transformer** and apply it on a **multi-class text classification** problem.
The model is fine-tuned using the **SMILE annotation dataset** and developed using the **PyTorch** framework.

---

2. Data Processing and Preparation

- **Dataset**: The SMILE annotation dataset was used, which contains text samples labeled across multiple sentiment categories.
- **Tokenizer**: A pre-trained **BERT tokenizer** (bert-base-uncased) was used to tokenize the text into input embeddings.
- **Encoding**: Text data was encoded into input IDs and attention masks suitable for feeding into the BERT model.
- **DataLoader**: Separate **train** and **validation** loaders were created to batch the data during training and evaluation.

---

3. Model and Training Setup

- **Model**: A pre-trained **BERT model** was used and fine-tuned by adding a classification head on top for multi-class prediction.
- **Loss Function**: CrossEntropyLoss was used, appropriate for multi-class classification tasks.
- **Optimizer**: **AdamW** optimizer was used for training.
- **Learning Rate Scheduler**: A linear scheduler with warm-up steps was used to adjust the learning rate during training.

---

4. Training and Evaluation

- A **training loop** was designed to fine-tune the BERT model for several epochs.
- The training process included:
  - Forward pass
  - Backward pass (gradients computation)
  - Optimizer step
  - Scheduler step
- **Evaluation metrics** like accuracy were computed after each epoch.
- The **fine-tuned model** showed good accuracy on the validation set, confirming successful learning and generalization.

---

5. Results

- The fine-tuned BERT model achieved strong performance on the sentiment classification task.
- It demonstrated the effectiveness of transfer learning using large-scale language models even on relatively small datasets.

---

6. Conclusion

This project successfully demonstrates:

- How to leverage **pre-trained BERT** models for downstream NLP tasks like **sentiment analysis**.
- Best practices for fine-tuning, including **data preprocessing, optimization, and scheduling**.
- The potential of deep learning approaches to significantly improve accuracy in natural language understanding tasks.