ML HACKATHON

DELIVERABLES DOCUMENT

AAKANKSH SEELIN – PES2UG22CS003

ABHINAV S DASU – PES2UG22CS017

ADVAITH B-PES2UG22CS038

AJAYBIR SINGH HARI-PES2UG22CS042

**Data Preprocessing Steps:**

* Loaded the training and test datasets from CSV files with appropriate encoding (ISO-8859-1).
* Checked for and handled missing values in the 'Utterance' and 'Sentiment' columns by dropping rows with missing data.
* Performed some basic data exploration, such as printing the first few rows and checking for missing values
* Splitting Data: Divided the dataset into training, validation, and test sets .
* Data Augmentation: Augmented video data using techniques to generate synthetic samples for greater model robustness.

**Feature Extraction Steps**:

* Video Features: Used Resnet50 to extract frame embeddings, aggregated to represent video clips.
* Text Features: Employed transformer models BERT) for subtitle embedding generation.
* Combined Features: Merged video and text features using late fusion techniques.

**Model Decisions:**

* The code uses a pre-trained BERT-based sentiment analysis model from the Transformers library to predict the sentiment of the utterances.
* It is a common and effective approach for text classification tasks.
* Designed multi-input models that processed features independently and combined them for final predictions.
* Chose models based on their ability to handle cross-modal dependencies and applied transfer learning for effective feature extraction.

**Comparison Between Early and Late Fusion Results:**

* Early Fusion: Combined video and text features at the input level, yielding moderate accuracy but higher model complexity.
* Late Fusion: Processed video and text data independently and combined outputs at a later stage, showing higher adaptability and better results.
* Analysis: Late fusion had better performance, with higher F1 score and AUC, demonstrating its advantage in maintaining separate data representations.

**Performance Analysis:**

* The model achieved an accuracy of 0.66
* Insights: Late fusion reduced overfitting and balanced learning due to independent feature refinement. Hyperparameter tuning further improved results.
* Challenges: Computationally intensive video analysis required parallel processing to mitigate resource constraints.
* Conclusion: The late fusion approach proved more effective for multimodal sentiment analysis on 'Friends' clips.