Project Log

Data Loading and Analysis (Week 1)

Combined data from all csv files and analysed the distribution

Feature Generation and Analysis (Week 1 and 2)

Generated features using NLP techniques such as word2vec, topic modelling, sentiment analysis and tf-idf. - Deeptansh

Analysed variance and inferenced results - Aayush

Running inbuilt models (Week 2 and 3)

Ran all baseline inbuilt models for RFC, CNN and LSTM to get a general benchmark for each model - Aayush (RFC) , Deeptansh (CNN and LSTM)

RFC (Week 4)

Built and ran RFC from scratch. Performed hyperparameter tuning using optuna. Ran on different versions of data - Aayush (Scratch RFC) , Deeptansh (Optuna)

CNN (Week 5 and 6)

Built and ran CNN model using generated features. Performed hyperparameter tuning on WandB. Ran the model on different versions of data - Aayush

LSTM (Week 6 and 7)

Built and ran LSTM on the data while simultaneously running on GRU and BiDirectional LSTM. Performed hyperparameter tuning on WandB. Ran the model on different versions of data - Deeptansh

distilBERT (Week 7 and 8)

Finetuned huggingface's distilBERT in pytorch. Ran on different subsets of data. - Deeptansh

Inferencing (Week 8)

Inferenced all the saved models on test data and compared results. Generated lime text explanation for RNN and distilBERT. - Aayush

Code cleanup and compilation of results (Week 8)

Cleaned up code and files. Compiled all results in tables. - Aayush and Deeptansh