

Process Explanation

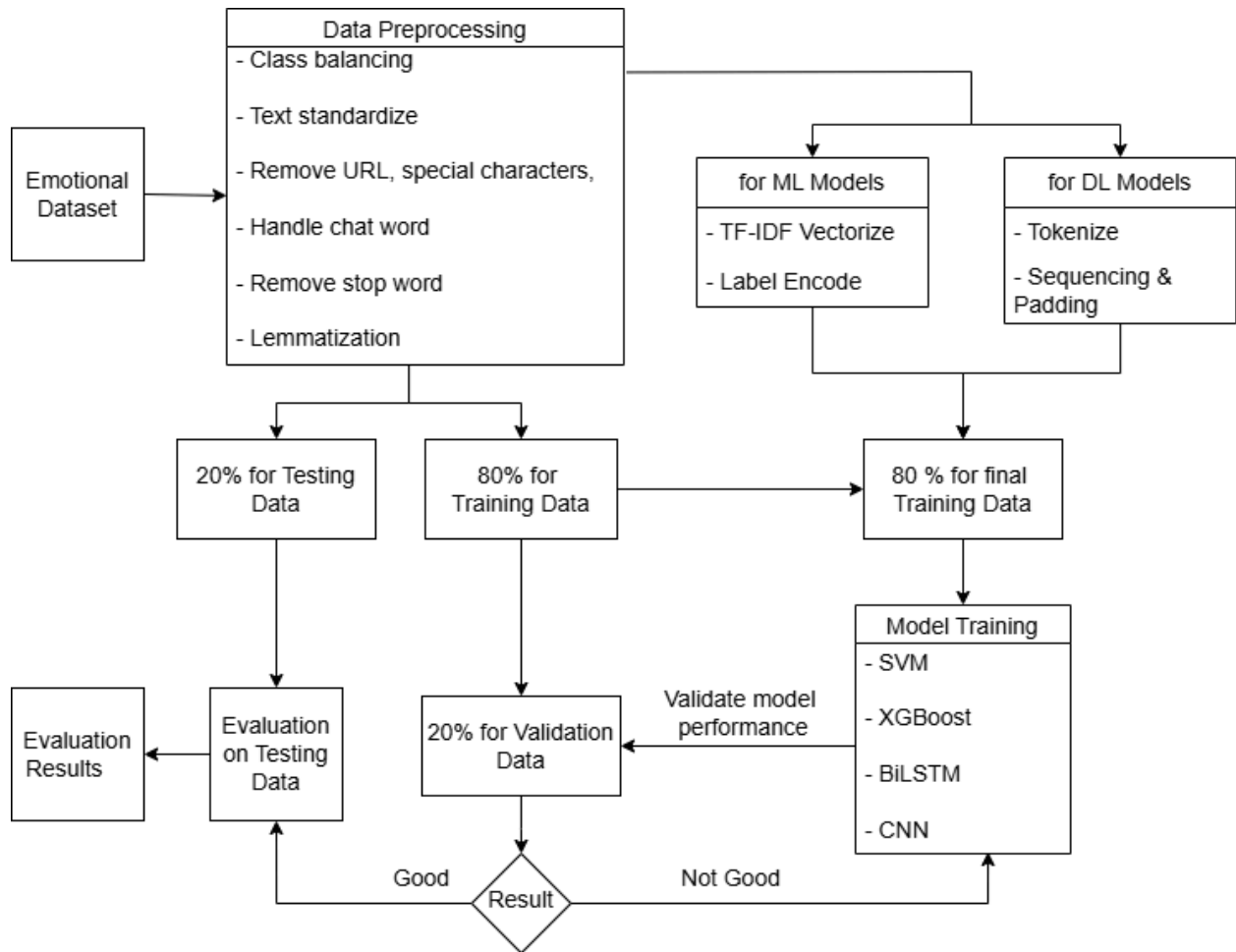


Fig 1 Process Flow Chart

The emotion dataset from Kaggle which contains around 600K text samples is used for this experiment. Data preprocessing is the first step, during which the dataset is cleansed and made ready for analysis. This includes managing certain language features like chat words and stop words, standardizing the content, and removing extra white space etc.. Following preprocessing, the data is divided into 20% training and 80% testing sets. To improve the model, 20% of the training set is separated into a validation set. After that, the data is ready for various deep learning (DL) and machine learning (ML) models. Tokenization, sequencing, and padding are applied for DL models, whereas TF-IDF vectorization and label encoding are applied for ML models. To maximize their performance, the models are trained using the training data with various strategies. After, assessed using the validation data. Once the model results are good enough and the best model is selected, it is finally evaluated on the testing set to assess its overall accuracy and effectiveness in classifying emotions from unseen text data.

Ps. More step by step details information of process is explained in the Methodology section of the draft paper.