FOML KAGGLE HACKATHON REPORT

Submission By: SM21MTECH12009

Team Members:

Prakriti Sahu (SM21MTECH12009), Shubham Ramesh Dangat (SM21MTECH14003)

FINAL MODEL USED

GRADIENT BOOSTING CLASSIFIER

Parameters used:

- max_depth = 7
- max_features = 28
- n_estimators = 103
- learning_rate = 0.1
- random_state = 1

Final Validation accuracy = 0.8523337864957597

Final Test accuracy (Public leaderboard score) = 0.86767

OTHER MODELS TRAINED

- XGBOOST (Validation accuracy = 0.849614811937593)
- RANDOM FOREST (Validation accuracy = 0.8376383763837638)
- DECISION TREE (Validation accuracy = 0.8376383763837638)

Initially, we started with the gradient boosting classifier, extreme gradient boosting classifier, random forest classifier and decision tree classifier. For every change in the parameters, feature engineering steps taken and every data processing steps taken, the gradient boosting and the extreme gradient boosting classifiers, outperformed the random forest and decision tree classifiers by a lot. The decision tree classifier and the random forest classifier gave almost the same validation accuracies for every change. Between the gradient boosting classifier and the extreme gradient boosting classifier, they gave almost the same accuracy on validation set, but the gradient boosting classifier did better on the test set than the extreme gradient boosting classifier.

After parameter tuning, feature engineering and data preprocessing, the gradient boosting classifier gave the best validation set accuracies of (0.8523337864957597, 0.8522043115167994, 0.8500679743639542), corresponding to the best test set accuracies of (0.86767, 0.86759, 0.86741) respectively.