Report
Programming Assignment 2

Answer 1) (20 points)
Describe all the pre-processing done for the given dataset. Were any columns removed?
Discuss about the change in F1score for any one pre-processing step added.

Answer 2) (15 points)
Briefly describe how much data is missing in the dataset provided.
How did you handle the missing data? Discuss at least 2 different methods used over here, and which one gives better increment in metrics and why?

Answer 3) (25 points)
Share results(F1score, precision, recall and accuracy score) from at least 2 models in your report.

Answer 4) (20 points)
What are the top 5 features and according to your reasoning, justify why the top features are relevant?
Starting with top 5 single features, find out the pair of features most relevant for the prediction.

Answer 5) (10 points)
Based on your understanding of the various models used, intuitively explain why you get such performance from the models and what it shows about the data?

Answer 6) (5 points)
Describe the model that gets the best kaggle metrics accuracy for you.