**User Guide for Cricket Winner Prediction**

There are three python files in the folder PreProcessing.py, Feature\_Selection.py and Predict.py.

PreProcessing.py – Purpose of this python script is to remove the null values from the data set and convert the textual data of the column in numerical format. This script requires path of ‘matches.csv’ and ‘deliveries.csv’ under ‘Data Set Raw’ folder.

Feature\_Selection.py – This script is for analysing the correlation between attributes and for finding the best value of the hyperparameters for 3 models. Before running this script, you need to provide the path of ‘processed\_data.csv’ file present under ‘Data set after processing’ folder.

Predict.py – This is the main python script file which is responsible for calculating the validation score of the 3 models using he K fold cross validation technique. Before running this script also, you need to provide the path of ‘processed\_data.csv’ and ‘ipl-2018.csv’ files present under ‘Data set after processing’ folder. In this we also predicted the outcomes of the 2018 IPL tournament matches. After running this script, the predicted outcomes would be saved in the newly created csv file ‘Predicted\_data.csv’.

Predict.py in the main script responsible for the prediction. You don’t need to run the other 2 scripts. You can directly run Predict.py to get the validation scores by mentioning path of required csv files in it as mentioned above. We have also added the output file ‘Predicted\_data.csv’, generated after running ‘Predict.py’, which you may need to delete before running ‘Predict.py’ so that new output file can be created.