The approach I used for predicting the score is linear regression. Since most of the test cases have around 10 overs or less, so for the test and training data sets to have similar conditions, I limited the training set to 10 overs per match. I assumed the following factors affect the total score of the game

* Batting team: A good batting team will generally put up a higher likewise a bad batting team will have a low total
* Bowling team: A good batting team will generally restrict the opponents to a lower likewise a bad bowling team will have a high target
* Total number of wide runs at a certain point: This gives the batting team an extra ball to play with and is slightly more advantageous than other extras(excluding no ball)
* Total number of no ball runs at a certain point: This gives the batting team an extra ball to play with and gives the batsman a free hit so it is slightly more advantageous than other extras(including wide)
* Total number runs at a certain point: Will tell us how the match is progressing in terms of run rate
* Total number of wickets at a certain point: More wickets mean lesser runs

Thus I used these values in the data frame while structuring the model.

Batting team and bowling team are represented by keeping a 0/1 against all the team names.

Also the teams used in the training set are limited to the ones present in the test set so in case there are extra or lesser teams in new test sets, the non essential columns must be removed from the pickle file.