Information Used to Predict The final Score:

Batting Team , Bowling Team, Overs, Current Score , Wickets

Used the information of which team plays against which team with the help of a One Hot encoded vector. Pre-Processing was required to get the Current score, wickets and Final Score for both the training and testing set.

Total data can be split into train and test to test for accuracy and then trained again on the test data for better results.

I used Random Forest Regressor Model because it can handle large datasets easily and because it ensembles multiple decision trees into its final decision which helps in increasing accuracy and efficiency.

I tried with different Machine Learning Models like Kth -Nearest Neighbours, Linear Regression , Decision Tree and Random Forest .

I was getting the maximum accuracy with Random Forest regression.

If given more time, I had included the scores of striker and non-striker since they play a major role in deciding Final Score.

Apart from this currently I am just using the data at a single moment, but we are given data of 8-12 overs. I could have used that to predict the score with much greater efficiency but due to the lack of time I could not. This could have been done with RNN using LSTM layers or I just could have used the Final Score predicted for previous balls to combine and predict the final score after 20 overs.