**IPL Powerplay score prediction**

**The Project**

The aim of this project was to predict the powerplay score in Indian Premier League 2021 for both innings before the starting of the match. We acquired the raw data from cricsheet.org which provided us with the ball-by-ball data of every match every played since 2008. We used data exploration, analysis and feature engineering to convert the data into more meaningful forms to gain further insights. Soon, we reduced the number of features to the ones with least correlation to other independent variables and most correlation with the output as they gave us the most valuable insights when performing Machine Learning. Lastly, we used linear regression techniques to obtain a predicted score.

**Libraries-Tools used**

We downloaded the files in csv format. We used Pandas and openpyxl libraries to work through the data frame. Pandas and matplotlib was used for all data exploration, feature engineering, correlation checking and data analysis. NumPy and scikit-learn libraries were used for splitting data into test train sets and performing the required linear-regression

**Progress**

We have achieved a machine learning model that predicts the score with an rms value of less than 5. The model usually predicts the scores with an error of 5 runs which is acceptable. We are currently working under the Guidance of Prof. V Prem Prakash to improve the model to get an error of not more than 1 or 2 runs.