Cricket: An Unfairly Random Game

Hitesh Mohite
hm7181@g.rit.edu

Dr. Zack Butler
zjb@cs.rit.edu

**Motivation**
- Often, it is the team that is going into the contest with better form of its players in the recent past with the same squad has the upper hand in the game of Cricket.
- Due to the increased uncertainty and randomness in the game since the advent T20I.

**Introduction**
- ODI format of the game is discussed
- In this project we try to prove this theory backed with the data obtained from a more scientific and systematically designed experiment.

**Challenges**
- The Data collection and pre-processing has been the tedious amongst all the challenges, as there were a very limited sources for collecting the data.
- As the data scrapped from web was in its crude form, it had a lot of inconsistencies. The data was not in its standard format and had many special characters attached to it.
- The structuring of data was a difficult task as the data was in its raw form and getting it standardized into the code readable format was a challenging.
- The input and the output scenarios for this project were tricky in the beginning of the project.
- Feature selection was a critical subject amongst all the challenges as there was not much data to implement a model.

**Implementation**
- The implementation part has different models – Team prediction model and Margin of victory model.
- Feature set for Team prediction model: Current form, Recent form and Cumulative win percentages for both the teams.
- Feature set for Team prediction model: Current form, Recent form and Cumulative win percentages for both the teams.
- Feature set for batting prediction model: Current SR, Recent Average, Recent SR, Current Average.
- Feature set for bowling prediction model: Cumulative Econ, Recent Econ, Recent Mdns, Cumulative Mdns.
- Team prediction model uses logistic regression to predict a winner out of two specific teams and Margin of victory uses Linear Regression algorithm to predict the model.

**Analysis**
- The Logistic Regression model for Winner prediction has not been able to perform better with the given set of features.
- The Linear regression model used to predict the batting runs has been able to predict the runs with difference of about 20 runs.
- The Linear regression model used to predict the bowling wickets has been able to predict the wickets with difference of about 1 wicket.

**Results**
- Get an accuracy of 53.33% on a predicting a team winner.
- Predicting the runs scored by a batsman in the upcoming game with difference in range of up to +20 to -20.

**Conclusion**
- The team prediction model could have performed better using more data instead of overfitting or underfitting.
- The batting and bowling prediction models is still inconsistent depending on the player’s instances.

**Future Work**
- The Game’s accuracy can be better predicted using more number of dynamic features.
- Batting and Bowling models can also be predicted with less margin of error.

**References**