MACHINE LEARNING TO PREDICT CRICKET SCORECARD

Abstract:

Cricket, over 2.5 Billion people watch this and 105 countries which plays game all over the world. The statistics of professional sports, including players and teams, provide numerous opportunities for research. The aim of this study is to investigate to what degree it is possible to predict the outcome the cricket matches by predicting individual performances in the team. The target competition is T20 cricket matches. The original features alongside engineered features gave rise to more than 8000 matches' statistics. In this study I will develop machine learning model using **neural networks** in order to predict the score card of a given cricket match.

Motivation:

Data science has recently become a popular term in the industry and providing plenty of opportunities for both professionals and organizations in growth. The two main reasons doing this project are **cricket** and **data science**.

Literature review:

An extensive online search produced very few articles related to players' performance prediction in the game of cricket. A very small number of researchers have studied the performance of cricket players. Muthuswamy and Lam predicted the performance of Indian bowlers against seven international teams against which the Indian cricket team plays most frequently. They used back propagation network and radial basis network function to predict how many runs a bowler is likely to concede and how many wickets a bowler is likely to take in a given ODI match. Wikramasinghe predicted the performance of batsmen in a test series using a hierarchical linear model. Barr and Kantor defined a criterion for comparing and selecting batsmen in limited overs cricket. Iyer and Sharda used neural networks to predict the performance of players where they classify batsmen and bowlers separately in three categories – performer, moderate and failure.

Although there are very few researches done in the past. Most of them are limited to a particular category either bowling or batting with very limited data. And aiming to predict while match is in progress. I am aiming to use each ball that has been bowled till the date, with an extensive usage of data, the accuracy of prediction will increase.

Objective:

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The objective of this project is to identify the best features that affects player performance and build an optimized model to predict each player individual contribution to the final score of the match.
To calculate the players features I have consider all the historical matches played by a player till date.

Modules of the project:

- Web scraping data from cricinfo
- Modelling Data (identifying features/calculating new features and modelling)
- Tuning the model with optimal hyper parameters.
- Prediction, furcating and visualization