

Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Links

Manuka Maduranga Hatharasinghe
Informatics Institute of Technology
No 57, Ramakrishna Road, Colombo 6, Sri Lanka
maduranga.manuka3@gmail.com

Guhanathan Poravi
Informatics Institute of Technology
No 57, Ramakrishna Road, Colombo 6, Sri Lanka
Guhanathan.p@iit.ac.lk

Abstract - Using Computer Intelligence to analyze and model the game of Cricket is a promising research area. The increased popularity and financial benefits have made Cricket an interesting sport to be subjected to statistical analysis and machine learning. The dynamic nature of Cricket, complex rules governing Cricket makes the task a challenging one. The various approaches taken and what has been disclosed from available work is neither very clear nor properly documented due to the differences in the approaches. If the good and the drawbacks of the existing work is properly analyzed and documented, it will assist in future researches. This paper presents an analysis of the existing work related to match outcome prediction in the Cricket domain. This paper is a result of an ongoing research, by the end of the research we hope to address the missing links and the drawbacks that will be explored in this paper.

Keywords – Cricket, Sabermetrics, Data Mining, Statistics, Social Media, Twitter, Machine Learning

I. INTRODUCTION

This section introduces the domain “Cricket match outcome prediction” with background information and why it is such a promising research area.

A. Background

Cricket is one of the most popular sports in the world, second only to football [2]. Cricket is played globally across 106 members of the International Cricket Council (ICC). A Cricket match is played between two countries mainly in three formats, Test Cricket, One Day International (ODI) and Twenty20 (T20). Test Cricket match is played over 5 days with 90 overs per day and One Day Cricket is played with 50 overs per each side while Twenty20 Cricket is played with 20 overs per each side. A 50-Over Cricket World Cup is held once every 4 years and a Twenty20 World Cup is held once every 2 years. In between the World Cups, teams tour other Cricket playing nations to play Test, ODI, T20 Cricket. Recently, with the introduction of domestic franchise Cricket tournaments like Indian Premiere League (IPL), Caribbean Premiere League (CPL), Pakistan Super League (PSL) etc., Cricket has become a heavy financial sport with billions of dollars involved in player auctions and prize money.

B. Motivation

Cricket can be considered as a very unpredictable sport. The whole outlook of a Cricket match can be changed within a few minutes. Due to the complex nature of the game, decisions on team selection, player performance prediction, match outcome prediction can be tough. As any other sport, every Cricket match leave behind a huge set of data that can be analyzed and modeled to extract data driven insights of the game. These insights can be very helpful to anyone who's involved in any decision-making process related to the game.

In Section 2, we explore the ways in which we gathered data and statistics in order to complete the task. In Section 3, we analyze the Cricket match outcome prediction in depth and the problems involved. In Section 4, we explore the existing solutions proposed under the category of approaches they have taken. In Section 5, we evaluate the existing solutions on a few evaluation metrics. In Section 6, we explore the identified missing links and drawbacks and in the last section we conclude the paper with identifying any possible future work.

II. STUDY SETUP

We started gaining the required domain knowledge with the literature survey and we identified 3 main problem domains that are linked with Cricket match outcome prediction,

1. Cricket Player Performance analysis.
2. Cricket match simulation.
3. Cricket Team Selection.

We explored more than 25 researches done under these domains and we identified 2 main approaches researchers have used in the past, which are discussed in Section 6. Under these approaches we identified 6 main solutions developed by researches, which are also discussed in Section 6.

III. CRICKET MATCH OUTCOME PREDICTION

Cricket is a dynamic game. A team might seem to be way ahead at the half way stage or at any stage of the game but an extraordinary performance from one player on the other team can change the outcome of the match within a few minutes. Also, various factors such as natural elements,