

**ODI Cricket Match Winning Prediction Using Data Mining
Techniques**

M.Sc. in Information Technology

2018

ODI Cricket Match Winning Prediction Using Data Mining Techniques

Subasingha SADP

158777V

Faculty of Information Technology

University of Moratuwa

June 2018

ODI Cricket Match Winning Prediction Using Data Mining Techniques

Subasingha SADP

158777V

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the fulfillment of the requirements of Degree of Master of Science in Information Technology.

June 2018

Declaration

We declare that is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student

Signature of Student

S.A.D.P Subasingha

.....

Date:

Supervised by

Name of Supervisor

Signature of Supervisor

S.C. Premarathne

.....

Date:

Acknowledgements

First and foremost, I would like to express my sincere gratitude towards my supervisor, Mr. Saminda Premarathne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his guidance, supervision, advices and sparing valuable time thorough the research project.

Furthermore, big thank should goes to Prof. Asoka Karunananda who taught Research Methodology and Literature Review and thesis writing subjects which were the basis for this research work.

I would also like to thank all the batch mates of the M.Sc. in IT degree program who gave their valuable feedbacks to improve the results of the research and my family for the support they provided me through my entire life and in particular. I must acknowledge my wife encouragement and editing assistance, I would not have finished this thesis.

Abstract

Cricket is one of the most popular games in many countries. More than 19 countries playing cricket as main game & population for cricket will increasing gradually. However there are no proper tool for analyzing pre outcome of the match stating to end, and available tools are not support to simulate match using batting partnerships. The ultimate goal of predicting pre outcome of cricket match is to identify key players and their batting performances also prevent wrong players selecting and toss decision by making statistical predictions.

This project is focusing on One Day International (ODI) cricket match and predicts the outcome of the particular match. Our proposed solution consists of three major modules namely; Web UI Module, CRIC-Win Analytic Engine and Backend Data Module. CRIC-Win Analytic Engine has two sub data models, one for predict overall match outcome based on given pre match data and next for predicting match outcome based on batting partnership both home team and opponent team. All sub models in the CRIC-Win Analytic Engine are developed base on Naïve Bayes algorithm and use for generating the classifier model which can be used to predict the outcome of the cricket match.

Mainly prediction result divided into two segments. First predict overall pre outcome of the match based on given details and next predict how each partnership will affect to win the match. This work suggests that the relative team strength between the competing teams forms a distinctive feature for predicting the winner. Modeling the team strength boils down to modeling individual players batting partnerships and bowling performances forming the basis of our approach. We use partnership statistics as well as the recent performances of home team & opponent team, partnership records, wicket falling pattern, remaining overs, required runs, toss decision, ground and day/night effects have also been considered in order to predict the outcome of a match. The solution can be simulated match outcome before the match start and while playing. Also this research describes in detail the different attribute selection techniques as well as the data mining algorithms used to solve this problem of outcome prediction in cricket. We have also used accuracy as the evaluation criteria to evaluate how well the prediction performs.

Contents

	Page
Declaration	i
Acknowledgements	ii
Abstract	iii
Contents.....	iv
List of Figures	vii
List of Tables.....	viii
Chapter 1	1
Introduction	1
1.1 Prolegomena	1
1.2 Problem Statement.....	1
1.3 Aim and Objectives	1
1.4 Background and Motivation	2
1.5 Problem in Brief	3
1.6 Proposed Solution.....	3
1.7 Structure of the Thesis	3
Chapter 2	4
Review of Using Data Mining Techniques for Wining Prediction in Cricket.....	4
2.1 Introduction	4
2.2 Data Mining Techniques	4
2.3 Data Mining Tools.....	6
2.4 Data Mining Process.....	7
2.5 Use of Data mining in sports	9
2.6 Sports Data Mining Applications	11
2.7 Summary of Challenges.....	12
2.8 Problem Definition	13
2.9 Summary.....	13
Chapter 3	14
Technology Adapted in Winning Prediction Analysis.....	14
3.1 Introduction	14
3.2 What is Data Mining?.....	14
3.3 Reasons for Classification Data Mining Algorithms for ODI-Cricket Match Winning Prediction.....	16

3.4 Java	16
3.5 WEKA / WEKA API.....	16
3.6 .NET Framework	17
3.7 MS SQL.....	17
3.8 Summary.....	17
Chapter 4	18
A Novel Approach to Predict Outcome of the ODI Cricket Match.....	18
4.1 Introduction	18
4.2 Hypothesis	18
4.3 Input.....	18
4.4 Output	19
4.5 Process	20
4.6 Users	20
4.7 Features.....	20
4.8 Summary.....	20
Chapter 5	21
Design of the CRIC-Win Predictor Tool.....	21
5.1 Introduction	21
5.2 Top Level Architecture (Design) of CRIC-Win Predictor	21
5.3 Data Collection and Preprocessing.....	22
5.4 User Interface Module	22
5.5 CRIC-Win Analytic Engine.....	22
5.6 Back-End Database.....	23
5.7 Summary.....	23
Chapter 6	24
Overall Implementation of CRIC-WIN Predictor Tool	24
6.1 Introduction	24
6.2 Overall Implementation.....	24
6.3. Data Collection and Preprocessing.....	25
6.3 Implementation of User Interface module.....	25
6.4 Implementation of CRIC-Win Analytical Engine	26
6.5 Implementation of Back-End Data Module.....	27
6.6 Summary.....	27
Chapter 7	28
Evaluation.....	28

7.1 Introduction	28
7.2 Evaluation of Classification Techniques	28
7.3 Evaluation of CRIC-Win Predictor Tool.....	29
7.4 Summary of Evaluation	30
Chapter 8	31
Conclusion and Further Work	31
8.1 Introduction	31
8.2 Limitations.....	32
8.3 Future Developments.....	32
8.4 Summary.....	32
References.....	33
Appendix – A	35
Attribute Selection.....	35
Appendix - B.....	37
Backend Data Module Implementation.....	37
Appendix - C.....	39
Model Evaluation Summary.....	39
Appendix - D.....	42
CRIC-Win Predictor Tool Test Result	42
Appendix - E.....	44
Code Snippet.....	44

List of Figures

Figure 2.1 Data Mining Techniques	4
Figure 2.2 Data Mining Process.....	8
Figure 3.1 Standard Process for Data Mining.....	15
Figure 3.2 Data Mining Techniques	15
Figure 4.1 ARFF Format Data Set.....	19
Figure 5.1 Top Level Architecture of CRIC-WIN Predictor solution.....	21
Figure 5.2 Design of CRIC-Win Analytical Engine.....	22
Figure 6.1 CRIC-Win Predictor Tool	24
Figure 6.2 Data Preprocessing	25
Figure 6.3 CRIC-Win Analytical Engine WSDL	26
Figure A: 1 Information Gain Attribute Evaluation	35
Figure A: 2 Attribute Selection.....	36
Appendix B: 1 Entity Framework Model Browser.....	37
Appendix B: 2 Backend Database Diagram	38
Appendix C: 1 Naive Bayes Model Creation Summary	39
Appendix C: 2 Decision Tree Model Creation Summary.....	39
Appendix C: 3 AdaBoost Model Creation Summary	40
Appendix C: 4 RandomForest Model Creation Summary.....	40
Appendix C: 5 Bagging Model Creation Summary.....	41
Appendix D: 1 SL vs India Match Outcome	42
Appendix D: 2 1st Inning 4th Wicket Partnership.....	42
Appendix D: 3 1st Inning 6th Wicket Partnership.....	42
Appendix D: 4 1st Inning Last Wicket Partnership	43
Appendix D: 5 2nd Inning 3rd Wicket Partnership	43
Appendix E: 1 Web Cricket Data Scraping	44
Appendix E: 2 Fetch Data from Backend Module.....	45
Appendix E: 3 Backend Service Module code Snippet.....	46
Appendix E: 4 Weka API Web Service Code Snippet	48

List of Tables

Table 2.1 Summary of Challenges	12
Table 4.1 Data Set Over-fitting Situation	18
Table 7.1 Classification Evaluation Measurements	28
Table 7.2 Comparison of Classification Techniques to Determine Overall Match Outcome	29
Table 7.3 Analysis Wining Factors Before Start the Match	29
Table 7.4 Sri Lanka vs. India Match Pre-Outcome	30
Table 7.5 Analysis of 1 st Inning Partnerships	30
Table 7.6 Analysis of 2nd Inning Partnerships	30