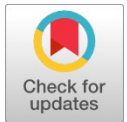


Novel Method for Cricket Match Outcome Prediction using Data Mining Techniques

S.A.D.P Subasingha, S. C. Premaratne, K. L. Jayaratne, P. Sellappan



Abstract: Cricket is one of the most popular games in many countries. As many as 19 countries play cricket as their main game, and the number is likely to increase in the future. However, there are no suitable tools for analyzing pre-outcome of the match from beginning to end. Existing tools do not support simulating match using batting partnerships. The ultimate goal of predicting pre-outcome of a cricket match is to identify key players and their batting performances. It is also to prevent wrong players from selecting and toss decision by making statistical predictions. This research focuses on One Day International (ODI) cricket match and predicts the outcome of a particular match. Our 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 predicting the overall match outcome based on a given pre-match data, and the other for predicting match outcome based on batting partnership of both home and rival teams. All sub-models in the CRIC-Win Analytic Engine are developed using the Naïve Bayes algorithm for generating the classifier model, which is used to predict the outcome of the cricket match.

Index Terms: Classification, Data mining, Cricket Match, Prediction.

I. INTRODUCTION

Today, sports are intensely competitive propositions. Motivated by huge financial rewards, sports professionals are engaged in huge competition, always try to take advantage over their opponents. Currently, sports professionals include not only the sportsmen actively participating in the game, but also their coaches, trainers, physiotherapists, and in many cases, strategists. Coaches, captains and team managers leverage their expertise and make decisions using their intuition. Such decisions can be biased by human impressions, and judgments of players and hence might overlook players' weaknesses. Moreover, interesting patterns in the game may hide the eye of the best tactician [1].

Effective formulation of strategies requires carrying out an extensive analysis of past games, current performance in the game in progress, and numerous other factors affecting a game. Players and team management (collectively often referred to as the team think-tank in sports) perform as a "human expert system", relying on experiences, expertise

and analytic ability to arrive at the best possible course of action before as well as during a game. The vast amount of raw data and statistics are available to aid in the decision-making process but determining what it takes to win a game is extremely challenging [1]. The trend noticeable both in an individual sport such as tennis and in team sports such as baseball and basketball are that this knowledge is used to determine pre-game strategy. Successful application of this pre-determined strategy often becomes an important factor in victory [2].

Cricket is the second most popular sports in the world with billions of fans across India, Sri Lanka, England, Pakistan, Africa, Australia, etc. It is an outdoor game played on a cricket field at the 22-yard rectangular long pitch, between two teams consisting each of 11 players. It is played in three formats, namely Test, One Day International (ODI) and Twenty over International (T20). In ODI, each team takes its chance to bat, trying to score as many amounts of runs, which can be scored in 50 overs while the other team fields for that much amount of overs. Each chance is termed as an innings [3]. Unlike other sports, cricket stadium's size and shape are not fixed except for dimensions of the pitch and inner circle, which are 22 yards and 30 yards respectively. The cricket rules do not mention the size and shape of the field of the stadium [4]. Pitch and outfield variations can have a substantial effect on batting and bowling. The bounce, seam movement and spin of the ball depending on the nature of the pitch. The game is also affected by atmospheric conditions such as altitude and weather. A unique set of playing conditions are created due to these physical differences at each venue. Depending on these set of variations a particular venue may be a batsman-friendly or a bowler-friendly [5].

Currently, in a cricket match, the projected scores can be seen displayed at the scorecard during the first innings, which is the final score of the batting team at the end of that innings if it scores according to the current run rate or a particular rate. Run rate is defined as the amount of runs scored per the number of overs bowled. However, the run rate is considered as the only criteria for calculating the final score. But there are other factors too which may affect the final score like the number of wickets fallen, the venue and the batting team itself [4]. Match data since the beginning of the ODI game is available. However, our literature search found no proper tool for predicting pre-match outcome and simulating the game in a given situation to make decisions. Some related work could be found on the topic of "A Classification Based Tool to Predict the Outcome in ODI Cricket" by Amal Kaluarachchi and Aparna S. Varde [3]. They used the Naïve Bayes algorithm to predict the outcome,

Manuscript published on 30 September 2019.

* Correspondence Author (s)

S.A.D.P Subasingha, Faculty of Information Technology, University of Moratuwa.

S. C. Premaratne, Faculty of Information Technology, University of Moratuwa.

K. L. Jayaratne, University of Colombo Schools of Computing.

P. Sellappan, School of Science and Technology, Malaysia University of Science and Technology (MUST).

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

but they have considered only basic factors and failed to simulate the game for making strategic decisions. So, this research aims to study the problem of predicting the game results before the game start, based on the statistics and data available from the data set and simulate pre outcome based on the given situation to make decisions for win the game.

II. LITARATURE REVIEW

The problem of match outcome prediction has been studied extensively in the context of baseball, basketball, and soccer. Bhandari et al. [6] developed the Advanced Scout system for discovering interesting patterns from basketball games, which has is now used by the NBA teams. More recently, Schultz [7] studies how to determine types and combination of players most relevant to winning matches. In soccer, Luckner et al. [8] predict the outcome of FIFA World Cup 2006 matches using live Prediction Markets. In baseball, Gartheepan et al. [9] built a data-driven model that helps in deciding when to 'pull a starting pitcher'.

In 2015 Khabir Uddin Mughal and P. Bhatia [10] were proposed that two methods, first predicts the score of the first innings on the current run rate, considering the number of wickets fallen, venue of the match and batting team. The second innings considering the same attributes as of the first method along with the target given to the batting team. Almost previous works [11] are worked in statistically predicting such as (average predicting techniques) [12] the scores or the outcome of the match. But this research uses two separate models one for the first innings and other for the second innings. Also, use linear regression classifier predictions for current data and Naïve Bayes classifier on the records. They observed that the error in linear regression classifier is less than the current run rates; another method in predicting the final score. Also, the accuracy of the Naïve Bayes is going from a higher percentage. However, further research would be conducted to improve the accuracy of both models. Also in 2016 Anik Shah, Dhaval Jha and Jaladhi Vyas introduced a tool that Wining and score predictor (WASP)[6] is a calculation tool used in cricket to predict the score and possible results of a limited over match format. According to analytics provided throughout this report, it can be derived that, there are loopholes in the existing method, and it needs to be resolved for better judgment. Also, different parameters analyzed during the new method suggest that it is an improved method than before and can be implemented at a higher platform like international cricket with some changes being done. Furthermore, in 2014 V. V. Sankaranarayanan, J. Sattar, and L. V. Lakshmanan developed separate models [11] for finding home runs and none-home runs using historical features as well as instantaneous match features from past games. Also, it demonstrated the quality and accuracy of their predictions with an extensive set of experiments on real ODI cricket data [5]. In addition to that, predicting runs for future segments. They proved that winner prediction accuracy is the highest reported in ODI cricket mining literature. However, future research should be conducted to predict the fall of wickets and improve prediction accuracy even further.

Rizwan khilaiq khan , Irian Manarvr and Mohay-ud-dhr have been established that data mining[13] could be

successfully used to evaluate the past performance of any cricket team such as data is continuously being collected all around the world. It analysis could be used successfully or making future strategies by team managers and players against other teams. It may also be concluded from the performance of New Zealand's team could perform much better at home grounds even against best teams of the world because of the support it gets from the crowd as well as having more practice on the same ground. However, this research not clearly mentioned how they analyzed data only represents previous data. So future research should be conducted on how to construct good modal to analyze cricket match records. In the recent past (in the last three decades/Since the 1980s) Data mining has shown a multifaceted development (growing interest) in many disciplines. Undisputedly R.P Singh & Rizwan khilaiq have produced the first reported web tool to predict the score and winning prediction called WASP [12] (winning and score predictor) in 2016. Subsequently, numerous applications in sports winning prediction were developed by many researchers. For instance (example), Brown and others [13] develop some model for predicting winning prediction in cricket match but considering only a few affecting factors. Such as run rate, pitch condition etc. However, there are much more affecting factors to be considered to a prediction about the cricket match. Because new rules and regulations are introduced by the international cricket council called ICC. It should be noted that these are rather industry-based and complex in nature.

In contrast, data mining techniques have also been used to predict cricket score and winning by many researchers. These researches are primarily targeted only to predict projected score anyhow some research predict winning percentage also. But the accuracy of prediction is very much low.

There is always a sport in everyone's life that takes over their lives and we don't mean that in a bad way. After being dwelled in work all day the human mind really needs to rest and that is where it needs a sport. This sport shouldn't be just any other boring sport, it should be something that thrills you and excites us. Something that makes our heart beat faster on every move. Such sports provide us an adrenaline rush which makes us forget about everything and focus on the game.

The one sport that can get us the adrenaline rush after football is none other than cricket! Cricket has been in our lives for quite some time now. Ranked as the second most watched sport in the world, cricket has its own fan following that goes lengths. With millions of followers from all over the world, the sport goes back to ages. From the rivalry between India Pakistan to the legendary matches between Australia and England, this is something that goes back ages and has been a part of everyone's life for a long time.

Match data since the beginning of the ODI game is available. Two team members are announced before the match. Some works could be found on cricket match scoring rates by Clarke [14] and Preston and Thomas [15]. They used dynamic programming methods to predict scoring rates.

Also, some studies, such as those conducted by De Silva [16] analyze the magnitude of the victory.

It is found that most of these studies describe the factors with the goal of predicting the probability of victory. In the real-world scenario, however, there are cases where the magnitude of the victory is important especially when betting is involved.

There are different ways to do the prediction. The prediction can be done taking into consideration the player's performance as well as the team performance. There are many unpredictable things that happen in a cricket game like matches being washed out due to rain, a key player getting injured before the game, players changing their teams, etc. Sometimes a key player also gets injured during the game and hence is not able to take further part in the game. All these factors do affect the prediction to some extent. The report discusses a methodology that I followed for the game result prediction. The methodology consists of first the attribute selection algorithms which trim down the list of attributes to only important ones and then the data mining algorithms which can be applied to those attributes. The game prediction problem that our study does not take into consideration the player's performance, but it does take into consideration the team's past performance at a high-level extent along with the other factors like toss winner, toss decision, home support, etc. The attribute selection techniques consist of the wrapper method and the ranker method. The data mining algorithms that are used are Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbor. The data mining tool used in the project is WEKA, and it is a freely available data mining tool which has good support for a number of different data mining algorithms.

III. DESIGN

In our design, we have decided to maintain a separate interface module to enable access from different sources of inputs. We have also included a database for data manipulation. The core of the proposed solution is defined as the CRIC-Win Analytics Engine. The top-level architecture of the CRIC-Win Predictor is shown in Fig. 1.

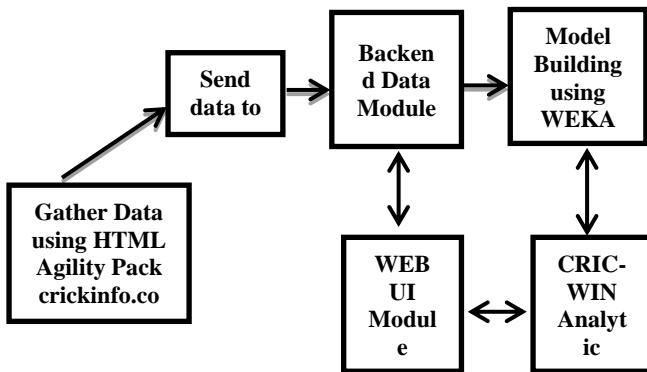


Fig. 1. Top Level Architecture of CRIC-WIN Predictor.

According to Fig. 1 we collect cricket match data using HTML Agility library. As we expected www.crickinfo.com sports web site had displayed all previous matches' data in HTML format, so we have to scrap data and push all raw data to the separate database. We do this purposely to reduce the weight on the main database. Then collected row data clean and sent to the main backend data module for easy query manipulation. Then we build classifier models by using WEKA tool, and also, we include testing and evaluation for model building. The core module of this solution is that CRIC-Win Analytical Engine. Mainly it acts as a mediator when a request comes, it analyzes the data coming from Web

UI and picked up appropriate models for making the prediction. Then visualized predicted result on Web UI.

A. Data Collection and Preprocessing

We have to use the HTML Agility Pack for data collection process through web data sources. HTML agility pack is the library that helps to scrap the matches' data from www.cricinfo.com. It speeds up the data collection process because we don't have freely available APIs for collection ODI cricket matches data. Initially, we stored scraped data into a separate database to minimize the weight of the primary database. Then cleaned data send to the primary database for future query manipulation.

B. User Interface Module

This module enables user interaction with the system in terms of input and output requirements. More importantly, this module has direct access to the main back-end database and the Analytical Engine. This module is not limited to handle input/output but execute certain pre-processing on input data and also the processing of results to appear in different forms.

C. CRIC-Win Analytic Engine

This engine primarily works as a data mining engine and performs the roles of data preprocessing, attribute selection, classification and finally analyzed results push to the main database. Fig. 2 shows phases of data mining that are covered by CRIC-Win Analytic Engine. Also, this engine directly connected with WEKA tool. The output of the WEKA tool directly captured by CRIC-ESS Analytic Engine and then parallel sends to both the back end and the user interface module.

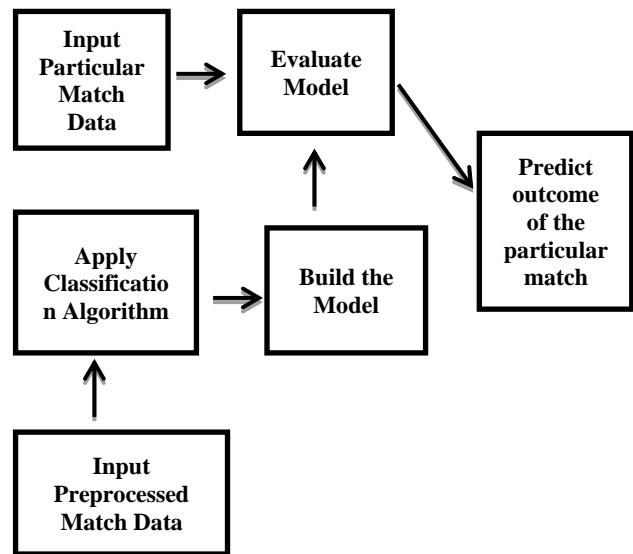


Fig. 2. Design of CRIC-Win Analytical Engine

Based on the request coming from the user interface, this module can generate output through the database. This module cannot be accessed directly by the users without going through Web UI. Because of the user have to choose two teams, players details, ground details, and other necessary stuff from Web UI. When user request coming from Web UI, will send to the backend data module, then backend data module send back JSON formatted data to the Web UI.

Then the only user can connect with CRIC-Win Analytic Engine and it responsible for evaluating the model based on given match data and predicted result send back to the Web UI. However, limited information related to the current session can be returned to the user interface module from the CRIC-Win Analytic Engine. In general, this module can be seen as an extension to and ordinary data mining tool except for the input/output mechanism.

The back-end database is the key to handle ongoing data mining with CRIC-Win Analytical Engine. This is also an essential component to work with the User interface. The design of this database is rather crucial activity in the solution. This affects the analytical engine as well as the user interface. The database consists of all analyzed summary data and also players' details, Match details, Ground Details, and other supported tables to minimize data redundancy.

According to this design without going through CRIC Win Analytical Engine, the system can facilitate sessions to query on the database, related to the past data. More importantly, the interaction between the user interface and back-end database create the opportunity for the CRIC-Win Analytical Engine work efficiently by knowing what exactly should be done for the specific user request. Thus, the workload on the analytic Engine will be reduced.

IV. IMPLEMENTATION OF CRIC-WIN PREDICTOR

The overall solution is a web-based one and has been implemented. Net based application is running on any web browser. The implementation is concerned about ASP.net MVC architecture and Java-based rest services for connecting WEKA tool with Web UI. The main module of our product CRIC-Win Analytical Engine implements as Java-based soap service. Because data mining tool called WEKA can be easily integrated with the Java platform. Web UI module develops based on .Net technology and Backend data module is developed by using Entity Framework 6.0 with MS SQL RDBMS technology for creating databases. Fig. 3 shows a snapshot of CRIC-Win Predictor tool. Next, we discussed module wise implementation with reference to the top-level architecture.

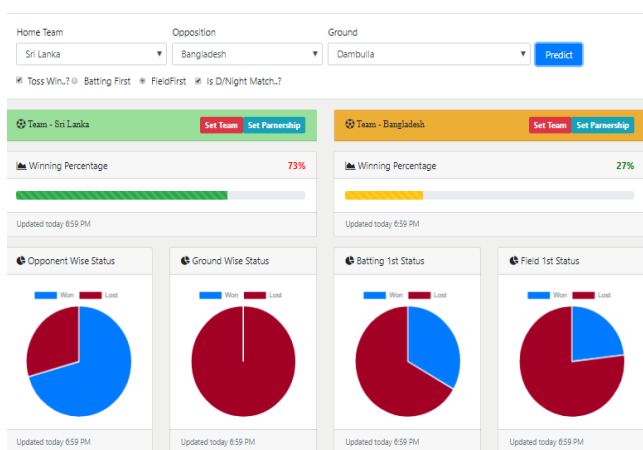


Fig. 3. CRIC-Win Predictor Tool

Data for the analysis is collected from www.cricinfo.com. However, we have to use special. The net library called HTML Agility Pack. We have created web data scrapping tool to gather cricket match related data and then scrapped data push into a separated database because scrapped data had a lot of missing values and noisy values in addition to that

redundancy data are the major problem when scrapping data from the web. After eliminating redundant data and cleaning noisy data, we stored them in a separated database which is used to create the main backend module. We do this purposely to reduce the weight of the main database module. Next, we use MS SQL Server Management studio for query manipulation and fetching up the necessary information to create the main data set. Finally, we use WEKA tool to apply the Attribute Selection filter to select only necessary attribute with the evaluator InfoGainAttributeEval, which evaluated the worth of an attribute by measuring the information gain.

CRIC-Win Analytical Engine is the major module of our solution. This module is developed by using Java technology with the WEKA API. WEKA tool can be easily integrated with Java, but our UI module was developed by using .Net technology so sorts out this language gap we introduced separate java based SOAP service module. After creating the complete dataset, we have to construct data models to predict the outcome of the selected match in a given situation. In model creation, we have tried different machine learning techniques, namely clustering, classification, and association rule mining. Selected algorithms from each technique were trained using the WEKA tool. Clustering and association rule mining did not make any contribution to predicting the outcome of the match, because we have used multiple independent attributes, therefore, placing them in groups based on their similarity did not seem feasible.

However, classification techniques produced significant results. In machine learning, Naïve Bayes and Decision Trees are two popular classification approaches. We also looked into two other algorithms, AdaBoost and Bagging. The results obtained were interesting in terms of improving performance. Likewise, we create separate data models and stored them in the backend data module in binary format. Then our CRIC-Win Analytical Engine is responsible for fetching the appropriate model from the Backend data module and evaluating the model based on user requests. Finally predicted outcome send back to Web UI. This CRIC-Win analytical engine can be easily plugged any web application without any modification.

Back end data module is the separate module that handles all data request coming from the Web UI. The reason for using a separate module for data requesting is that reduce the dependency of the central database with Web UI and CRIC-Win Analytical Engine. This module was implemented by using Entity Framework 6.0, which is very easy to use with MS SQL database and easy to maintain.

V. EVALUATION

We have trained the collected data set using different classification techniques, namely Naïve Bayes, J48, Bagging, and RandomFores, by using WEKA tool. ODI Cricket match outcome prediction model is the main input for the CRIC-Win Analytical Engine. For evaluating a classifier quality, we used the confusion matrix, which evaluated various measurements such as accuracy, recall, and precision. These measurements and their definition are given in Table I.

Table I. Classification Evaluation Measurements

Measurement	Formula	Description
Precision	$TP / (TP + FP)$	The percentage of positive predictions those are correct.
Recall / Sensitivity	$TP / (TP + FN)$	The percentage of positive labeled instances that were predicted as positive.
Specificity	$TN / (TN + FP)$	The percentage of negative labeled instances that were predicted as negative.
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$	The percentage of predictions those are correct.

Using previously mentioned measurements, we can derive TP rate, FP rate, F-measure, and ROC area. TP rate is equal to sensitivity, while the FP rate is equal to 1- specificity F-measure calculated by precision and recall.

The area under a ROC curve quantify the overall ability of the test to discriminate between usefulness and uselessness a truly useless test as an area of 0.5. A perfect test has an area of 1.00. Usually, best models have higher TP rate, lower FP rate, and ROC space close to 1.00. The classification results of each technique are given in Table II.

Table II. Comparison of classification techniques to determine overall match outcome.

Technique	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC
Naïve Bayes	0.982	0.082	0.982	0.982	0.981	0.978
Decision Tree	0.975	0.092	0.975	0.975	0.974	0.998
AdaBoost	0.772	0.746	0.691	0.772	0.704	0.867
Bagging	0.696	0.809	0.6	0.696	0.644	0.402
RandomForest	0.722	0.759	0.641	0.722	0.674	0.829

Table II shows that Naïve Bayes and Decision Tree represent almost equal accuracy, but Naïve Bayes having higher accuracy value than Decision Tree. It was found that Naïve Bayes produced the best results in ODI Cricket match winning prediction. Other classification techniques, as shown in the above table, did not perform significantly well in our research.

We tried some match-winning factors including toss, day/night effect, and ground advantages for predicting match outcome before starting the match, and next we consider batsman combination with their partnership runs to predict match outcome.

As shown in Tables III - VI, our CRIC-Win Predictor tool produced some interesting results, so winning predictor tool can be used to predict the outcome of future matches using the knowledge discovered from our research. Also, cricketers, coaches, team selectors, and other interesting parties who interested to gain competitive advantages to win the game, can be used our predictor tool for decision making.

Table III. Analysis of winning factors before starting the match.

Teams	Win Toss	Batting 1st	Day Match?	Ground	Win Probability
Sri Lanka	Yes	Yes	No	Dambulla	52%
Bangladesh	No	No	No		48%
Sri Lanka	Yes	No	No	Dambulla	73%
Bangladesh	No	No	No		27%
Sri Lanka	No	Yes	Yes	Mohali	25%
India	Yes	No	Yes		75%
Sri Lanka	Yes	No	Yes	Melbourne	60%
India	No	Yes	Yes		40%
Sri Lanka	Yes	No	No	Hambantota	79%
Zimbabwe	No	Yes	No		21%

Table IV. Sri Lanka vs. India Match Pre-Outcome

Teams	Win Toss	Batting 1st	Day Match?	Ground	SL Win Probability
Sri Lanka	Yes	No	No	Dharmasala	51%
India	No	Yes	No		49%

Table V. Analysis of 1st Inning Partnerships.

Wicket No	Partnership combination	Runs	Total Run	Overs	SL Win Probability
04	MK Pandey SS Layer	8	16	12.5	75%
06	MS Dhoni HH Pandya	12	28	15.2	73%
10	MS Dhoni YS Chahal	35	112	38.2	85%

Table VI. Analysis of 2nd Inning Partnerships.

Wicket No	Partnership combination	Runs	Total Run	Overs	Target	Win Probability
03	AD Mathews WU Tharanga	53	65	12.3	112	87%

VI. CONCLUSION

In this research, we have addressed the problem of predicting the chances of victory in a One Day International cricket match. By analyzing different attributes related to the ODI game, we have been able to predict the winning criteria formulated using attributes from the dataset. We have developed a web-based software tool called CRIC-Win Predictor based on our study.

The main goal of this research is to learn a model for predicting game progression and outcome in ODI cricket match. We separate our prediction into two segments. First one is for predicting winning possibility by considering toss effect, ground condition, day-night effect, and opponent. Next, we consider batting performances by taking into players combination, wicket no, partnership runs, total runs or target and overs to predict the progression of the game. This way help us to simulate the game properly. Naïve Bayes algorithm is used to predict the match outcome in any given situation. We demonstrated the quality and accuracy of our predictions with an extensive set of experiments on real ODI cricket data. In addition to predicting winning chances for future segments, our winner prediction accuracy is indicated higher value because of calculating winning possibility before starting the match and while playing the match.

Currently, team strategists rely on a combination of personal experience and team constitution. Inherently, the methodology employed by human experts is to extract and leverage important information from both past and current game statistics. So, match outcome prediction can also be helped to make strategic decisions to increase winning chances. The main contributions of our work can be listed as follow.

- Comparison of machine learning techniques which revealed that classification is the best approach to solve the problem.
- Evaluation of various classifiers over real data which proved that Naïve Bayes works best over the concerned datasets.
- Partnership combinations, partnership runs and wicket falling pattern is taken into predict progression of the game.
- Development of the web-based CRIC-Win Predictor tool that can be used in real-world scenarios to predict the chances of victory in a given match.

VII. FUTURE WORK

As future work, we are planning to expand our analysis using more attributes such as the previous match result of the selected team and the opponent team, the number of known batsmen in the selected team and the opponent team and more. It is also possible to apply association techniques to predict partnership broken bawler.

We are currently working to further reduce the prediction error. Furthermore, to make the prediction engine functionally complete, we intend to predict the fall of wickets, overcoming the challenges presented in the current data set and expanding match prediction other countries as well. Finally, we aim to leverage bowler's features (in addition to the batsmen's) to improve the prediction accuracy even further.

REFERENCES

1. V. Veppur Sankaranarayanan, "Towards a time-lapse prediction system for cricket matches," PhD Thesis, University of British Columbia, 2014.
2. D. Prasetyo and others, "Predicting football match results with logistic regression," in *Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, 2016 International Conference On, 2016, pp. 1–5.
3. A. Kaluarachchi and S. V. Aparna, "CricAI: A classification based tool to predict the outcome in ODI cricket," in *Information and Automation for Sustainability (ICIAFs)*, 2010 5th International Conference on, 2010, pp. 250–255.
4. Y. Saito, M. Kimura, and S. Ishizaki, "Real-time prediction to support decision-making in soccer," in *Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, 2015 7th International Joint Conference on, 2015, vol. 1, pp. 218–225.
5. T. Singh, V. Singla, and P. Bhatia, "Score and winning prediction in cricket through data mining," in *Soft Computing Techniques and Implementations (ICSCTI)*, 2015 International Conference on, 2015, pp. 60–66.
6. I. Bhandari, E. Colet, and J. Parker. *Advanced Scout: Data mining and knowledge discovery in NBA data*. *Data Mining and Knowledge Discovery*, 1(1):121{125,1997.
7. D. Lutz. A cluster analysis of NBA players. In *MIT Sloan Sports Analytics Conference*, 2012.
8. S. Luckner, J. Schroder, and C. Slamka. On the forecast accuracy of sports prediction markets. In *Negotiation, Auctions, and Market Engineering*, International Seminar, Dagstuhl Castle, volume 2, pages 227{234,2008.
9. G. Gartheeban and J. Gutttag. A data-driven method for in-game decision making in mlb: when to pull a starting pitcher. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '13, pages 973{979, New York, NY, USA, 2013. ACM.
10. Khabir Uddin Mughal. Top 10 Most Popular Sports In The World. http://sporteology.com/top-10-popular-sports-world/Accessed_2_February_2015.
11. V. V. Sankaranarayanan, J. Sattar, and L. V. Lakshmanan, "Auto-play: A data mining approach to ODI cricket simulation and prediction," in *Proceedings of the 2014 SIAM International Conference on Data Mining*, 2014, pp. 1064–1072.
12. A. Shah, D. Jha, and J. Vyas, "WINNING AND SCORE PREDICTOR (WASP) TOOL."
13. R. K. Khan, I. Manarvi, and others, "Evaluating performance of Blackcaps of New Zealand vs. global cricket teams," in *Computers & Industrial Engineering*, 2009. CIE 2009. International Conference on, 2009, pp. 1500–1504.
14. Allsopp, P., & Clarke, S. R., Rating teams and analyzing outcomes in one-day and test cricket. *Journal of the Royal Statistical Society A*, (2004).
15. S.R. Clarke, "Dynamic programming in one-day cricket— optimal scoring rates", *Journal of the Operational Research Society*, 1988, Vol. 39, No. pp. 331–337.
16. B.M De Silva, and T.B. Swartz, Estimation of the magnitude of the victory in one-day cricket. *Australia and New Zealand Journal of Statistics*, 2001, Vol. 43, pp. 1369-1373.

AUTHORS PROFILE



Dinesh Priynkara Subasingha is an MSc student in the University of Moratuwa Sri Lanka. He obtained his B. Sc in Management & Information Technology from the South Eastern University of Sri Lanka in 2013. Currently, working as a Senior Software Engineer, specializing in Microsoft Azure Cloud platform in the reputed IT Company in Sri Lanka. His research interests are Data Mining special, sports sector and business sector data for AI chat bot. This is the first research paper is going to publish by him. Now he is conducting Research & Development to create a model for text analysis and classification



Saminda Chandika Premaratne is a senior lecturer in the Department of Information Technology, Faculty of Information Technology, University of Moratuwa Sri Lanka. He obtained his B.Sc. in Computer Science from the American University of Asia in 2002 and obtained his M.Phil. in 2008 from University of Colombo School of computing. Currently, conducting his Ph.D. studies in Malaysia University of Science and Technology. His research interests are Data mining, Video processing, and multimedia information retrieval. He has published his research work in several conferences and journals. He is currently conducting research mainly on event resolution on sports videos for effective content retrieval. The goal of this research is to automate the indexing of sports videos based on the essential events that occurred in the full-length sports video. There are several postgraduate researches carried under his supervision in the areas of data mining and multimedia systems.

Dr. Lakshman Jayaratne is a Senior Lecturer attached to the University of Colombo School of Computing (UCSC), University of Colombo. Dr. Lakshman Jayaratne obtained his B.Sc (Hons) in Computer Science from the University of Colombo, Sri Lanka in 1992. He obtained his PhD in Computer Science from the University of Western Sydney, Sydney, Australia in 2006. Dr. Jayaratne is currently the principal investigator of the Automated Content Based Audio Music Monitoring Approach for Radio Broadcasting Channels in Sri Lanka, in which his team is investigating an approach to identify songs for radio channels in Sri Lanka in order to ensure the intellectual property rights of the song creators. Further, he is currently the principal investigator of the Computational Approach to Train on Music Notations for Visually Impaired in Sri Lanka where his team is using music notations to enable visually impaired musicians to overcome the barriers they had in learning music, most highlighted in reading and regenerating music notations. For his research contributions, Dr Jayaratne was awarded Vice Chancellor's Award for "Recognition of Excellence in Research (Research Award)" in the year 2013 and 2015 at Postgraduate Convocation of University of Colombo.



Dr. Sellappan is currently Professor of IT and Dean of School of Science and Engineering and Provost of Malaysia University of Science and Technology. He holds a PhD in Interdisciplinary Information Science from University of Pittsburgh (USA), a MSc in Computer Science from University of London (UK), and a Bachelor in Statistics from University of Malaya. His research interests include Data Mining, Machine Learning, Health Informatics, Blockchain, Cybersecurity, Data Analytics and IoT.