Design and Implementation of Statistical Estimation Based Model for Fair Assessment of Rain Interrupted Cricket Matches

Praveen D Chougale
Ramaiah University of Applied Sciences, Bangalore
Praveenmaths89@gmail.com

Abstract—

Cricket has achieved the status of a religion in India due to its huge popularity. The huge amounts of money and interest that cricket garners is increasing the spotlight on making the cricket experience for avid fans more seamless and enjoyable. There is a immediate requirement to come up with a fair assessment method which at any point of the game can decide the winner considering all relevant factors influencing the match. The current model used in rain interrupted matches is the Duckworth-Lewis (D/L) method. In interrupted matches a decision has to be reached within an allocated time of the game and the game cannot be postponed to another day. It has been reported that the D/L method delivers unrealistic target scores for certain cases exhibiting its unfairness and bias towards teams batting second.

The proposed algorithm formulated is an alternate approach that could serve well to reset the target score overcoming this intrinsic problem of the D/L method. This algorithm demands extensive data cleaning and structuring of the raw available data, followed by feature extraction. Exploratory analysis and statistical tests have then been carried out on the independent variables. The developed mathematical functions work for both batting and bowling teams and the neural networks are trained to learn these functions. The developed algorithm is trained and validated for all the completed ODI matches as well as for D/L matches. Accuracy of the model tested on completed ODI matches and for rain interrupted matches is 57 % and 61% respectively. The implemented algorithm can be extended to player selection, modelling using other features (apart from batting and bowling related) to improve the prediction for the rain interrupted matches implementing a D/L method - for fairer evaluation of outcomes.

Keywords-ODI cricket matches, Rain interrupted, Anova, Duck worth Lewis method, Neural network.

I. Introduction

In comparison to other sports, limited overs cricket is particularly vulnerable to inclement weather – when it rains, or becomes too dark, cricket becomes too dangerous to play. Consequently, when a One-Day International

(ODI) or Twenty-20 International (T20I) match is interrupted by rain or bad light, either or both of the competing teams can often not complete their allotted overs. Incomplete games are unsatisfactory for the players and fans alike and, to some extent negate the purpose of the shorter formats since an abandoned match offers Minimal levels of excitement. Furthermore, to enable knockout tournament play, such as the ODI and T20I World Cups, games must reach a positive conclusion.

Therefore, the cricket authorities have adopted quantitative methods to adjust scores and reset targets in order to ensure interrupted matches are concluded with positive results. Rain is the major reason for the interruption of International cricket matches. Interruptions also have a cost and resource implication as the match has to be continued on the next day (reserve day) which may not be as convenient for the fans and organisers. Continuation on another day also implies brand new conditions and environment for the players which could favour a particular team or players. This would give an undue advantage to them and constitute an unfair advantage.

Currently, ICC has approved and accepted Duckworth Lewis method as the way of evaluating and resetting the scores of interrupted matches. For the D/L to be applied, at least 20 overs of the game should have been played in the second innings of the match (Section 12.4.2.B iii of the ICC Rules). Only then, can the target score for the team playing second be reset or the winner of the match can be declared. This proves a huge impediment to the team management and the players as the game strategy has to be drastically altered in the new scenario, which could be favourable to any one team.

II. Related Work

A. Papers /Journals Related To D/L Methods

Research publications are limited to just domestic cricket or only player performances or only one format of the game and also had results with a generalized accuracy of around 50-55%. The researchers in (Bhattacharya, Ghosal and Ghosh, 2018) Bayesian Inference is applied to build a resource table which overcomes the non-monotonicity problem of the current D/L resource table to show that it gives better prediction for teams in first innings score and hence it is more suitable for using in rain affected matches.

For each match they have defined $R(u,w(u))$ as the run scored from the stage in first innings where u overs are available and $w(u)$ wickets are lost until the end of the first innings. They have also calculated $R(u,w(u))$ for all values of u that occurred in the first innings. The estimated resource percentage table is then calculated by averaging $R(u,w(u))$ over all matches where $w(u) = w$ and dividing by the average of $R(50, 0)$ (which is the average first innings score) over all matches. Just like D/L table, this non-parametric resource table suffers from the lack of monotonicity.

Authors of (Shah *et al.*, 2015) have used isotonic regression method to overcome this issue, whereas in (Bhattacharya, Ghosal and Ghosh, 2018) they have taken a parametric Bayesian approach.

Non-parametric model-based resources decay as overs remaining decrease for different wickets. Throughout this paper in resource decay plots index w ∈ W indicates loss of w wickets. Instead of throwing out those columns or rows that have missing entries, (Bhattacharya, Ghosal and Ghosh, 2018) have used the Bayesian inferential framework that provides a natural way of imputing the missing entries using the posterior predictive distribution once a full hierarchical model is specified. Adopting the following nonlinear regression model:

$$\boldsymbol{R}(\boldsymbol{u},\boldsymbol{w})\sim\boldsymbol{N}(\boldsymbol{m}(\boldsymbol{u},\boldsymbol{w};\boldsymbol{\theta}),\boldsymbol{\sigma nuw}),\in\mathbf{u},\mathbf{w}\in\mathbf{W} \quad \textbf{(1)}$$

where $\bar{R}$(u,w) is the sample average of runs scored by a team among the total number of matches considered in the data set and m(u,w; θ) is the corresponding modelled population average of runs scored by a team when a large number of games are taken into consideration and θ denotes a vector of parameters to be specified later in our model. As R(u,w) is not observed for each of the match (in the sample), the average is taken over all those matches, denoted by nuw, over which the sample mean $\bar{R}$(u,w) is calculated. If there is no observation for R(u,w) across all of the matches sampled.

Residual Sum of Squares than the D/L method specially when the match is interrupted in situations where there are lots of overs left is shown. Under the MAR assumption, the proposed Bayesian model provides a natural method to carry out imputations using the posterior predictive distributions which is an advantage over many existing methods (e.g., compared to the non- parametric method). This method is broadly applicable in the sense that it is not restricted to only 50-overs cricket match interruption problem and can be applied many similar sports events. Moreover, the model can be used to estimate the nonlinear mean function of two variables under bi-monotonicity constraint. One future direction for research can be to develop a nonparametric approach for modelling such constrained bivariate functions that is not necessarily based on an exponential decay model. Another alternative method to calculate the revised target in interrupted 50 overs ODI matches is found in (Singh and Adhikari, 2015). Existing D/L method and its modified versions only take available batting resources of the batting team into account and ignore the individual player's excellence to calculate the revised target. Here, it is worth mentioning that individual player's excellence varies in reality, and therefore quality of the available resources may affect the revised target significantly. Furthermore, in D/L method the revised target calculation depends only on the available batting resources of the batting team and does not consider the available bowling resources of the fielding team. Their method overcomes these two shortcomings by taking individual player's excellence and available bowling resources of the fielding team into account. Individual player's excellence has been determined by Data Envelopment Analysis (DEA), a well-known non parametric mathematical programming technique.

Analyzing the D/L method using graphical and mathematical methods and find out the root cause of this unfairness is done in (Scarf and Shi, 2005). Here, the reason for the unfairness of D/L method using graphical methods and chi-square tests is shown. It has been shown that this is due to the inherent nature of the D/L method that use graphs produced using past statistics of all teams of the world. The expected performance specified by those graphs deviate considerably from the actual performance of the teams participating in the game resulting in the unfairness. In each innings, there were at most 50 data points. As such, the degrees of freedom was less than or equal to 49.

B. Papers /Journals related to other variants.

To facilitate the comparison, the absolute values of the differences between the two tables was imputed, and a heat map is produced. The darker shades of the heat map indicate the greatest disagreement between the two tables. On investigating these areas of disagreement, it is observed that the greatest absolute differences occur in three regions. First, large differences occur in the top- right hand corner and bottom-left hand corner of the table. These are precisely the regions where very little or no data are available. These regions are not viewed as too important as the resetting of targets would rarely use these entries. It is interesting however that the non- parametric approach provides more resources in these regions than the D/L approach.

The more interesting discrepancy occurs in the 'middle' of an innings (8–13 overs available with 3–6 wickets lost). In this stage of an innings, the non-parametric approach based on Gibbs sampling suggests that there is up to 5% fewer resources remaining than provided by the D/L method. In 1-day cricket, a team needs to protect its wickets over a longer period of overs. Consequently, up until the middle stage, more resources are conserved in the 1-day game than in Twenty20. They remark that a difference of 5% resources may be very meaningful as a target of 240 runs diminished by 5% gives 228 runs. As more Twenty20 matches become available, authors of (Bhattacharya, Gill and Swartz, 2011) endorse a review of the use of D/L in Twenty20 and the estimation techniques used in the construction of the associated resource table. The method is based on a simple model involving a two-factor relationship giving the number of runs which can be scored on average in the remainder of an innings as a function of the number of overs remaining and the number of wickets fallen (Duckworth and Lewis, 1998). It is shown how the relationship enables the target score in an interrupted match to be recalculated to reflect the relative run scoring resources available to the two teams, that is overs and wickets in combination. The method was used in several international and domestic one-day competitions and tournaments in 1997.

Therefore, need a two-factor relationship between the proportion of the total runs which may be scored and the two resources, overs to be faced and wickets in hand. To obtain this it is necessary to establish a suitable mathematical expression for the relationship and then to use relevant data to estimate its parameters The basis of this method is that it recognizes that the batting side has two resources at its disposal from which to make its total score; it has overs to face and it has wickets in hand. The number of runs that may be scored from any position depends on both of these resources in combination. Clearly, a team with 20 overs to bat with all ten wickets in hand has a greater run scoring potential than a team that has lost, say, eight wickets. The former team have more run scoring resources remaining than have the latter team although both have the same number of overs left to face.

The mechanisms of other methods used for resetting target scores in interrupted one-day cricket matches is explained in (Kampakis and Thomas, 2015). Each of these methods yields a fair target in some situations. None has proved satisfactory in deriving a fair target under all circumstances. We have presented a method which gives a fair revised target score under all circumstances.

This is based on the recognition that teams have two resources, overs to be faced and wickets in hand, to enable them to make as many runs as they can or need. They have derived a two-factor relationship which gives the average number of runs which may be scored from any combination of these two resources and hence have derived a table of proportions of an innings for any such combination. This enables the proportion of the resources of the innings of which the batting team are deprived when overs are lost as a result of a stoppage in the play to be calculated simply and hence a fair correction to the target score to be made.

Though the examples given, both hypothetical and real, it is shown that this method gives sensible and fair targets in all situations. They include the circumstances where overs are lost at the start of the innings, part way through, or at the end of an innings and where the game is abandoned requiring a winner to be decided if Team 2's innings is terminated. The examples have shown the importance of taking into account the wickets that have been lost at the time of the interruption and the stage of the innings at which the overs are lost. Our method was adopted by the England and Wales Cricket Board for the 1997 domestic and Texaco one-day international competitions and the International Cricket Council has used it for several international one-day competitions.

C. Proposed Work

After extensive evaluation and research of the existing models the proposed algorithm has been developed.

The Figure 1 represents the overview of the proposed algorithm. These steps cover end-to-end development, implementation and evaluation of the model from raw data to end user application. The phases of this process are:

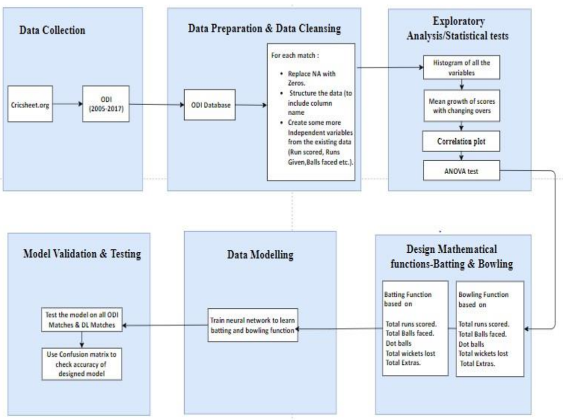

Figure 1 Process flow of proposed algorithm

*1) **Data collection:*** The raw unstructured data was collected from **cricsheet** for 1348 ODI matches. The raw data required to be first deciphered and structured to convert all information into columns with headings and sequential match data that contains the details for teams playing, innings details, number of balls bowled, non- striker, bowler, runs scored by batsmen, extras and the batsman who got out along with the mode of dismissal. All data for all the matches was collated to form a single database.

2) Data Cleansing and Preparation: Raw data requires extensive cleansing and preparation to suit the modelling aspects. Missing data and NA’s are replaced with zeroes. The Column names are renamed and structured and additional columns are derived to give a proper structure to the data. Eg: The column total runs scored on any ball is the sum of runs scored by batsman and extras. The cumulative scores on any for these factors namely, Dot Balls, Extras, Runs Scored and Balls Bowled are also calculated.

3) Exploratory Analysis: To explore the data and interpret the relationship between the variables the data is plotted and statistical testing – ANOVA is performed. This analysis has been done for each and every match.

4) Statistical tests: determine which of the independent variable is affecting the dependent variables. The Dependent variable is “Innings” at two levels – Innings 1/Innings 2. The Independent variables are – “Total runs”, “Total wickets”, “Dot balls” and “Extras”.

In Table 1 the stars (***) represent the significance level. The higher the number of stars the greater is the relevance of the variable. Eg: there is one star (*) for Team1$Totalwickets which has a p-value of 0.0395. Since this value is less than 0.05 it means that this feature is more relevant for the model to use.

Table1

Anova-Innings-1	Df	Sum Sq	Mean Sq F value	Pr(>F)
Team1$Totalruns	1 8.500e-32	8.482e-32	1.734	0.1889
Team1$Totalwickets	1 2.090e-31	2.094e-31	4.281	0.0394 *
Team1$zeros	1 0.000e+00	3.100e-34	0.006	0.9362
Team1$zeros	1 2.300e-32	2.331e-32	0.477	0.4905
Residuals	305	1.492e-29	4.891e-32	
Signif. codes	0 ‘***’ 0.001	‘**’ 0.01	‘*’ 0.05 ‘.’	0.1 ‘ ’ 1

Anova-Innings-2	Df	Sum Sq	Mean Sq F value	Pr(>F)
Team2$Totalruns	1 1.000e-29	1.049e-29	0.409	0.5232
Team2$Totalwickets	1 1.900e-29	1.924e-29	0.750	0.3874
Team2$zeros	1 1.720e-28	1.724e-28	6.720	0.0101 *
Team2$Extra	1 0.000e+00	3.000e-32	0.001	0.9710

Residuals	229	5.875e-27	2.565e-29	
Signif. codes	0 '***' 0.001	'**' 0.01	'*' 0.05 '.'	0.1 ' ' 1

5) S5) Scoring Pattern: At any given point in the match the comparative performance of each team can be known and analyzed at any specific over. This trend analysis of the two teams shows the exact point in the match/ over where the team 2 (Blue – England) overtook the batting scores of the other team (Red – Ireland). Figure 2 shows the comparative scoring pattern of team 1 and team 2 batting trends.

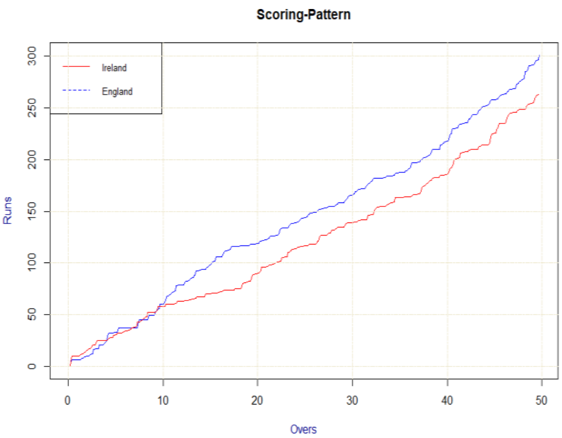

Figure 2 Comparative scoring pattern of both the teams while batting

6) Correlation test*:* A correlation plots between all the variables for the both innings has also been performed and it can be interpreted as:

High correlation existing between variables:
• Total wickets and number of balls faced – 76%

• Zeros and total wickets - 82% Low

correlation between variables:
• Extras and number of balls – 0%

• Sixes and number of balls – 51%

• Wickets and runs – 2%

• Wickets and extras – 0%

The inference from this correlation matrix can be derived that the high importance variables for this sample match are:
• Number of balls

• Runs scored

• Zeros

• Fours

• Twos

The remaining variables hardly contribute to the explanation of the variance of dataset.

7) Design Mathematical Functions (Batting & Bowling): Various functions of different combinations of variables were tried and tested and only the equations 1 and 2 were found to be aptly suitable. Functions for both batting and bowling are designed using the following four variables:
• Number of Runs scored

• Number of Extras

• Number of Dot balls

• Number of wickets lost

$$F(R,W,D,E) = a1 * \frac{Total\ runs}{Total\ Balls\ faced} + a2 * \frac{Total\ extras}{Total\ Balls\ faced} + a3 * \frac{Total\ wickets}{Total\ Balls\ faced} + a4 * \frac{Total\ number\ of\ dot\ balls}{Total\ Balls\ faced}$$

$$F(R,W,D,E) = b1 * \frac{Total\ runs}{Total\ Balls\ faced} + b2 * \frac{Total\ extras}{Total\ Balls\ faced} + b3 * \frac{Total\ wickets}{Total\ Balls\ faced} + b4 * \frac{Total\ number\ of\ dot\ balls}{Total\ Balls\ faced}$$

Where,
$a1$, $a2$, $a3$, $a4$, $b1$, b2, b3, b4 are the constants and should be calculated for the data.

Subsequently a neural network is trained to learn both batting and bowling related functions.

8) Proposed algorithm*:* Algorithm designed for choosing the winner are elaborated in the following steps:

- ✓ ***Step 1****:* Compute the batting and bowling related function values for both the Innings.
- ✓ ***Step 2****:* If the match stops in between during the second Innings, calculate the total function value for both Innings of the teams.

Table 2: Abbreviations used for Mathematical function for both innings

Innings 1
F1 (Batting function-Team 1)
F2 (Bowling function-Team 2)
Innings 2
F3 (Batting function-Team 2)
F4 (Bowling function-Team 1)

- Next is to calculate the total function value of both the Innings.

Total function value for Team 1 = F1 + F4
Total function value for Team 2 = F2 + F3

- ✓ ***Step 3:*** Winner can be decided based on any one the measures
  - Comparing total function value for Team 1 and Team 2.
  - Based on the function value calculate frequency of instances where Team 1 and Team2 are the winners.

9) Data Modelling:

9.1) Train Neural Network: A simple ANN consisting of one input layer two hidden layers and one output layer is used to learn the proposed batting and bowling functions. The four input nodes takes input of total runs, total wickets, extras and dot balls respectively. At a time the output node gives either the batting function value or the bowling function value.

Table 3 Comparison of training parameters used for batting and bowling functions

Parameters	Batting Function	Bowling Function
Inputs	Total runs, Total wickets, Dot balls, Extras	Total runs, Total wickets, Dot balls, Extras
Output	Batting function value	Bowling function value
Hidden Layers	2 Hidden layers (2-3)	2 Hidden layers (4-2)
Steps	52	5623
Mean square error	2.7	6.7

The mean square error is optimal with a value of 2.7 at the 52nd iteration for the batting function and the mean square error is optimal with a value of 6.7 at the 5623rd iteration for the bowling function. The architecture of batting and bowling function is 4-2-3-1, and 4-4-2-1.

9.2) Model Validation & Testing: The designed algorithm after implementation is validated to check its performance accuracy. Such validation is done for 1311 completed ODI matches and also 120 D/L matches.

Table 4 Sample of validation results for 20 completed ODI matches

S.No	Date	Actual result	Predicted result
1	16-08-2016	Scotland	Scotland
2	14-08-2016	Scotland	Scotland
3	19-07-2016	Ireland	Afghanistan
4	17-07-2016	Afghanistan	Afghanistan
5	12-07-2016	Afghanistan	Afghanistan
6	01-09-2016	Australia	Australia
7	31-08-2016	Australia	Australia
8	28-08-2016	Australia	Australia
9	24-08-2016	Sri Lanka	Australia
10	21-08-2016	Australia	Australia
11	06-01-2016	Afghanistan	Afghanistan
12	04-01-2016	Zimbabwe	Afghanistan
13	02-01-2016	Zimbabwe	Afghanistan
14	29-12-2015	Afghanistan	Afghanistan
15	25-12-2015	Afghanistan	Afghanistan
16	10-02-2017	South Africa	South Africa
17	07-02-2017	South Africa	South Africa
18	04-02-2017	South Africa	South Africa
19	01-02-2017	South Africa	South Africa
20	28-01-2017	South Africa	South Africa

Table 5 Confusion matrix for completed ODI matches result against proposed model

Confusion Matrix and Statistics		
	Reference	
Prediction	0	1
0	255	306
1	266	484
Accuracy	0.5637	
95% CI	(0.5363, 0.5907)	
No Information Rate	0.6026	
P-Value [Acc > NIR]	0.9981	
Kappa	0.1008	
Mcnemar's Test P-Value	0.1030	
Sensitivity	0.4894	
Specificity	0.6127	
Pos Pred Value	0.4545	
Neg Pred Value	0.6453	
Prevalence	0.3974	
Detection Rate	0.1945	
Detection Prevalence	0.4279	
Balanced Accuracy	0.5511	
'Positive' Class	**0**	

This confusion matrix which shows the accuracy of the model.

Observations:

- The overall accuracy rate is computed along with a 95% confidence interval is 56.37 %.
- A p-value from McNemar's test is 0.10, which statistically significant.
- The sensitivity and specificity of the model are 48.94 % and 61.27 % respectively, from which it can interpreted that the algorithm is biased towards team batting in innings2.

Table 6 Sample of validation results for 20 ODI matches where D/L was applied

S.No	Date	Actual result	Predicted result	Method
1	03-10-2015	Zimbabwe	Pakistan	D/L
2	04-11-2015	Sri Lanka	Sri Lanka	D/L
3	01-11-2015	Sri Lanka	Sri Lanka	D/L
4	31-01-2016	New Zealand	New Zealand	D/L
5	29-06-2016	England	England	D/L
6	21-06-2015	Bangladesh	Bangladesh	D/L
7	03-12-2014	England	England	D/L
8	30-08-2014	Sri Lanka	Pakistan	D/L
9	23-08-2014	Pakistan	Pakistan	D/L
10	17-06-2014	India	Bangladesh	D/L
11	15-06-2014	India	Bangladesh	D/L
12	27-08-2015	Australia	Australia	D/L
13	20-06-2015	England	England	D/L
14	12-06-2015	New Zealand	England	D/L
15	09-05-2014	England	England	D/L
16	16-11-2013	Sri Lanka	New Zealand	D/L
17	12-11-2013	New Zealand	New Zealand	D/L
18	29-10-2013	Bangladesh	Bangladesh	D/L
19	22-05-2014	England	England	D/L
20	30-08-2014	India	England	D/L

Table 7 Confusion matrix for matches where D/L was applied against proposed model

Confusion Matrix and Statistics		
	Reference	
Prediction	0	1
0	29	16
1	33	42

Accuracy	0.5917
95% CI	(0.4982, 0.6805)
No Information Rate	0.5167
P-Value [Acc > NIR]	0.05989
Kappa	0.1901
Mcnemar's Test P-Value	0.02227
Sensitivity	0.4677
Specificity	0.7241
Pos Pred Value	0.6444
Neg Pred Value	0.5600
Prevalence	0.5167
Detection Rate	0.2417
Detection Prevalence	0.3750
Balanced Accuracy	0.5959
'Positive' Class	**0**

This confusion matrix which shows the accuracy to be 59.17%. in this case also specificity is higher i.e. 72.41% from which it can interpreted that the completed ODI matches and D/L matches the batting in second innings is more difficult in D/L.

Observations:

- The overall accuracy rate is computed along with a 95% confidence interval is 59.17 %
- A p-value from McNemar's test is 0.05, which statistically significant (Reject the null hypothesis and accept the alternate hypothesis)
- The sensitivity and specificity of the model are 46.77 % and 72.41 % respectively, from which it can interpreted that the algorithm is biased towards team batting in innings 2.

III.Conclusion

There are several competing methods for interrupted matches among which the D/L is the most popular and widely accepted. This method too has its pitfalls.

This paper attempts to overcome these pitfalls by developing a statistical function related to both batting and bowling considering four factors which influence match outcomes. There are other factors like pitch conditions, winner of the toss, opponent team composition etc. which are not considered due to unavailability of sufficient data. These factors can be incorporated in the proposed model to improve the model outcomes.

A Machine learning algorithm is then developed by applying the formulated statistical estimation technique to decide the match outcome. The implemented model is then validated /tested on the past D/L applied rain interrupted ODI matches. The accuracy of the model for both completed ODI matches and D/L matched is 57 % and 61% respectively.

References

[1] Bailey, M. and Clarke, S. R. (2006) 'Predicting the match outcome in one day international cricket matches, while the game is in progress', in *Journal of Sports Science and Medicine*, pp. 480–487.

[2] Collins, M., Schapire, R. E. and Singer, Y. (2002) 'Logistic regression, AdaBoost and Bregman distances', *Machine Learning*, 48(1–3), pp. 253– 285. doi: 10.1023/A:1013912006537.

[3] *Cricsheet*(nodate).Available:https://cricsheet.org/downloads/allmatches.

[4] Dangeti, P. (2017) Statistics for Machine Learning: Techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R, Development.

[5] Emrouznejad, A. and Anouze, A. L. (2010) 'Data envelopment analysis with classification and regression tree - A case of banking efficiency', *Expert Systems*, 27(4), pp. 231–246. doi: 10.1111/j.1468-0394.2010.00516.x.

[6] Kampakis, S. and Thomas, W. (2015) 'Using Machine Learning to Predict the Outcome of English County twenty over Cricket Matches', in. doi: 10.1360/04yd0007.

[7] Kevin Markham (2014) 'Simple guide to confusion matrix terminology', Data School. Available at: http://www.dataschool.io/simple-guide-to- confusion-matrix-terminology/.

[8] Lokhande, R. and Professor, A. (no date) 'Live Cricket Score and Winning Prediction', International Journal of Trend in Research and Development, 5(1), pp. 2394–9333. Available at: www.ijtrd.com.

[9] S., G., R., H.-W. and H.P., S. (2010) 'A support vector machine for decision support in melanoma recognition', Experimental Dermatology, 19(9), pp. 830–835. doi: 10.1111/j.1600-0625.2010.01112.x.

[10] Science, C. (2015) 'Predicting a T20 cricket match result while the match is in progress Authors Name Fahad Munir (11201014) Md . Kamrul Hasan (11201032) Sakib Ahmed (11201009) Sultan Md . Quraish (11201017) Supervisor Rubel Biswas Moin Mostakim A thesis presented fo', (11201014).