International Journal of Knowledge Based Computer Systems
7 (1), June 2019, 01-09
http://www.publishingindia.com/ijkbcs/

Context Based Cricket Player Evaluation Using Statistical Analysis

Vaibhav Khatavkar[1*] and Parag Kulkarni[2]
[1]Research Scholar and Assistant Professor, Department of Computer Engineering and Information Technology, College of Engineering, Pune, Maharashtra, India. Email: vkk.comp@coep.ac.in
[2]Adjunct Professor, Department of Computer Engineering and Information Technology, College of Engineering, Pune, Maharashtra, India. Email: paragindia@gmail.com
*Corresponding Author

Abstract: In business, making decisions on the spot is a critical task since data is available online. Same applies for sports. For managers and coaches selection of best fit player in the team has to be made. Many times, team managers have to bid for the players. Many statistical methods are applied in order to evaluate a player for sports, specially, baseball, and football. This paper focuses on Cricket since it has been deprived of good statistical methods. Typically, for cricket, Deep Performance Index (DPI) method is used to evaluate a player. But it has some limitations. This work aims to assimilate the best features of the previous approaches and losing the drawbacks. An algorithm is devised for statistical analysis of cricket data and particularly Indian Premier League (IPL) data which can be later extended to other sports with minor customizations. The algorithm considers context of the player in terms of indices like most valuable player index for batsman and for bowler. The important context introduced is winning contribution ratio of the player. The algorithm is tested on IPL and proves to be comparable with DPI. Hypothesis testing is performed to validate the results.

Keywords: Context, Cricket data mining, Hypothesis testing, Indian Premier League (IPL), Statistical analysis.

I. Introduction

In today's era data analysis plays an important role in the field of Information Retrieval (IR). IR is widely used in business applications where in data can be collected and decisions can be made in real time. Khatavkar and Kulkarni identified context of a document with the help of Latent Semantic Analysis [11]. In their work they used Singular Value Decomposition of Semantic Analysis. The idea of representing documents in term of forest of graphs is proposed [12]. The researchers [15] focused on the current trends in document analysis. They concluded that context of a document plays a crucial role in data analysis. In their work [14] presented the use of context vectors in document analysis. When a query is fired for a term the systems computes context vectors and gives terms related to the query fired. Deepali extracted named entities from a document in order to find theme of a document which is context of a document [13].

Statistical analysis is instrumental in determining the trends and predicting the future results. The sports sector is ultra-dynamic and hence the statistical analysis of sports data is a challenging task at the same time this analysis is a necessity because creating a strategy against a team or for a tournament demands an in-depth study of opponent team(s) as well as of the team players on both sides. Baseball has adapted to this need the earliest and hence is ahead of the curve. There are many approaches available for statistical analysis of baseball data and there a dearth of options for other sports.

Football (one of the most popular sports across the globe) has always resisted the use of technology but now with the advent of goal line technology and the investment in the development of statistical analysis methods it's catching up. Some attempts have been made to develop and implement an approach for statistical analysis of cricket data, but much work is needed. Indian Premier League (IPL), which started in 2008 has gained immense popularity and is a commercial hit. IPL structure involves an auction where team owners and executives bid on a player with certain base price. The auction in itself is a complex game which demands readily available statistics with in depth analysis. The player ratings, the player's current form and the highest bid the player is worthy of are to be determined, this motivated us to develop an approach for determining the value of a player by statistically analyzing the data available and predicting his performance in near future.

The current methods available heavily depend on the individual contribution of the player and the parameters considered are found lacking while determining the players contribution to the result. Currently the parameters considered are average, runs and strike rate for a batsman and average, wickets and economy for a bowler. We aim to devise an approach for evaluating a player based on the venue of match, overall average, current season average and additional parameters. The results of this approach will vary, because every sport has a complex time-line which changes the result, pertaining to every minute change in the event time-line. For example, in cricket every ball has more than seven possible outcomes. Thus, this work aim to devise an approach with defined constraints and highest possible accuracy to evaluate a player and his contribution to the team. We intend to develop an approach for statistical analysis of IPL data based on the earlier approaches in various sports, which can be worked upon to be applied to the overall sports scenario.

This motivates us to derive context of the player in sports specifically in cricket. The context will be defined with various indices like the most valuable batsman, bowler etc.

The paper is arranged as follows: Section II gives the work done in the area of context and sports. Section III explains the proposed system followed by experimentation and results in Section IV. Section V concludes the paper with future scope.

II. Literature Survey

A lot of statistical analysis has been done on two games namely baseball and cricket. The empirical analysis of baseball is known as Sabermetrics. It is baseball statistics that is used to measure in-game activity. This data obtained from in-game activity is collected by Sabermetricians. Player Empirical Comparison and Optimization Test Algorithm, also known as PECOTA is a sabermetric system for predicting the performance of major league baseball. PECOTA also became the inspiration for some similarly functioning projection systems for other professional sports [1]. KUBIAK for the national football league, SCHOENE and CARMELO for the National Basketball Association and VUKOTA for the National Hockey League are some examples. The four major categories of attributes used for determining a player's comparability, according to this tool are Production metrics, Usage metrics, phenotypic attributes and fielding position. Usually, the database is very large, enough to provide a meaningfully huge set of suitable comparables. In case of absence of an appropriate comparable, the designers design program to 'cheat' by extending its tolerance [4] i.e., by including dissimilar players for comparison to reach a reasonable sample size [1]. Primarily, a three-year window of a pitchers performance is used to predict the similarity scores in PECOTA [1]. Hence, it enables us to look at a pitchers performance in age 35 to 37, and compare the same with the most similar performances of the same age (35 to 37). An adjustment has to be made for parks, league effects, and a lot of other factors. On the other hand, James Method designs the program to evaluate the totality of a player's performance throughout his career. The similarity scores obtained by this method are thus different from those of PECOTA. A set of "comparables" once established for a player are used to predict his future performance based on the past performance of his "comparables". For example, the past performance of the most comparable major league 26-year-old in their subsequent seasons will form the basis for predicting the future performance of 26-year-old player in the coming season [1].

Another tool used in Baseball statistical analysis is Value Over Replacement Player (VORP). VORP is a relative value statistics used to present the offensive contribution of a hitter or the contribution of a pitcher to his team in comparison to a theoretical "replacement player", considered to be an average fielder at his position and also a below average hitter.

VORP is not a projected statistic, but a form of cumulative statistics (counting statistics). Take for an example, a player having a VORP of +30 runs after 90 matches, has contributed 30 more runs of offense to his team compared to what the fictitious replacement player would have, over 90 games. The players VORP will either increase or decrease based his performance throughout the rest of the season and will then attain a final figure, for example +40.

A non standardized tool named WARP is also used in Baseball. To ascertain a player's total contributions to his team, a non-standardized statistical tool has been developed, known as WAR (Wins above Replacement) or Wins Above Replacement Player (WARP). This value is determined by determining the number of additional wins achieved by the player's team above the expected number of team wins if a replacement-level player substituted the player in question. (A replacement player may be known as a player who costs minimal cost and effort when added to the team). The number and success rate of the on-field performance by a player determines the WAR value of the player. Larger contributions of the player to the team's success, which can be either in terms of playing time or in terms of successful matches, are indicated by higher WAR values. The WAR value is determined on the basis of the appraised number of runs won by a player in batting and base running and runs saved by the player in fielding and pitching. Ultimate Zone Rating (UZR), Ultimate Base Running (UBR) and Fielding-Independent Pitching (FIP) are some more statistics used to get an approximate of the runs made and runs defended by the player. This can be estimated by multiplying these statistics and the players playing time. But, this WAR method is prejudiced to favor past players because of the higher variance in skill levels at the time.

Sometimes Averaging is also used in baseball analysis. Clay Davenport invented this method. This method was developed to determine the production of hitters exclusive of league and park factors. It uses the same scale as batting average to project a hitter's productivity. The league average is defined to be an EqA of 0.260. Baseball Prospectus renamed equivalent average as True Average (TAv).

$$EqA = \frac{H + TB + 1.5*(BB + HBP) + SB + SH + SF}{AB + BB + HBP + SH + SF + CS + (SB/3)} \quad (1)$$

Raw EqA (REqA) = (H + TB + 1.5* (BB + HBP + SB) + SH + SF) / (AB + BB + HBP + SH + SF + CS + SB) (2)

The expression for equivalent average is in Equation (1).

Equation uses the minor league statistics to calculate the same for the major league statistics. The level of competition and influencing parameters in minor leagues are not the same as those of major league. Hence, to get a better estimate for minor league, equation is derived from the raw equation, which is given in equation (2). Baseball prospectus uses Equation enormously and predicts the equation for each hitter in its annual PECOTA forecasts.

A Markov chain can be considered to be represented by a finite state machine.

$$(Xiow0,\ldots.Xiow7)\sim$$
$$multinomail(miow;piow0,\ldots,piow7) \quad (3)$$

The Markov chain approach for baseball is given [2]. Firstly, they defined the states of baseball with identical players from the view of a whole game. It is essential to determine the State space for formulation and calculation of the probability of winning in a whole nine innings which is given as:

*433 statesredefined - innings (9 * 2possibilities) * numberofoutspossible (3possibilities) * the patternofbasesoccupied (8possibilities) + theendofthegame (1possibility)*

For cricket Davis *et al.* in their work [3] used Relative value of statistics. These statistics consider the run differential to be the key measure of a team's performance. A player can be evaluated by considering the teams run differential in his presence and absence from the team lineup. But practically there will be other factors that will change too, hence in the simulation the other changes are ignored focusing only on the player in question. Every ball can have one of these 8 outcomes:

Outcome j = 0 0 runs scored

Outcome j = 1 1 runs scored

Outcome j = 2 2 runs scored

Outcome j = 3 3 runs scored

Outcome j = 4 4 runs scored

Outcome j = 5 5 runs scored

Outcome j = 6 6 runs scored

Outcome j = 7 dismissal

Extras occur at the rate of 5.1% in twenty-20 cricket and hence in this simulation the states have been limited to the non extras scenario and adding up the extras based on probability later.

Enumeration of batting outcomes mentioned above gives the model mentioned in equation (3). Xiowj is the number of occurrences of outcome j by the i^{th} batsman during the oth over when w wickets have been taken, miow is the number of balls faced by the batsman i. The data was obtained by parsing the commentary logs of matches listed on cricinfo website. The dataset considered all the twenty-20 matches involving full member nations of the ICC played in the period from January 2005 to August 2014. This approach succeeded in giving additional weightage to real time considerations of match status. The approach restricted the parameters in play to an ideal condition of assuming a particular lineup. It used a dataset which is based on individual contributions and used a team approach for evaluation which has led to major deviations in results obtained.

Another machine learning based approach which created a new index, named as Deep Performance Index (DPI), which reflects the performance of the batsmen and bowlers on a deeper analysis of the focused requirements of twenty-20 cricket is given [5]. The recursive feature elimination algorithm based on machine learning is used for extracting meaningful features and their relative importance towards designing the DPI after assigning the required weights.

The researchers [5] considered IPL performance data up to IPL 7 and the overall T20 career data up to the end of IPL 8 of all the players participating in IPL [5]. Batsman scoring more than 500 runs in twenty-20 internationals with a strike rate exceeding 100 and who have played in at least 25 matches were considered for batting ranking. Applying these criteria, 89 batsmen were considered for this rating system.

The constraints for assimilating a bowler into this system where that he must have bowled at least 30 over in their twenty-20 international careers and must be playing in IPL 8. Applying these criteria, 120 bowlers were considered for this rating system.

Various indices have been obtained by using a machine learning approach. Machine learning based approach is used to determine the correlation between the indices to determine the indices which can represent the effect of other indices by using a weight. Caret package in R was used for feature selection which resulted in selection of 5 indices that are more relevant to the target measure. A weighted function is obtained by computing the relative weights of the features. It is an improvement over the older approaches but fails to take into consideration the effect of venue on the player's performance. Venue is an important factor while evaluating a player, taking into consideration the home and away performance.

Other approaches are also used like MVPI (Most Valuable Player Index) is often quoted index in the case of IPL. It has been popularized by rediff cricket [7]. One more approach is MCDM (Multiple Criteria Decision Matrix). This approach by Dey *et al.* uses a MCDM approach for evaluating a player's performance [8]. This approach uses normalized matrices for evaluating multiple criteria. MCDM approach gives almost 25% weightage to factors that are a function of the length of a player's career rather than the performance. Principle Component Analysis (PCA) is also performed on cricket analysis [9]. For batsmen higher value indicates better performance. PCA gives nearly equal weightage to all the parameters (unlike MVPI) but the values vary widely, For example, the number of boundaries scored by a batsman may be a lot more than the number of centuries scored. In this approach smaller value of component indicates better performance for a bowler. Haseeb *et al.* applied generative and discriminative machine learning algorithms to predict rising stars in cricket [16].

III. Proposed Work

The proposed system is shown in Fig. 1. The proposed system first mines the data from official IPL cricket website and creates a database of 9 seasons. When a query for player evaluation is fires the database is accessed and then context player evaluation engine will execute algorithm 1 in order to rank the player. The player ranking will be the output of the system. For ranking the player following metrics proposed are:

A. Batting Rating Points System (BRPS)

BRPS = *WCR*+*AV I* +

(*BCSA*/*BOA***BCSSR*/*BOSR*) + *HPR*

Where,

WCR = Winning Contribution Ratio,

AVI = Average Venue Index,

BCSA = Batsman Current Season Average,

BOA = Batsman Overall Average,

BCSSR = Batsman Current Season Strike Rate,

BOSR = Batsman Overall Strike Rate,

HPR = Hitting Power Ratio.

WCR is the contribution of the player to his team's victory; this parameter has never been considered earlier. WCR is equal to the batsman average in winning matches divided by his overall average. There have been repeated instances where a player has performed well but his team lost the match, and there have been instances where a player has performed when his team needed it but, since he scored less runs and came down the order (the finishers) his average is affected, to avoid this we have calculated WCR.

WCR = (*BatsmanAverageinwinningMatches*/*OverallAverageofthebatsman*) * (*Strikerateinwinningmatches*/*OverallStrikerate*)

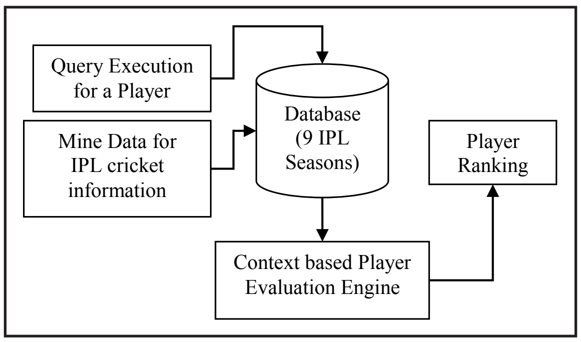

Fig. 1: Proposed System

Venues have been known to affect a player's performance significantly. AVI helps us determine the effect of venue on a player's performance. To calculate the average of a batsman at each venue is needed and his overall average. VI for a particular venue is determined by dividing the batsman's average at that venue by the overall batting average at that venue.

VI = (Batsman's Average at that venue) / (Overall Batting Average at that venue)

It's a known fact that overall average can be misleading in many cases and hence some other parameter is needed and that parameter is BCSA. BCSA gives the current season average of a batsman which is more effective.

BOA is the overall average of a batsman. BOA has significant dependence on the number of matches played by the player and most of the times decrease as the number of matches played increases. Strike rate is an important factor, especially in IPL, twenty-20 and ODIs. We have used two parameters BOSR, the overall strike rate and BCSSR, the current season strike rate. BCSSR is more effective because the current form is what defines the player's current efficiency and value.

For accurate analysis of the hitting power of a player i.e. his fast scoring probability we have devised another parameter HPR, the hitting power ratio of a player, its calculated by multiplying the number of boundaries by 4 and number of sixes by 6 and then dividing the product by total number of balls faced by the batsman.

HPR = (4 * No. of 4s + 6 * No. of 6s) / No. of balls faced

Applying the formula to these parameters, a positive value is obtained according to which the players are ranked on a scale of 10. Higher the rating points, better the performance, higher the value. Batsman having rating points 6 and above are deemed to be high performing players and between 5 and 6 are deemed to be moderately high performing players and those between 4 and 5 are considered to be average players while those having points below 4 are considered to be low performing players.

B. Bowling Rating Points System (BoRPS)

BoRPS = *WCR* + *AV I* + (*BoOA*/*BoCSA* * *BoOE*/*BoCE*) + *WMR*

Where,

WCR = Winning Contribution Ratio,

AVI = Average Venue Index,

BoCSA = Bowler Current Season Average,

BoOA = Bowler Overall Average,

BoOE = Bowler Overall Economy,

BoCE = Bowler Current Season Economy,

WMR = Wickets Matches Ratio.

It's difficult to determine the actual contribution of a bowler to a team's performance and hence for better results, WCR for bowlers is found out using the overall average, average in the matches won by his team, overall economy and the economy in the matches won by his team. The important thing to be considered while calculating the ratings points for a bowler is that lower the average, better the bowling performance which is contrasting to the case of batsman.

WCR = (*Overall Average of the Bowler/Bowler's Average in winning Matches*) + (*Overall economy of the Bowler /Economy in winning matches*)

Venue index is helpful in taking into consideration the effect of pitch conditions and venue on a player's performance. VI for a bowler at a particular venue is the ratio of the average Bowling Average of all bowlers on that venue and the bowling average of the bowler in consideration at that venue. AVI is the average of venue indices of all the venues the bowler has bowled at. VI helps to determine the performance of a particular bowler at a particular venue which helps in team selection.

V I = Bowling Average at that venue /Bowler's average at that venue

In cricket, especially in twenty-20 the current form matters the most and hence the parameters BoCSA and BoCE are used. It's often said that form is temporary but class is permanent and hence the parameters BoOA and BoOE are used. Wickets matches ratio is the ratio of total wickets taken by a bowler and the number of matches played by the bowler; it helps in determining the consistency of a bowler over a period of time.

WMR = no. of wickets taken / no. of matches played

All Rounder Rating Points System (ARPS)

ARPS = (*BRPS* + *BoRPS*)/2

For an all rounder, the rating points can be easily found by taking the average of those players Bowling rating points and Batting rating points.

Algorithm 1: Context Based Player Evaluation

1. Get data of batsman and bowler, store them in database from the official cricket website for IPL.
2. Fire a query about a player.
3. Perform rating using metrics namely: BRPS, BoRPS, and ARPS.
4. Rate the player for which query was fired.
5. Use Hypothesis Testing in order to find correctness of result.

IV. Experimentation and Results

The approaches used earlier have been based on strike rate or average for a batsman and on average or economy for a bowler, problem being the strike rate of a batsman will be high even if he hits a six of the first ball and gets out of the next and overall average can't determine the current form of a batsman.

Keeping into mind this point the approach using run differential was developed but it has unnecessarily high number of constraints. In order to increase the efficiency of the approach we have focused on IPL and applied it to the same.

After mining data from the IPL cricket website for all 9 seasons was collected and validated. Then the context based player evaluation algorithm was applied for each player which used rating system proposed in the paper for evaluating the performance of a particular player. Instead of considering only the normal parameters, the number of parameters were expanded thereby increasing the accuracy and efficiency of proposed approach.

The proposed system considered the players who have played at least 15 matches and recorded their data from cricket information's website [10]. Overall 46 bowlers and 39 batsmen were included in the dataset.

The data of 9 seasons of IPL was assimilated for 39 batsmen and for 46 bowlers from espncricinfo website [10] and exported to a .csv file for further operations. The input file for batsmen had information regarding the players name, number of matches played, runs scored, overall strike rate, overall average, strike rate in winning matches, average in winning matches, season average and strike rate, number of fours and number of sixes, venues and respective averages, using these parameters we calculated the ratings points of each batsman. The input file for a bowler had information regarding the players name, number of matches played, number of wickets taken, overall average, overall economy, average in winning matches, economy in winning matches, previous season average and economy, 3 wicket hauls, 5 wickets hauls, venues and respective bowling averages, using these parameters we calculated the ratings.

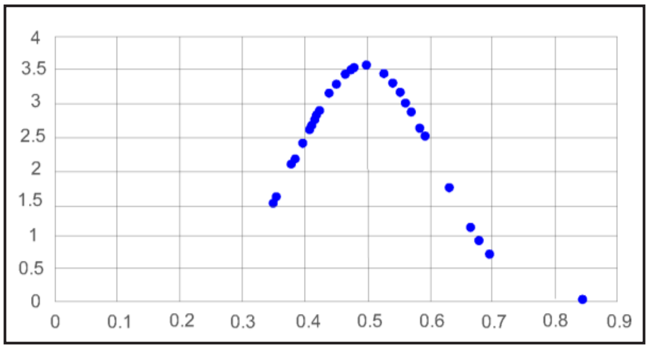

Fig. 2: Normal Distribution for Batsman Data

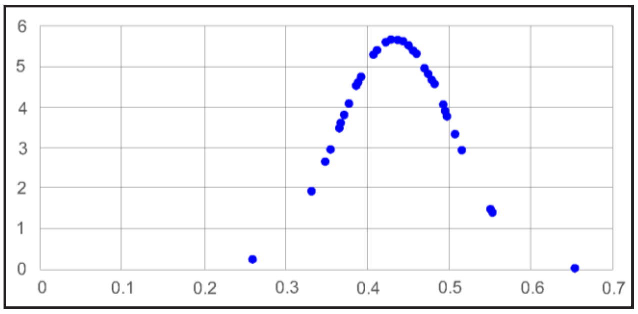

Fig. 3: Normal Distribution for Bowler Data

The results obtained were exported to a .csv file with all the input parameters duly mentioned. For calculating the ratings points for an all-rounder, we took the average of that players batting and bowling ratings points. We calculated the VI for every venue for a player, WCR and HPR. Each metric was calculated. The values of BRPS vary from 3.47 to 8.42 and the values of BoRPS vary from 2.58 to 6.54. The parameters used vary too and hence the dataset has been normalized. Normalization of data restricts the range of values from 0 to 1. Normalization of data made it possible for us to use central limit theorem to accurately determine the parameter dependency. As evident from Fig. 2 a bell shaped curve is obtained with mean being 0.494466601 and the standard deviation being 0.111345553 for normalized data for batsman. As evident from Fig. 3 a bell-shaped curve is obtained with mean being 0.434047477 and the standard deviation being 0.070081851 for normalized data for bowlers.

Hypothesis is used in case of two mutually exclusive statements to determine the statement most supported by the sample data. To determine the correctness and relative accuracy of the results obtained by our approach and those obtained by statistical methods.

a. Null Hypothesis: $H0 : \mu = 0.49$

b. Alternate Hypothesis: $H1 : \mu = 0.49$

c. $\alpha = 0.01$

d. Critical Region: $z < -2.575$ *and* $z > 2.575$
where $z = (x - \mu)/(\sigma/\sqrt{n}\,)$

Computation: x = 0.513, n = 10, $\sigma = 0.11$

Hence, z = 0.5720

Decision: Accept $H0$. Implying the average batsman rating is 0.49

1. Null Hypothesis: $H0 : \mu = 0.43$

2. Alternate Hypothesis: $H1 : \mu = 0.43$

3. $\alpha = 0.01$

4. Critical Region: $z < -2.575$ *and* $z > 2.575$
where $z = (x - \mu)/(\sigma/\sqrt{n}\,)$

Computation: x = 0.45, n = 10, $\sigma = 0.11$
Hence, z = 0.4597

Decision: Accept $H0$. Implying the average bowling rating is 0.43

Two tailed test is used to test if the sample is greater than or less than certain range of values. The range of values for batsmen varies from 0.34 to 0.84 and the range of values for bowlers varies from 0.25 to 0.68 and to accept these values we have determined the critical region using two tailed test.

Various metrics were operated on and the results were stored in csv files along with the input data for reference. The data obtained was normalized as explained in Section IV for the sake of efficient analysis. The normalized data was validated using hypothesis testing and two tailed testing. The players were ranked according to the output obtained.

According to the analysis the batsman rankings are depicted in Table III. Andre Russell has been ranked first thanks to his high average, strike rate and HPR. David Warner, the explosive Australian opener has secured the second spot thanks to his consistent big hitting at the start of the innings. The first two batsmen are contrasting in the number at which they come to bat. Russell comes to bat late in the innings while Warner opens the innings this confirms the unbiased nature of our approach as every contribution is evaluated without any bias of the position. Virat Kohli, who has been in a great form since 2014 has been ranked fifth because of his decent performance in first few seasons of IPL as only the IPL data has been considered. AB de Villiers is at ninth position due to the lack of consistent performance in IPL while Dwayne Smith made it to the top ten due to his consistent run in IPL when he opened for Chennai Super Kings (CSK). According to the analysis the bowler rankings are depicted in Table IV. Mitchell Starc the Australian fast bowler is ranked first as a result of his excellent and consistent bowling performance both at the start of innings and in the death overs irrespective of the venue. Ashok Dinda has been ranked second despite his high economy rate.

Because of his better performance in the death overs and his contribution to team's victory. Tim Southee has made it to the top three due to his excellent economy rate and adaptability in the lineup. JP Dumniy has performed well every time he has been called on to bowl, he has got fewer wickets than other bowlers because he has bowled less but his performance has been excellent. Mitchell Johnson has been a consistent wicket taker and has done so on crucial occasions and hence has been ranked fifth.

Mohammad Shami has played less number of matches due to injuries and other reasons but has performed exceedingly well in whatever matches he has bowled, his excellent death over performance has landed him the sixth position in our rankings. Ishant Sharma has been ranked seventh due to his performances in the earlier editions of IPL, despite of the fact that he has not performed well in last season. Because he is a bankable player. Jaidev Unadkat has been ranked eighth because of his excellent performances in the recent IPL seasons and his performance in IPL 10 confirms his improving performance. Andre Russell has been ranked ninth because of his excellent bowling during death over for KKR and his match winning capability. Nathan Coulternile is a tricky customer and hence hasn't been ranked higher. He has been amongst the wickets and has been a match winning player and that has landed him a place in top ten but his high economy rate has kept him at tenth position.

There isn't a separate ranking system for IPL in particular and thus it's difficult to correlate our results. DPI was an improvement over the earlier approaches and hence we compared our results with theirs. It considered the parameters as denoted in Table I. DPI has used a weighted approach and was done for IPL 8 we applied their approach to the dataset we compiled of the nine seasons of IPL and the results we obtained were as Table I and II. It can be observed from the tables that its weighted approach for average and strike rate has resulted in players who have played more matches have been ranked in top 10. The 2017 year's performance of these players has been below par [6]. The bowling ratings too have been dominated by the players who have played more number of matches. Irfan Pathan who has been ranked tenth in this index, went unsold in 2017 year's auction [6]. Our results too have been found flawed for particular cases due to the lack of data available for many players. But some results have been spot on For Example: David Warner, who is a top contender for the orange cap 2017 year.

Table I: Top 10 Batsman According to Existing DPI

Batsman	DPI Ranking
Suresh Raina	2.7391
Chris Gayle	2.5882
Shane Watson	1.5833
Robin Uthappa	1.4034
Rohit Sharma	1.3265
MS Dhoni	1.2366
Shikhar Dhawan	1.0077
Virat Kohli	0.9688
Brendon McCullum	0.8829
Parthiv Patel	0.83510

Table II: Top 10 Bowlers According to Existing DPI

Bowler	DPI Ranking
Harbhajan Singh	3.881
Lasith Malinga	3.4142
DJ Bravo	3.2453
Ravichandran Ashwin	3.1094
Sunil Narine	2.8515
Amit Mishra	2.7556
Mitchell Johnson	2.66637
Umesh Yadav	3.468
James Faulkner	3.379
Irfan Pathan	3.10

Table III: Top 10 Batsman According Proposed Method

Batsman	Rating Points	Ranking
Andre Russell	8.42	1
David Warner	6.94	2
Faf du Plessis	6.785	3
Quinton de Kock	6.76	4
Virat Kohli	6.63	5
George Bailey	6.28	6
Wriddhiman Saha	5.88	7
Ajinkya Rahane	5.686	8
AB de Villiers	5.598	9
Dwayne Smith	5.494	10

Table IV: Top 10 Bowlers According Proposed Method

Bowler	Rating Points	Ranking
Mitchell Starc	4.63	1
Ashok Dinda	4.62	2
Tim Southee	4.37	3
JP Duminy	4.33	4
Mitchell Johnson	4.04	5
Mohammad Shami	3.91	6
Ishant Sharma	3.9	7
Jaidev Unadkat	3.79	8
Andre Russell	3.74	9
N. Coulternile	3.70	10

Table V: Comparison of Existing DPI and Proposed Ranking System for Bowlers

Bowler	DPI Rank	Bowler	Proposed System Rank
Harbhajan Singh	1	Mitchell Starc	1
Lasith Malinga	2	Ashok Dinda	2
DJ Bravo	3	Tim Southee	3
Ravichandran Ashwin	4	JP Duminy	4
Sunil Narine	5	Mitchell Johnson	5
Amit Mishra	6	Mohammad Shami	6
Mitchell Johnson	7	Ishant Sharma	7
Umesh Yadav	8	Jaidev Unadkat	8
James Faulkner	9	Andre Russell	9
Irfan Pathan	10	N. Coulternile	10

Table VI: Comparison of Existing DPI and Proposed Ranking System for Batsmen

Batsman	DPI Rank	Batsman	Proposed System Rank
Suresh Raina	1	Andre Russell	1
Chris Gayle	2	David Warner	2
Shane Watson	3	Faf du Plessis	3
Robin Uthappa	4	Quinton deKock	4
Rohit Sharma	5	Virat Kohli	5
MS Dhoni	6	George Bailey	6
Shikhar Dhawan	7	Wriddhiman Saha	7
Virat Kohli	8	Ajinkya Rahane	8
Brendon McCullum	9	AB de Villiers	9
Parthiv Patel	10	Dwayne Smith	10

V. Conclusion

The proposed work lies in the domain of statistical analysis and performance evaluation. While researching about the different approaches for the objective attainment the flaws with these approaches surfaced. An extensive study of these approaches and the various methodologies implemented by these approaches was carried out to determine the best possible approach. A dataset with parameter information hitherto uncompiled was built. An approach to efficiently evaluate a player's performance based on statistical analysis was proposed and it was implemented upon a dataset considering 39 batsmen and 46 bowlers with their statistics for the 9 seasons of IPL.

The result of DPI for concerned dataset was obtained and then compared with the obtained results of our approach.

Table V and VI compares the ranking of Bowlers and Batman with existing Deep Performance Indexing (DPI) and our proposed ranking algorithm. As shown in Table V and VI, the ranking of players change when context is taken into consideration.

Additional parameters WCR, HPR, WMR and VI were devised and calculated. For the first time the effect of venue was quantified. These parameters are considered as context along with the venue in order to evaluate the player.

Acknowledgements

I acknowledge all authorities of College of Engineering, Pune for their kind support and encouragement for research work. I am also thankful to Mr. Sushant Chavan, Mr. Anurag Bhikane, Mr. Shrikrishna Kulkarni and Mr. Pruthiraj Parade Patil who gave their time for brain jamming sessions.

References

1. "PECOTA," En.wikipedia.org, [Online]. Available: https://en.wikipedia.org/wiki/PECOTA
2. N. Hirotsu, and M. Wright, "A markov chain approach to optimal pinch hitting strategies in a designated hitter rule baseball game," *Journal of the Operations Research*, vol. 46, no. 3, pp. 353-371, 2003.
3. J. Davis, H. Perera, and T. B. Swartz, "A simulator for twenty-20 cricket," *Australian and New Zealand Journal of Statistics*, vol. 57, no. 1, pp. 55-71, 2015.
4. Method and apparatus for detecting selected absence of digital logic synchronism, by J. L. Williams. (1991, May 28). Patent US5020024A [Online]. Available: https://www.google.si/patents/US5020024
5. C. D. Prakash, C. Patvardhan, and S. Singh, "A new machine learning based deep performance index for ranking IPL t-20 cricketers," *International Journal of Computer Applications*, vol. 137, no. 10, pp. 42-49, March 2016.
6. "Indian premier league twenty-twenty cricket matches." Iplt20.com. [Online]. Available: https://www.iplt20.com
7. "Cricket rankings," Rediff.com, [Online]. Available: https://www.rediff.com/cricket/report/icc-world-cup-de-villiers-maintains-big-lead20150320.htm
8. P. K. Dey, D. N. Ghosh, and A. C. Mondal, "An MCDM approach for evaluating bowlers performance in IPL," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 11, pp. 563-573, November 2011.
9. A. B. W. Manage, and S. M. Scariano, "An introductory application of principal components to cricket data," *Journal of Statistics Education*, vol. 21, no. 3, 2013.
10. ESPNcricinfo, [Online]. Available: https://www.espncricinfo.com
11. V. Khatavkar, and P. Kulkarni, "Document context identification using latent semantic analysis," 3rd *Int. Conf. on Data Management, Analytics and Innovation*, Springer Conference, January 2017.
12. V. Khatavkar, and P. Kulkarni, "Context vector machine for information retrieval," *Advances in Intelligent Systems Research*, vol. 137, pp. 375-379, Atlantis Press, 2017.
13. D. Nagarale, V. Khatavkar, and P. Kulkarni, "Document theme extraction using NER," *ICCASP*, Springer, January 2018.
14. V. Khatavkar, and P. Kulkarni, "Usage of context vectors in document analysis," *Proc. of 3rd Int. Conf. on Computing, Communication, Control and Automation*, Pune, Maharashtra, India, 17-18 August 2017.

15. V. Khatavkar, and P. Kulkarni, “Trends in document analysis,” in *Int. Conf. on Data Management, Analytics & Innovation*, Springer, January 2018.
16. H. Ahmad, A. Daud, L. Wang, H. Hong, H. Dawoods, and Y. Yang, “Prediction of rising stars in the game of cricket,” *IEEE Access*, vol. 5, pp. 4104-4124, 2017.