An update to this article is included at the end

Decision Analytics Journal 2 (2022) 100025

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A new in-form and role-based Deep Player Performance Index for player evaluation in T20 Cricket

C. Deep Prakash *, Sanjay Verma

Information Systems Department, Indian Institute of Management Ahmedabad, Vastrapur, Ahmedabad 380015, India

ARTICLE INFO

Keywords:
Sports analytics
Player performance evaluation
Data mining
Machine learning
K-Means clustering
Random Forest algorithm

ABSTRACT

Sports analytics has benefited immensely from the growth and popularity of Machine Learning algorithms. Machine Learning and Data Mining advances have enabled sports analysts to evaluate a player's performance more effectively. A review of existing literature on player performance evaluation methods shows the need to develop a new performance evaluation index for Twenty20 (T20) Cricket. We propose a Deep Player Performance Index (DPPI) to evaluate a T20 Cricket player based on batting and bowling strengths. DPPI captures a player's current form and role in the team. DPPI serves a dual purpose. First, it enables sports fans and researchers to compare players playing a similar role in different teams. Second, the aggregated DPPI values of players playing at different positions in a team give the approximate team strength. To build DPPI, we first modify the existing Fédération Internationale de Football Association (FIFA) player performance evaluation guidelines. We then use the modified guidelines in the context of T20 Cricket. We propose DPPI based on K-Means clustering and Random Forest algorithm and compare our results with the existing player performance evaluation indexes for the Indian Premier League (IPL) 2019 season. Our empirical results show that DPPI captures a player's batting and bowling strength better than other indexes. Thus, DPPI serves as a helpful index for fantasy Cricket users, Cricket fans, coaches, and managers to gain better insights into a player's performance.

1. Introduction

The sports industry is a multi-billion-dollar industry [1]. In today's sporting world massive battery of coaches, analysts, physios, dietitians and trainers assist the players and teams worldwide [2]. The emergence and popularity of fantasy sports, especially the analysis behind these fantasy sports leagues, attracts many sports fans [3]. Therefore, sports analysts are of great assistance to the players, coaches and are also sought after by the media and sports fans to provide more meaningful and in-depth analysis. The growth and popularity of Machine Learning algorithms provide a considerable boost to Sports Analytics [4]. The technological advancements in the Internet of Things (IoT) enable sports analysts to capture data more rapidly and accurately [5]. As a result, in every sport, analysts use Machine Learning algorithms to perform various tasks ranging from player performance evaluation, team selection to maximizing the chances of their team's win [6–8].

In India, Cricket is the most passionately followed sport [9]. One can read the rules and laws of Cricket in detail [10]. In this paper, only a brief description of the game is provided. Cricket is eleven players a side game in which the team scoring more runs wins the match. International Cricket takes place in three different formats: Test Cricket, One Day Internationals (ODI) and Twenty20 (T20) Cricket [11]. A game comprises at least one innings where each team takes turns in batting and bowling/fielding. The fielding team uses a bowler to get a batter out, whereas the batter tries to score as many runs as possible before getting out. An innings of Cricket is further broken down into several overs where each over consists of six legitimate deliveries bowled from one end of a cricket pitch to the batter playing at the other end. In Test Cricket, both the teams get a maximum of two innings and play for a maximum of five days, where ninety overs are bowled each day. In ODI format, both the teams get a maximum of fifty overs to bat. On the other hand, in T20 Cricket, both the teams get a maximum of twenty overs to bat. An innings ends when either the batting team loses ten wickets or plays for the maximum number of overs, or the captain of the team declares the innings and asks the opposition to bat. Thus, T20 Cricket is the shortest and most followed format by Cricket fans who prefer rapid action and thrills [12].

Interestingly, Cricket's shortest format is also the most data-driven format [13–15]. Therefore, in this work, we focus on the T20 format of the game because of its popularity and the advent of many T20 Cricket leagues worldwide where data-based decisions occur the most [16]. Additionally, India's estimated 150 billion dollars fantasy Cricket market transforms Indian cricket fans from mere spectators to one of the main

* Corresponding author.
E-mail addresses: phd18cdeepp@iima.ac.in (C. Deep Prakash), sverma@iima.ac.in (S. Verma).

https://doi.org/10.1016/j.dajour.2022.100025
Received 14 October 2021; Received in revised form 25 December 2021; Accepted 11 January 2022
Available online 25 January 2022
2772-6622/

stakeholders of the game [17]. As a result, the vast growth in fantasy cricket and sports industry investments in India [18] makes it necessary for sports analysts to develop better player performance evaluation indexes for fantasy Cricket users, fans, coaches, and managers.

Researchers often study match outcome prediction and player performance evaluation as two separate research problems [19] because player performance evaluation metrics may not yield better accuracy when directly used to predict match outcomes. The argument supporting this statement is that the way analysts and researchers look at a player's rating in a team sport is remarkably like rating an individual in an individual sport. However, while building match outcome prediction models, the Key Performance Indicators (KPIs) must be taken at the team level and not individually.

It is important to realize that Cricket is a team game that involves highly individual duels between a particular bowler and a batter at any given time during the match. Each bowler and batter have well-designated roles to play in the team and how well they fulfill those expectations go a long way in deciding the outcome of the match. Thus, any index that can capture this information precisely with corresponding downstream analysis can assist in performance prediction at the team level. The challenge is to do so effectively. This challenge motivated us to develop a player performance evaluation index for T20 Cricket called the 'Deep Player Performance Index' (DPPI) that, instead of evaluating a player based on position in the team, evaluates a player according to current form and role in the team. DPPI is based on Machine Learning algorithms and serves a dual purpose. First, instead of comparing players playing at similar positions, it compares T20 players playing a similar role in different teams. Second, for predictive modeling, DPPI values of these T20 players playing different roles in their team can predict the match outcome with a suitably trained Machine Learning model. Our empirical analysis shows that capturing players' role is a better indicator of their KPIs than their positions. Thus, this paper aims to achieve the following research objective:

Research Objective: To develop an in-form and role-based performance evaluation index for T20 Cricket players.

To build DPPI, we study the existing widely accepted FIFA player performance evaluation guidelines [19] and suitably modify them for our purpose. We selected the Indian Premier League (IPL), the most popular T20 Cricket league in India and considered the grand gala jamboree of Cricket [20] for our empirical analysis. Based on K-Means Clustering [21] and Random Forest algorithm [22], we propose a novel player performance evaluation index called DPPI. However, by selecting the suitable sport-specific Key Performance Indicators (KPIs), our proposed idea of DPPI can be extended to other sports as well. We compare our results with the existing methodologies based on the IPL 2019 season statistics. Our empirical results show that DPPI captures a T20 player's performance better than other existing indexes. Thus, DPPI serves as a helpful index for fantasy Cricket users, Cricket fans, coaches and managers to gain better insights into a player's performance.

The organization of the rest of the paper is as follows. Section 2 provides a literature review on player performance evaluation indexes. Section 3 describes our proposed methodology in detail. Section 4 discusses our results and Section 5 presents theoretical implications, practical implications, limitations and future extensions of our work. Finally, Section 6 concludes our work.

2. Literature review

2.1. Use of machine learning algorithms in sports analytics

From knowing one's body [31] to the usage of medicine [32], from player performance analysis [33] to match result prediction [34], from injury prediction [35] to fitness tracking [36], from intelligent sports commentary [37] to social media content generation strategy [38] and from team selection [39] to fantasy sports [40], Machine Learning algorithms are being used in every domain of sports. Especially in Cricket, researchers have primarily used Machine Learning algorithms to predict the match outcome [41–43] and evaluate players [9,30].

As far as supervised Machine Learning algorithms are concerned, researchers have used the Random Forest algorithm in evaluating Baseball [44], Cricket [9], Football [45] and Basketball players [46]. Further, the Random Forest algorithm has been used in predicting the result of a Football match [47], Major League Baseball season [48] and Cricket match [49]. However, we use the Random Forest algorithm to learn feature importance in this work.

Researchers have used an unsupervised Machine Learning algorithm, K-Means clustering, to create training groups for American Football students [50], create team profiles in the Australian football league using performance indicators [51], create mental toughness profiles in cricketers [52] and select the best possible team for a Cricket match [53]. However, in this work, we use the K-Means clustering algorithm to find the roles of batters and bowlers based on KPIs.

2.2. Player performance evaluation methods

Every team management aims to evaluate and find the players who can bring the right combination into the squad and maximize the likelihood of their team's win [54]. Therefore, Data Mining techniques play an essential role in sports for player performance evaluation. For example, a Machine Learning-based Deep Performance Index (DPI) [9] uses different KPIs and evaluates Cricket players for the IPL. A relative performance ranking index uses Data Envelopment Analysis and evaluates Baseball players [55]. In Basketball, a Data Science-based approach evaluates the impact of an injury on a player's performance [56]. In Soccer, the F-MARC test battery [57] assesses the relative importance of playing conditions, technical performance and tactical performance. However, for assessing a player's performance in Soccer, the EA Sports Fédération Internationale de Football Association (FIFA) player performance index [19] is the most used methodology in English Premier League (EPL) and English Championship matches. Further, the player performance evaluation guidelines [19] act as a tool for researchers to develop new indexes. FIFA player performance evaluation guidelines are as follows:

(1) The proposed model should be a statistical model.
(2) The proposed methodology should be able to compare players across different roles.
(3) The number of goals scored by a player should be a key metric in the final model.
(4) Points should be awarded to players of a team whose team wins a match with a clean sheet.
(5) Points should be assigned to players if they assist in goal-scoring.
(6) A trade-off must be made between the simplicity and complexity of the model.
(7) The model should be explainable.

Nevertheless, there are some limitations to the FIFA player performance index [19]. First, the importance given to the number of goals and assists provides more advantage to forward players and midfielders than defenders and goalkeepers. For example, in the EA Sports FIFA 19 ratings, Lionel Messi has been awarded 94, followed by Cristiano Ronaldo (Real Madrid) with 94 and Neymar Jr (Paris Saint-Germain) with a rating of 92. If we see the positions of these players, then Messi, Ronaldo and Neymar Jr are all forward position players (Right Forward, Striker and Left Wing, respectively). In Soccer, it is necessary to keep the goals scored in the performance evaluation guidelines because a team aims to score goals and win the game. Similarly, in T20 Cricket, the number of runs scored and wickets taken are essential KPIs. In addition to that, the runs scoring capability (how fast runs are scored, how many balls did the batter face, how many boundaries did the batter score wherein the ball crossed the perimeter of the playing field that generally yields four or six runs depending on whether the ball is hit to or beyond the perimeter) and wickets taking

Table 1
Summary of batting performance evaluation indexes.

Method	Batting performance indexes
1	(Average * Strike Rate)/100 [23]
2	[(Player Average/Tournament Average) + (Player Strike Rate/Tournament Strike Rate)] * Runs [24]
3	KPIs: Runs, Average, Strike Rate, Fours, Six, HF = (2*Century) + half-century Methodology: • Use correlation matrix and apply Principal Component Analysis. • Take the weights of the first principal component. • Multiply the weights with the normalized KPI values. • Get the aggregated batting strength score. [25]
4	KPIs: (1) Hard Hitter = (4*Fours + 6*Sixes)/Balls faced by player (2) Finisher = Number of times Not Out/Total number of played innings (3) FastScorer = Strike Rate (4) Consistent = Average (5) RunningBetweenWickets (RBW) = (Runs −(4*Fours + 6*Sixes))/Number of balls faced without boundary Methodology: • Calculate MVPI Index [26]. • Apply Random Forest Algorithm (take MVPI Index as target vector) • Calculate feature importance for each KPI. • Multiply the calculated feature importance values with normalized KPI values. • Get normalized index values between 0 and 1 for each player. [9]
5	Mike Hussey Number for Batting = Average + Strike Rate [27]
6	[(Player Average/Tournament Average) + (Player Strike Rate/Tournament Strike Rate)^2] * Runs [26]
7	Average * Strike Rate [28]
8	Average^{1}-α * Strike Rate$^{\alpha}$ where $0 \leq \alpha \leq 1$ ($\alpha = 0.75$) [29]
9	KPIs: (1) Hard Hitter (2) Finisher (3) FastScorer (4) Consistent (5) RunningBetweenWickets (RBW) Methodology: • Calculate MVPI Index [26]. • Divide the batsman into four categories: Openers, Middle Orders, Finishers And Bowlers. • Apply Random Forest algorithm for each category (Take MVPI Index as target vector) • Calculate feature importance for each KPI. • Multiply feature importance with normalized KPI values. • Get normalized index between 0 and 1. [30]

capability (how many wickets did the bowler take, how many runs did the bowler concede, how many balls did the bowler bowl) also matter depending on the game situation. Second, the FIFA Player Performance Index aims to develop a rating index that can work across different roles, not for a specific role. However, in T20 Cricket, players play according to a strategy and definite role. A player's current form (recent performances) and historic statistics both capture different skillsets of a player. Therefore, we present a player's in-form and role-based performance evaluation index called DPPI. We modify and use the modified player performance evaluation guidelines [19]. The modified guidelines for in-form and role-based player performance evaluation in T20 Cricket are given below:

(1) The proposed model should be a statistical model.
(2) The model should evaluate each player according to their current form and role in the team.
(3) The number of runs scored and wickets taken by the player should be a part of the performance evaluation index. However, the performance index should also capture the player's run-scoring ability and wicket-taking ability according to the role in the team.
(4) A trade-off must be made between the simplicity and complexity of the model.
(5) The model should be explainable.

We use the above guidelines to critically analyze the current literature on player performance evaluation in Cricket. Tables 1 and 2 summarize the player's batting and bowling performance evaluation indexes, respectively.

Table 1 shows that the most used performance evaluation KPIs to assess a player's batting performance are average and strike rate. However, these two KPIs are not enough to evaluate a player's batting ability [9,30]. Similarly, Table 2 shows that the most used performance evaluation KPIs are bowling average, strike rate and economy rate for bowlers. Nevertheless, these three KPIs are not enough to evaluate a player's bowling ability [9,30]. Therefore, from the literature review, we deduce that the traditional methods of averages and strike rates do not work well in evaluating a player's true potential. As a result, one needs to consider additional KPIs to capture a player's batting and bowling ability.

We evaluate the existing player performance evaluation indexes based on the modified guidelines to see whether they follow them. We consider a model to be a statistical model if it follows the statistical

Table 2
Summary of bowling performance evaluation indexes.

Method	Bowling performance indexes
1	(Average * economy rate)/100 [23]
2	[(Tournament average/player average) + (Tournament economy/player economy)] * Wickets [58]
3	KPIs: Wickets, bowling average, strike rate, economy Methodology: • Use correlation matrix and apply Principal Component Analysis. • Take the weights of the first principal component. • Multiply the weights with the normalized KPI values. • Get the aggregated batting strength score. [25]
4	KPIs: (1) Economy (2) Wicket Taker = Strike Rate (Balls bowled/Wickets Taken) (3) Consistent = Average (4) BigWicketTaker = Number of times four wickets or five wickets taken/Number of innings played (5) ShortPerformance = (Number of wickets taken – 4* Number of times four wickets – 5* Number of times five wickets taken)/(Number of innings played –Number of times four wickets or five wickets taken) Methodology: • Calculate MVPI Index. • Apply Random Forest algorithm (Take MVPI Index as target vector) • Calculate feature importance for each KPI. • Multiply feature importance with normalized KPI values. • Get normalized ranks between 0 and 1 [9]
5	[(Tournament average/player average) + (Tournament economy/player economy)^2] * Wickets [26]
6	Strike Rate * Average [28]
7	Strike Rate^{1}-α * Average$^{\alpha}$ where $0 \leq \alpha \leq 1$ ($\alpha = 0.75$) [59]
8	Bowling Index (CBR) = 3* Runs Conceded/(Wickets Taken + Overs Bowled + Strike Rate*Runs Conceded) [60]
9	KPIs: (1) Economy (2) Wicket Taker (3) Consistent (4) BigWicketTaker (5) ShortPerformance Methodology: • Calculate MVPI Index [26]. • Divide the bowlers into four categories: Pacers, Spinners, Pace Allrounder and Spin Allrounder. • Apply Random Forest algorithm for each category (Take MVPI Index as target vector) • Calculate feature importance for each KPI. • Multiply feature importance with normalized KPI values. • Get normalized index between 0 and 1. [30]

modeling assumptions during our evaluation. We consider a model to provide role-based evaluations if it identifies different player roles and evaluates the players accordingly. We consider a model to capture the run-scoring ability if it considers KPIs other than batting average and strike rate. Similarly, we consider a model to capture the wickets taking ability for bowlers if the model captures KPIs other than average, strike rate and economy. Further, we determine the model's simplicity and complexity based on the model's linearity or nonlinearity. The explainability of the model is determined by evaluating if it can explain why some KPIs are more important than others. Lastly, form consideration is evaluated by checking if the methodology treats recent performances data differently. Tables 3 and 4 show whether the batting and bowling indexes follow the modified player performance evaluation guidelines in T20 Cricket.

A literature review of existing player performance indexes (Tables 3 and 4) in Cricket highlights that existing performance evaluation indexes do not satisfy all the modified player performance evaluation guidelines. Therefore, in this work, we propose additional KPIs and show how to use these KPIs to evaluate a player's batting and bowling ability. Additionally, unlike the existing methods, in DPPI, we also consider a player's current form. Thus, this paper identifies the gap mentioned earlier and addresses it through the proposed DPPI.

3. Methodology

Fig. 1 shows our proposed methodology. We consider the IPL 2019 season to evaluate our methodology because, unlike IPL 2020 and IPL 2021, IPL 2019 happened in India just like the earlier four IPL seasons. As highlighted earlier that the proposed DPPI is an in-form and role-based player performance evaluation index. Therefore, we extract a player's T20 international career's batting and bowling statistics until the start of the IPL 2019 season (i.e., 22 March 2019) to consider a player's international experience. We then extract a player's IPL career's batting and bowling statistics for the last four seasons (i.e., IPL 2015, IPL 2016, IPL 2017, IPL 2018) and consider a player's IPL experience. For those Indian players who did not play any IPL or international match, we extract the batting and bowling statistics of their last Syed Mushtaq Ali Tournament (i.e., SMAT 2018), the domestic T20 league of India. We do not consider the domestic statistics of international players playing in IPL for the first time and have no international

Table 3
Evaluation summary of existing batting performance evaluation indexes.

Method	Statistical model	Role-based evaluation	Run scoring ability	Simplicity/ complexity	Explainability	Form consideration
1	Yes	No	No	Simple	Yes	No
2	Yes	No	No	Simple	Yes	No
3	Yes	No	Yes	Complex	No	No
4	Yes	No	Yes	Complex	Yes	No
5	Yes	No	No	Simple	Yes	No
6	Yes	No	Yes	Simple	Yes	No
7	Yes	No	No	Simple	Yes	No
8	Yes	No	No	Simple	Yes	No
9	Yes	Yes	Yes	Complex	Yes	No

Table 4
Evaluation summary of existing bowling performance evaluation metrics.

Method	Statistical model	Role-based evaluation	Wickets taking ability	Simplicity/ complexity	Explainability	Form consideration
1	Yes	No	No	Simple	Yes	No
2	Yes	No	No	Simple	Yes	No
3	Yes	No	Yes	Complex	No	No
4	Yes	No	Yes	Complex	Yes	No
5	Yes	No	Yes	Simple	Yes	No
6	Yes	No	No	Simple	Yes	No
7	Yes	No	No	Simple	Yes	No
8	Yes	No	Yes	Simple	Yes	No
9	Yes	Yes	Yes	Complex	Yes	No

Table 5
Batting and bowling KPIs used in DPPI.

Batting KPIs	Bowling KPIs
Average (BA) = Runs scored per dismissal	Average (BA) = Runs conceded per wicket
Strike Rate (SR) = Runs scored per hundred balls	Strike Rate (SR) = Number of balls bowled per wicket
Balls Faced Index (BFI) = Balls faced per innings	Balls Bowled Index (BBI) = Balls bowled per innings
Running Between the Wickets Index (RBWI) = (Number of runs scored without a boundary)/(Number of balls played without a boundary)	Economy Rate (ECON) = Runs conceded per over
Boundary Index (BI) = (Number of boundaries)/(Number of innings)	Big Wicket Index (BWI) = (Number of times four or more than four wickets taken in an innings)/(Number of innings in which bowled)
Big Innings Index (BII) = (Number of fifties + hundreds)/(Number of innings)	Short performance Index (SPI) = (Number of wickets without considering the big wickets)/(Number of innings without considering the big wicket innings)
Finishing Index (FI) = (Number of not out innings)/(Number of innings)	

experience. The reason is that for Indian players, the conditions remain similar in IPL, whereas for the international players, the game conditions change entirely.

While evaluating a player, we highlight that it is necessary to consider international statistics and domestic statistics separately because some players play for their respective national teams and bring that international experience into a franchise-driven league such as IPL. On the other hand, some players are yet to make their international debut. Therefore, finding the right combination of domestic and international statistics through different dataset importance values becomes necessary to capture a player's international and domestic experience. The subsequent subsections describe the details of each step.

3.1. Defining the Key Performance Indicators (KPIs)

We define our KPIs and calculate the KPIs for all the datasets mentioned above (domestic statistics of last four IPL seasons, last SMAT season and international T20 career statistics). Using the KPI values for each dataset, we identify different roles for each player. Table 5 describes the batting and bowling KPIs we use in this work. As modified guideline 4 suggests, we need to make a trade-off between the simplicity and complexity of the model. Therefore, instead of defining KPIs for each role separately, we define these KPIs irrespective of the players' roles. First, we convert the KPIs to capture a player's form. We then use a Data Mining approach to convert these KPIs to role-based KPIs by assigning different feature importance values depending on a player's cluster-based batting and bowling roles.

3.2. Capturing a player's current form using different dataset importance values

We consider the overall international T20 career statistics for every player until the start of the IPL season to capture the international experience. However, since we also want to capture a player's current form, we assign different importance values (heuristic-based) to each KPI extracted from different domestic season statistics. Table 6 shows the importance given to each IPL season to capture a player's current form. The rationale behind these different weights is that a player's recent performances are more important than past performances. Therefore, depending on whether a player has played all four or three or two or one out of the last four IPL seasons, we assign different importance values to IPL season statistics.

3.3. Identifying player roles using the K-means clustering algorithm

Cricket is eleven players a side game and every team member has a specific role [9]. Role-based categorization of players and performance evaluation of players is not new in the Cricket literature [30]. The category-based Deep Performance Index (CDPI) method categorizes the batters into openers, middle order players, finishers and bowlers. Similarly, CDPI [30] categorizes the bowlers into Spinners, Pacers, Spin Allrounders and Pace Allrounders. CDPI assumes that all those players who play for different teams but play at the same batting positions have a similar role in the team. For example, the feature importance values for Rohit Sharma (batter from India) and Jos Buttler (batter from England) are the same because they both are Opening batters. Thus,

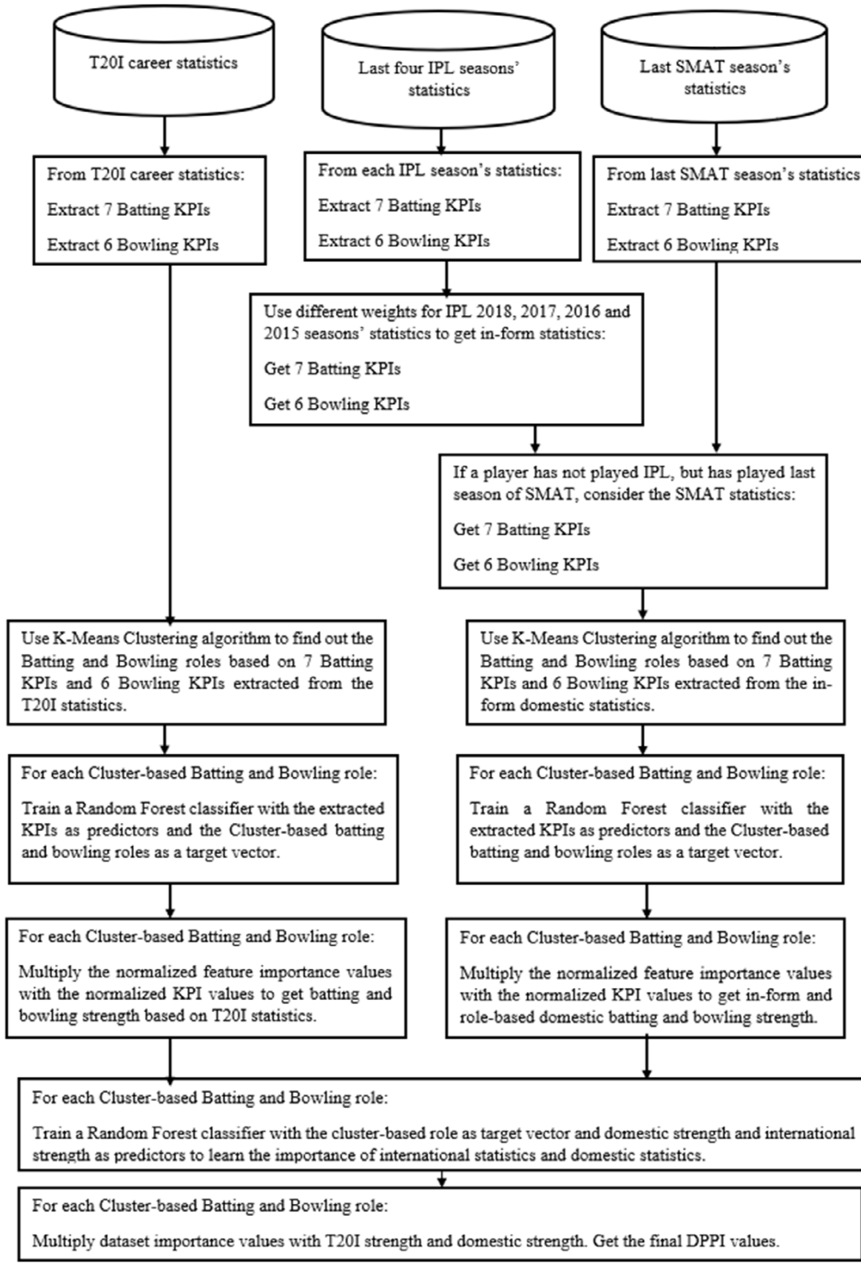

Fig. 1. The proposed methodology for DPPI.

CDPI restricts a player as an Opener or a Middle Order player, or as a Finisher based on their batting positions. However, roles are far more than just the batting position in Cricket. Another limitation of CDPI is that it does not capture the player's current form because it treats all games equally. Unlike CDPI, which only looks at the player's position in the team, we capture a player's current form and role in the team.

First, we normalize the values for each KPI. Higher values are preferred for some KPIs, and for some, lower values are preferred. Therefore, we use different approaches to normalize the KPI values. After normalization, for all the KPIs, higher values are preferred. Eqs. (1) and (2) show our normalization method.

Case 1, if higher values are preferred for the KPI 'i' for a player 'j'

$$KPI_{ij} = (KPI_{ij} - Min(KPI_i))/(Max(KPI_i) - Min(KPI_i)) \quad (1)$$

Case 2, if lower values are preferred for the KPI 'i' for a player 'j'

$$KPI_{ij} = (Max(KPI_i) - KPI_{ij})/(Max(KPI_i) - Min(KPI_i)) \quad (2)$$

After normalization, we use Elbow Method [61] to find the number of clusters (K) based on international batting, international bowling, domestic batting and domestic bowling statistics. The Elbow Method is a heuristic in which the explained variation in the data is plotted as a function of the number of clusters (K) and the elbow of the curve is picked as the optimal value of K. Thus, based on the Elbow Method heuristic for all four datasets we arrive at K = 5. Afterward, based on the extracted KPIs for different datasets and the value of K (number of clusters), we use the K-Means Clustering algorithm [21] to categorize the players into different clusters. We select the K-Means clustering algorithm in our work because our modified guidelines suggest that we need to make a trade-off between the simplicity and complexity of the model and the model should also be explainable. K-Means Clustering algorithm satisfies both guidelines. The K-Means clustering algorithm is order-independent which means that it generates the same partition of data irrespective of the way the observations are presented. Thus, it becomes easy to explain the clustering results with the help of cluster means [62]. Second, the algorithm's time complexity is O(n*k*i), where

Table 6
Capturing the form of a player based on different dataset importance values.

Case	Case detail (2018–2015 IPL seasons)	IPL season weights
1	If a player has played all four last IPL seasons	Average Case (x) = 0.25 weight for each season However, for capturing form: • IPL 2018 (x + 3 + 4) = 0.32 • IPL 2017 (x + 3) = 0.28 • IPL 2016 (x − 3) = 0.22 • IPL 2015 (x − 3 − 4) = 0.18
2	If a player has played only three out of the last four IPL seasons	Average Case (x) = 0.33 weight for each season However, for capturing the form: • Last Season (x + 4) = 0.37 • Second Last Season (x) = 0.33 • Third Last Season (x − 3) = 0.30
3	If a player has played only two out of the last four IPL seasons	Average Case (x) = 0.5 weight for each season However, for capturing the form: • Last Season (x + 3) = 0.53 • Second Last Season (x − 3) = 0.47
4	If a player has played only one out of the last four IPL seasons	Average Case (x) = 1 weight for the season However, for capturing the form: • Last Season (x) = 1

'n' is the number of observations, 'k' is the number of clusters and 'i' is the number of iterations taken by the algorithm converge [62]. Therefore, the K-Means Clustering algorithm allows us to make a trade-off between the simplicity and complexity of the model for smaller values of 'k'. Based on the cluster means, we identify each player's batting and bowling roles in the team. We also compare the cluster-based DPPI approach with the position-based player categorization approach [30] to find which method best works. For cluster-based batting role identification, we select a player's international and domestic statistics only if the player has faced a minimum of 25 deliveries. The reason for selecting a minimum cut-off value is to filter our batters from those who can influence the overall KPIs with just one inning. For example, a bowler Shardul Thakur played one innings in 2018 in which he scored fifteen runs in five balls and remained not out. This inning results in a strike rate value of 300. The sample of 5 balls is not enough to declare Shardul Thakur as the best hitter. One can also use different minimum-ball-faced criteria.

Tables 7 and 8 show different batting clusters based on KPIs extracted from international and domestic statistics, respectively. From Tables 7 and 8, one can see that the batters are categorized into Specialist Batters (SB), Specialist Finishers (SF), Less Contribution (LC), Handy with Bat (HWB) and Floaters (F). Cluster mean values for international and domestic stats-based batter clusters differ because of the differences in average scores in international and domestic games. However, the roles of batters remain the same. Specialist Batters (SB) are those for whom, apart from the FI, all six KPI values are better than other batters. Specialist Finishers (SF) have better values of FI than Specialist Batters (SB) but have average values of the remaining six KPIs. Floaters (F) do not have FI values as good as Specialist Finishers (SF) and do not have the remaining six KPI values as good as Specialist Batters (SB). They are called Floaters (F) because they do the job of a Specialist Batter (SB) sometimes and can also do the job of a Specialist Finisher (SF) if required. Handy with Bat (HWB) are those batters who can remain not out and contribute something with the bat based on their low values on other KPIs. Less Contribution (LC) category players have the least values for BA and SR.

Tables 9 and 10 show different clusters (bowling roles) based on KPIs extracted from international and domestic performance statistics. We only select a player's international and domestic statistics for cluster-based bowling role identification if the player has bowled a minimum of 60 deliveries (10 overs).

From Tables 9 and 10, one can see that the bowlers are categorized into Specialist Bowlers (SB), Short Performance Bowlers (SPB), Impact Bowlers (IB), Handy with Ball (HWB) and Less Contribution (LC). Specialist Bowlers (SB) are better than the rest based on all KPIs. Short Performance Bowlers (SPB) are those who do not guarantee as many overs as Specialist Bowlers (SB), but they can take wickets in the overs that they bowl (High SPI). Impact Bowlers (IB) do not bowl as many overs as Specialist Bowlers (SB) or Short Performance Bowlers (SPB), but they focus on maintaining their economy when they bowl. Handy with Ball (HWB) bowlers are those who bowl occasionally. Less Contribution (LC) bowlers do not take wickets as expected. Thus, the K-Means Clustering algorithm allows us to find players who have a similar role for their team.

3.4. *Calculating feature importance using random forest algorithm*

We borrow most of our KPIs from the CDPI Index [30] in the proposed work, but the methodology adopted for computing DPPI is very different. CDPI Index uses the "MVPI Index" [26] as the target variable for training the Random Forest [22]. Since MVPI Index is a continuous variable, the Random Forest algorithm works as an explanatory model in the CDPI Index. However, we use the Random Forest algorithm as a classifier with K-Means Clustering results as a target vector in this work. We use a binary variable (1/0) in the target vector for each batting or bowling cluster-based role to represent whether a player belongs to that cluster-based role and then learn the feature importance [63]. We follow a similar approach for learning the feature importance for bowlers.

We use the Random Forest algorithm because it typically yields better predictive results than linear modeling approaches. We can explain the feature importance through the decrease in the mean Gini Index [64]. Our overall approach can be summarized as follows. First, we take advantage of the unsupervised learning algorithm K-Means Clustering, which does not depend on a predefined set of classes and groups players together. Then we use the supervised classification algorithm Random Forest which uses a set of labeled players and learns the model to predict the unclassified players into different clusters that are role-based. This combination of unsupervised and supervised learning [65] allows us to accurately determine the relative importance of different features for each role. It makes our methodology better than the existing player performance evaluation methods in T20 Cricket.

The feature importance learning methodology for batters is as follows. Let there be total 'm' batters in our dataset. Such that, there are 'a' batters who belong to Cluster 1, 'b' batters who belong to Cluster 2, 'c' batters who belong to Cluster 3, 'd' batters who belong to Cluster 4 and 'e' = m – (a + b + c + d) who belong to Cluster 5. The goal of Random Forest classifier training is to find a function 'g' for each cluster 'a', 'b', 'c', 'd', 'e' based on the KPIs: (BA, SR, BFI, RBWI, BI, BII, FI) such that given a feature set $(BA, SR, BFI, RBWI, BI, BII, FI\})_i$ for a batsman 'i' whose cluster is yet to be determined, the function 'g' can tell whether batter 'i' belongs to 'a' or 'b' or 'c' or 'd' or 'e'. Since we use a 'one vs.

Table 7
International T20 career data-based batting clusters, cluster means and example players.

Cluster (Number of batters)	Batting Role (Example player)	BA	SR	BFI	BI	RBWI	BII	FI
1 (8)	Specialist batters (KL Rahul)	0.702 0.860	0.780 0.830	0.800 0.910	0.872 1.000	0.455 0.590	0.815 0.910	0.214 0.270
2 (19)	Floaters (Steven Smith)	0.449 0.490	0.593 0.590	0.520 0.550	0.465 0.480	0.601 0.750	0.280 0.440	0.323 0.390
3 (7)	Less contribution (Joe Denly)	0.206 0.070	0.327 0.110	0.434 0.300	0.388 0.230	0.177 0.140	0 0	0.057 0
4 (29)	Specialist finishers (Andre Russell)	0.216 0.270	0.530 0.730	0.187 0.220	0.231 0.330	0.462 0.300	0.030 0	0.555 0.520
5 (10)	Handy with bat (Nicholas Pooran)	0.471 0.410	0.641 0.730	0.642 0.470	0.684 0.540	0.362 0.410	0.440 0.560	0.142 0.280

Table 8
Domestic (IPL + SMAT) performance data-based batting clusters, cluster means and example players.

Cluster (Number of batters)	Batting Role (Example player)	BA	SR	BFI	RBWI	BI	BII	FI
1 (33)	Floaters (Steven Smith)	0.437 0.580	0.596 0.590	0.573 0.750	0.726 0.870	0.508 0.580	0.357 0.370	0.160 0.200
2 (34)	Less contribution (Jasprit Bumrah)	0.168 0.140	0.453 0.370	0.133 0.090	0.699 0.720	0.134 0.070	0.014 0	0.567 0.790
3 (23)	Handy with bat (Marcus Stoinis)	0.203 0.320	0.462 0.450	0.334 0.380	0.621 0.790	0.266 0.250	0.056 0.130	0.120 0.310
4 (9)	Specialist finishers (MS Dhoni)	0.615 0.720	0.747 0.570	0.353 0.560	0.829 0.750	0.377 0.430	0.112 0.240	0.644 0.580
5 (6)	Specialist batters (KL Rahul)	0.688 0.670	0.641 0.630	0.815 0.740	0.750 0.760	0.767 0.700	0.755 0.590	0.230 0.350

Table 9
International T20 career data-based bowling clusters, cluster means and example players.

Cluster (Number of bowlers)	Bowling Role (Example player)	BA	SR	ECON	BBI	BWI	SPI
1 (11)	Specialist Bowlers (Rashid Khan)	0.904 0.980	0.897 0.950	0.631 0.900	0.894 0.920	0.349 0.520	0.626 0.790
2 (11)	Handy with ball (Suresh Raina)	0.697 0.770	0.666 0.690	0.556 0.680	0.304 0.330	0 0	0.222 0.200
3 (27)	Short performance (Bhuvneshwar Kumar)	0.881 0.850	0.837 0.770	0.713 0.770	0.852 0.850	0.075 0.120	0.557 0.400
4 (3)	Impact Bowlers (Lungi Ngidi)	0.970 0.980	0.957 0.960	0.777 0.850	0.520 0.580	0.747 0.560	0.430 0.560
5 (18)	Less contribution (Colin de Grandhomme)	0.702 0.720	0.685 0.700	0.485 0.490	0.712 0.530	0.028 0	0.325 0.280

Table 10
Domestic (IPL + SMAT) performance data-based bowling clusters, cluster means and example players.

Cluster (Number of bowlers)	Bowling Role (Example player)	BA	SR	ECON	BBI	BWI	SPI
1 (8)	Handy with ball (Suresh Raina)	0.346 0.310	0.344 0.160	0.582 0.690	0.561 0.110	0 0	0.237 0.010
2 (35)	Short performance (Bhuvneshwar Kumar)	0.864 0.880	0.866 0.860	0.687 0.790	0.875 0.940	0.094 0.160	0.661 0.700
3 (8)	Less contribution (Colin de Grandhomme)	0.740 0.620	0.656 0.580	0.550 0.640	0.105 0.320	0 0	0.081 0.140
4 (3)	Specialist Bowlers (Andrew Tye)	0.953 0.990	0.963 1.000	0.773 0.800	0.953 0.940	0.827 1.000	0.673 0.740
5 (33)	Impact Bowlers (Ravindra Jadeja)	0.736 0.700	0.724 0.660	0.632 0.710	0.731 0.660	0.032 0.050	0.374 0.300

all' classification method [66] to learn the function 'g', we learn five different functions g_1, g_2, g_3, g_4, g_5, corresponding to each cluster-based batting role. To summarize, in the conceptualization of g_1, g_2, g_3, g_4, g_5, we learn the feature importance of BA, SR, BFI, RBWI, BI, BII, FI in classifying a batter 'i' into the respective cluster that is role-based.

We measure the feature importance through Gini Index [67]. The mean decreases in Gini Index while constructing the Random Forest and splitting the dataset from one level to another level based on a feature value $f_j <= t$ where f ∈ {BA, SR, BFI, RBWI, BI, BII, FI}, 'j' is the node split level and 't' is the cut-off value of feature 'f', gives us the measure of the feature importance. Node splits across all levels measure how important feature 'f_j' is in classifying the batters into different cluster-based roles. Thus, we get a feature importance vector corresponding to each KPI (BA, SR, BFI, RBWI, BI, BII, FI) in functions g_1, g_2, g_3, g_4 and g_5. The intuition behind this idea is to know how important each feature is in splitting the batters into different partitions so that by splitting a Random Forest node on that feature, we get a purer class. The pseudo-code of the feature importance learning process is given below:

for learning feature importance for a particular ‘g’ (g_1, g_2, g_3, g_4, g_5)
divide data into 80% training and 20% testing
for number of Decision Trees ‘N’ = 500
for number of features randomly selected and considered at each node split ‘k’
(1) Randomly select a subset of players out of the available players from the training dataset
(2) Randomly select **“k”** KPIs from a total **“7”** KPIs for batters (6 KPIs for bowlers).
(3) Among the **“k”** features, calculate the node **“d”** using the best split point
(Highest decrease in Gini Index)
(4) Split the node "d" into two child nodes using the best split criteria
(5) Repeat steps 1-4 until a minimum of 20 observations are available for a node split
Check the classification performance of the Random Forest (N, K) on the test dataset
for each KPI ‘F’ (F_1, F_2, F_3, F_4, F_5, F_6, F_7)

$$\text{Mean Decrease in Gini (F)} = \frac{1}{N} * \left(\sum_{n=1}^{N}(\text{Decrease in Gini (F)n})\right)$$

Calculate the normalized weight of each Mean Decrease in Gini(F) in corresponding g

Finally, let [w_{11}, w_{12}, w_{13}, w_{14}, w_{15}, w_{16}, w_{17}], [w_{21}, w_{22}, w_{23}, w_{24}, w_{25}, w_{26}, w_{27}], [w_{31}, w_{32}, w_{33}, w_{34}, w_{35}, w_{36}, w_{37}], [w_{41}, w_{42}, w_{43}, w_{44}, w_{45}, w_{46}, w_{47}] and [w_{51}, w_{52}, w_{53}, w_{54}, w_{55}, w_{56}, w_{57}] be the normalized feature importance values corresponding to KPIs {BA, SR, BFI, RBWI, BI, BII, FI}, in g_1, g_2, g_3, g_4 and g_5 for Cluster 1, 2, 3, 4 and 5 batters respectively. Then the role-based and in-form DPPI for each Cluster-based batting role is calculated as given in Eq. (3).

$$\text{In-form DPPI for Bat Role}_i(\text{DPPI\_Bat}) = \sum_{j=1}^{7} \text{Wij*fij} \tag{3}$$

Where DPPI_Bat is the In-form and Role-Based Batting DPPI score for player ‘i’. Role $_i$ is the cluster-based batting role of the player ‘i’. W_{ij} is the feature importance for KPI ‘j’ using the corresponding to player i’s batting role. ‘f_{ij}’ is the normalized jth KPI value for player ‘i’, based on the dataset. A similar analysis is done for cluster-based bowling roles to get DPPI Bowl Role$_i$. It is necessary to mention that other factors such as weather conditions, pitch conditions, home or away conditions, food habits, psychological factors, and injury probability can affect a player’s performance. However, DPPI is limited to a player’s on-field performance and does not consider these other factors. In the future, the model can incorporate other factors also.

4. Results

4.1. Case 1: Feature importance for cluster-based role identification approach

Tables 11 and 12 show the Random Forest (RF) classifier training results for batting and bowling roles depending on different clusters and datasets. Tables 11 and 12 also show the normalized feature importance values for each batting and bowling dataset and cluster-based roles for each KPI. For example, for SB (Specialist Batters, Table 11), their BI and BII are higher than other roles. Similarly, for SB (Specialist Bowlers, Table 12), their SPI and BBI matter more than other features. Thus, by using the same KPIs across different roles and finding the feature importance, we evaluate each player based on their role in the team.

4.2. Case 2: Feature importance for position-based role identification approach

Tables 13 and 14 show the Random Forest classifier training and normalized feature importance values for position-based batting roles (Opener (O), Middle Order (MO), Finisher (F) and Bowler (B), respectively) and bowling roles (Pacer (P), Spinner (S), Pace Allrounder (PA), Spin Allrounder (SA)), respectively. For example, for O (Openers, Table 13), their BI and BII are higher than other roles. Similarly, for P (Pacers, Table 14), their SPI and BBI matter more than other features. Thus, by using the same KPIs across different roles and finding the feature importance, we evaluate each player based on their role in the team.

Figs. 2–5 compare Random Forest classifier’s training accuracies between cluster-based and position-based role identification approaches for different datasets.

From Figs. 2–5, it is evident that the Random Forest classifier’s training accuracies are better in the cluster-based role identification

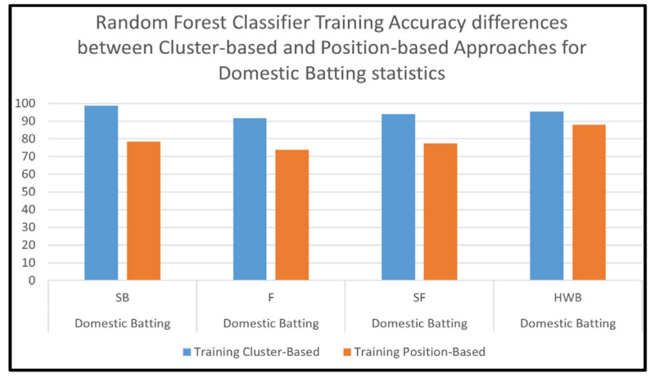

Fig. 2. Comparison of Random Forest classifier’s training accuracies between cluster-based and position-based role identification approaches based on domestic batting statistics.

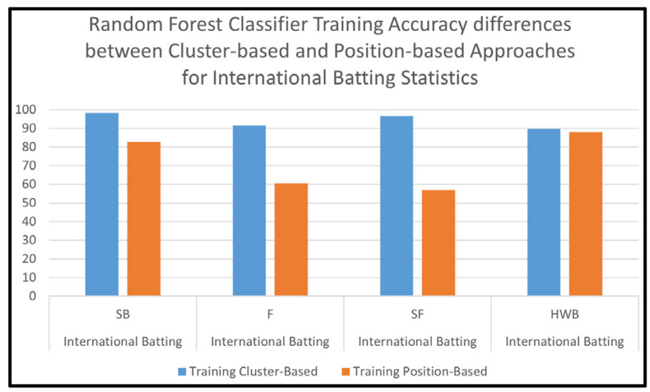

Fig. 3. Comparison of Random Forest classifier’s training accuracies between cluster-based and position-based role identification approaches based on international batting statistics.

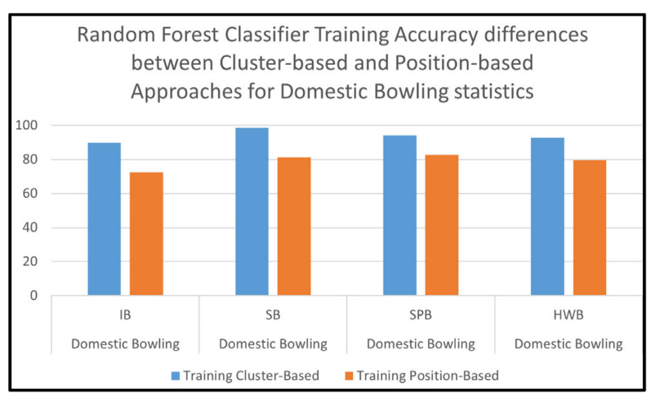

Fig. 4. Comparison of Random Forest classifier’s training accuracies between cluster-based and position-based role identification approaches based on domestic bowling statistics.

approach than the position-based role identification approach. Low training accuracy of the position-based role identification approach indicates that the current model configuration cannot capture the complexity of the data compared to the cluster-based role identification approach. Thus, our empirical results show that DPPI includes additional KPIs and uses a cluster-based role identification approach that performs better than the existing role-based CDPI index. The practical significance of this result is that limiting a player’s role to batting positions or bowling roles without considering the actual KPI values

Table 11
Random Forest classifier training for cluster-based Batting Roles.

Dataset	RF classifier training and feature importance	Parameters/KPIs	Cluster-based batting roles				
			SB	F	LC	SF	HWB
International batting	RF accuracy	Training (%)	98.28	91.38	94.83	96.55	89.66
		Testing (%)	100	93.33	93.33	100	93.33
	Feature importance	AVE	0.20	0.10	0.10	0.12	0.15
		SR	0.06	0.08	0.14	0.03	0.06
		BFI	0.17	0.21	0.11	0.35	0.13
		BI	0.27	0.21	0.10	0.23	0.22
		RBWI	0.01	0.15	0.26	0.03	0.12
		BII	0.26	0.18	0.04	0.06	0.16
		FI	0.03	0.08	0.24	0.18	0.16
Domestic batting	RF accuracy	Training (%)	98.81	91.67	96.43	94.05	95.24
		Testing (%)	100	100	90.48	100	90.48
	Feature importance	AVE	0.13	0.19	0.10	0.20	0.16
		SR	0.03	0.04	0.05	0.18	0.08
		BFI	0.22	0.18	0.20	0.11	0.13
		BI	0.23	0.17	0.25	0.13	0.15
		RBWI	0.02	0.03	0.03	0.09	0.09
		BII	0.35	0.28	0.09	0.06	0.13
		FI	0.03	0.11	0.28	0.22	0.28

Table 12
Random Forest classifier training for cluster-based Bowling Roles.

Dataset	RF classifier training and feature importance	Parameters/KPIs	Cluster-based Bowling Roles				
			SB	HWB	SPB	IB	LC
International bowling	RF accuracy	Training (%)	94.64	98.21	80.36	98.21	87.50
		Testing (%)	100	100	92.86	92.86	85.71
	Feature importance	AVE	0.08	0.13	0.18	0.18	0.28
		SR	0.18	0.14	0.16	0.21	0.18
		ECON	0.08	0.12	0.15	0.04	0.14
		BBI	0.14	0.12	0.17	0.13	0.22
		BWI	0.32	0.07	0.15	0.41	0.03
		SPI	0.19	0.43	0.20	0.03	0.15
Domestic bowling	RF accuracy	Training (%)	98.55	92.75	94.12	89.86	95.65
		Testing (%)	100	100	100	100	100
	Feature importance	AVE	0.20	0.40	0.17	0.21	0.09
		SR	0.22	0.34	0.19	0.18	0.09
		ECON	0.07	0.05	0.07	0.07	0.08
		BBI	0.06	0.12	0.15	0.19	0.42
		BWI	0.42	0	0.07	0.03	0
		SPI	0.03	0.08	0.35	0.31	0.32

Table 13
Random Forest classifier training for position-based Batting Roles.

Dataset	RF classifier training and feature importance	Parameters/KPIs	Position-based batting roles			
			O	MO	F	B
International Batting	RF accuracy	Training (%)	82.76	60.34	56.90	87.93
		Testing (%)	100	73.33	86.67	93.33
	Feature importance	AVE	0.13	0.14	0.14	0.12
		SR	0.09	0.14	0.14	0.11
		BFI	0.17	0.17	0.17	0.19
		BI	0.20	0.17	0.15	0.14
		RBWI	0.10	0.12	0.21	0.10
		BII	0.13	0.10	0.07	0.05
		FI	0.19	0.16	0.13	0.30
Domestic Batting	RF accuracy	Training (%)	78.57	73.81	77.38	88.10
		Testing (%)	85.71	52.38	80.95	80.95
	Feature importance	AVE	0.10	0.11	0.17	0.20
		SR	0.13	0.11	0.13	0.09
		BFI	0.15	0.17	0.19	0.30
		BI	0.22	0.17	0.16	0.17
		RBWI	0.11	0.12	0.12	0.06
		BII	0.13	0.18	0.11	0.10
		FI	0.14	0.13	0.11	0.09

Table 14
Random Forest classifier training for position-based bowling roles.

Dataset	RF classifier training and feature importance	Parameters/KPIs	Position-based Bowling Roles			
			P	S	PA	SA
International bowling	RF accuracy	Training (%)	57.14	83.93	71.43	76.79
		Testing (%)	78.57	100	100	78.57
	Feature importance	AVE	0.13	0.14	0.14	0.17
		SR	0.16	0.11	0.16	0.17
		ECON	0.18	0.32	0.19	0.22
		BBI	0.27	0.18	0.29	0.17
		BWI	0.08	0.08	0.07	0.08
		SPI	0.18	0.17	0.15	0.19
Domestic bowling	RF accuracy	Training (%)	72.46	81.16	82.61	79.71
		Testing (%)	72.22	72.22	88.89	88.89
	Feature importance	AVE	0.13	0.18	0.16	0.15
		SR	0.12	0.13	0.14	0.14
		ECON	0.20	0.33	0.19	0.13
		BBI	0.29	0.16	0.26	0.31
		BWI	0.07	0.08	0.04	0.05
		SPI	0.19	0.12	0.22	0.22

Table 15
Random Forest classifier training accuracies for learning dataset importance for cluster-based batting and Bowling Roles.

Batting/bowling	Cluster-based role	Training accuracy (In %)	Testing accuracy (In %)	Domestic statistics importance	International statistics importance
Batting	F	52.08	91.67	0.56	0.44
	LC	81.25	83.33	0.55	0.45
	HWB	75.00	100	0.55	0.45
	SF	85.42	91.67	0.57	0.43
	SB	95.83	100	0.67	0.33
Bowling	HWB	91.49	91.67	0.49	0.51
	SP	54.55	83.33	0.75	0.25
	LC	95.65	83.33	0.68	0.32
	SB	100	91.67	0.59	0.41
	IB	60.47	91.67	0.66	0.34

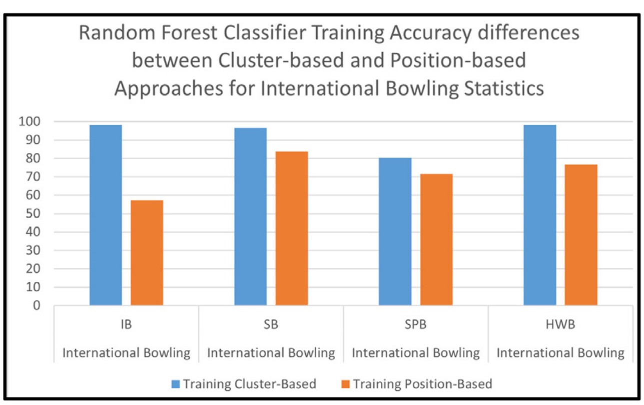

Fig. 5. Comparison of Random Forest classifier's training accuracies between cluster-based and position-based role identification approaches based on international bowling statistics.

does not justify the true potential of that player as it does not capture the player's contributions correctly to the team.

4.3. Calculating dataset importance using random forest algorithm for cluster-based role identification approach

Table 15 shows the Random Forest classifier's training results and the normalized dataset importance values for different cluster-based roles. From Table 15, one can see that depending on the skill set and cluster-based role, different importance values are assigned to the international statistics and domestic statistics.

4.4. Top ten batters and bowlers

Table 16 shows the top ten batters according to DPPI and other methods, as mentioned in the literature review section.

Table 17 shows the top ten bowlers according to DPPI and other methods, as mentioned in the literature review section.

5. Discussion

5.1. Theoretical contributions and implications

Figs. 6 and 7 show the comparison of DPPI with existing batting and bowling performance evaluation methods. We select the top fifteen run-scorers and wicket-takers in the IPL 2019 season to compare the results of different methods. First, we check on which position each method has ranked these fifteen run-scorers and wicket-takers according to the runs scored and wickets taken by them. Then, we calculate the Root Mean Squared Error (RMSE) for each method, as shown in Eq. (4), to compare the performance of different methods.

Let 'i' be a player among the top fifteen run-scorers (or top fifteen wicket-takers). N = 15 because we evaluate the method based on the top fifteen run-scorers and wicket-takers. x_{i_act} is the actual rank of the player in the IPL 2019 season among the highest run-scorers (or wicket-takers). x_{i_pred} is the predicted rank by the performance evaluation method for IPL 2019 season among the run-scorers (or wicket-takers). Figs. 6 and 7 show that DPPI captures a player's capability better than other existing methods as the RMSE value for DPPI is the least compared to all other methods. Thus, DPPI shows that capturing a player's form and role is necessary to identify a player's true performance.

$$\mathrm{RMSE} = \sqrt{\left(\sum_{i=1}^{15}(\mathrm{xi\_act} - \mathrm{xi\_pred})^2\right)/N} \tag{4}$$

Table 16
Top ten batters according to each method.

Method	Top Ten Batters
DPPI (Cluster-based)	Virat Kohli, David Warner, KL Rahul, AB de Villiers, Chris Gayle, Quinton de Kock, Evin Lewis, Kane Williamson, Rohit Sharma, Shikhar Dhawan.
Method 1	Virat Kohli, KL Rahul, Evin Lewis, Colin Munro, Manish Pandey, Faf du Plessis, Dinesh Karthik, MS Dhoni, Chris Gayle, Martin Guptill.
Method 2	Virat Kohli, Rohit Sharma, Martin Guptill, MS Dhoni, Chris Gayle, David Warner, Colin Munro, Suresh Raina, AB de Villiers, Faf du Plessis.
Method 3	Rohit Sharma, Martin Guptill, Virat Kohli, Chris Gayle, Colin Munro, David Warner, Shane Watson, AB de Villiers, Suresh Raina, Faf du Plessis.
Method 4	Virat Kohli, KL Rahul, Dinesh Karthik, Manish Pandey, MS Dhoni, Colin Munro, Evin Lewis, Moises Henriques, David Miller, Faf du Plessis.
Method 5	Evin Lewis, Colin Munro, KL Rahul, Virat Kohli, Dinesh Karthik, Chris Gayle, Shane Watson, Faf du Plessis, Hardik Pandya, David Miller.
Method 6	Virat Kohli, Rohit Sharma, Martin Guptill, Colin Munro, Chris Gayle, MS Dhoni, David Warner, Suresh Raina, Shane Watson, AB de Villiers,
Method 7	Virat Kohli, KL Rahul, Evin Lewis, Colin Munro, Manish Pandey, Faf du Plessis, Dinesh Karthik, MS Dhoni, Chris Gayle, Martin Guptill.
Method 8	Evin Lewis, KL Rahul, Colin Munro, Virat Kohli, Dinesh Karthik, Chris Gayle, Shane Watson, Faf du Plessis, Rohit Sharma, David Miller.
Method 9	Virat Kohli, Colin Munro, KL Rahul, Dinesh Karthik, Manish Pandey, MS Dhoni, Evin Lewis, Moises Henriques, David Miller, Faf du Plessis.

Table 17
Top ten bowlers according to each method.

Method	Top Ten Bowlers
DPPI (Cluster-based)	Sandeep Lamichhane, Lungi Ngidi, Andrew Tye, Rashid Khan, Mujeeb Ur Rahman, Kuldeep Yadav, Imran Tahir, Nathan Coulter Nile, Lasith Malinga, Mitchell McClenaghan.
Method 1	Rashid Khan, Joe Denly, Lungi Ngidi, Kuldeep Yadav, Mujeeb Ur Rahman, Amit Mishra, Lockie Ferguson, Imran Tahir, Washington Sundar, Dale Steyn.
Method 2	Rashid Khan, Lasith Malinga, Shakib Al Hasan, Imran Tahir, Dale Steyn, Mohammad Nabi, Tim Southee, Sunil Narine, Jasprit Bumrah, R. Ashwin.
Method 3	Lasith Malinga, Rashid Khan, Shakib Al Hasan, Imran Tahir, Dale Steyn, Tim Southee, Mohammad Nabi, Kuldeep Yadav, Jasprit Bumrah, Yuzvendra Chahal.
Method 4	Joe Denly, Lockie Ferguson, Lungi Ngidi, Washington Sundar, Amit Mishra, Rashid Khan, Kuldeep Yadav, Mujeeb Ur Rahman, Jason Behrendorff, Beauren Hendricks.
Method 5	Rashid Khan, Lasith Malinga, Shakib Al Hasan, Imran Tahir, Dale Steyn, Mohammad Nabi, Sunil Narine, Jasprit Bumrah, R. Ashwin, Tim Southee.
Method 6	Joe Denly, Lungi Ngidi, Rashid Khan, Kuldeep Yadav, Lockie Ferguson, Imran Tahir, Amit Mishra, Jason Behrendorff, Beauren Hendricks, Yuvraj Singh.
Method 7	Joe Denly, Lungi Ngidi, Rashid Khan, Kuldeep Yadav, Lockie Ferguson, Imran Tahir, Amit Mishra, Jason Behrendorff, Washington Sundar, Dale Steyn.
Method 8	Joe Denly, Lungi Ngidi, Lockie Ferguson, Kuldeep Yadav, Rashid Khan, Beauren Hendricks, Imran Tahir, Jason Behrendorff, Chris Morris, Amit Mishra.
Method 9	Joe Denly, Lockie Ferguson, Rashid Khan, Kuldeep Yadav, Amit Mishra, Washington Sundar, Lungi Ngidi, Mujeeb Ur Rahman, Imran Tahir, Chris Morris.

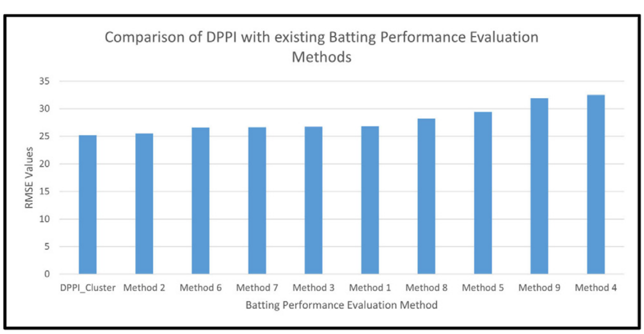

Fig. 6. Comparison of DPPI with existing batting performance evaluation methods.

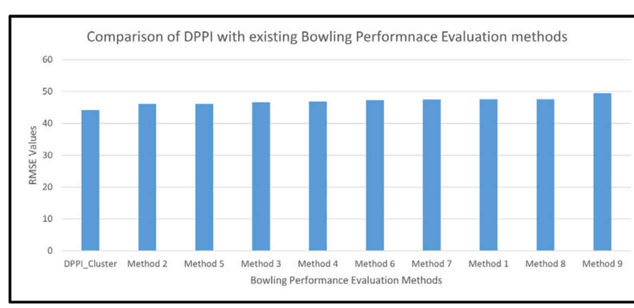

Fig. 7. Comparison of DPPI with existing bowling performance evaluation methods.

The first modified guideline for player performance evaluation suggests that the proposed model should be a statistical model. DPPI methodology follows a statistical approach where the players are evaluated based on KPIs. Further, the K-Means Clustering algorithm identifies player roles and the Random Forest classifier determines the feature importance. The second guideline suggests that the model should evaluate each player according to their current form and role in the team. DPPI does the same. The third guideline states that the run-scoring and wicket-taking capability should be captured apart from the number of runs scored and wickets taken. In DPPI, the number of runs scored and the number of wickets taken by the player are essential for the KPIs. Moreover, DPPI also captures the player's run-scoring ability and wicket-taking ability according to the role in the team. Fourth, as it is required to make a trade-off between the simplicity and complexity of the model, DPPI does it by considering common KPIs for all batters and bowlers and then converting them to role-specific KPIs through different feature importance values. Finally, as far as the explainability of DPPI is concerned, at every stage, DPPI explains why a KPI is given more importance for a specific cluster-based role. Thus, DPPI satisfies all the modified player performance evaluation guidelines and, more importantly, provides future research directions to develop performance evaluation indexes that follow these guidelines.

Apart from the modified player performance evaluation guidelines, DPPI also makes the following theoretical contributions. First, DPPI shows how to find the cluster-based roles and use the supervised classification techniques to build classifiers based on cluster-based roles. The cluster-based role identification approach can also be used in other sports by using sport-specific KPIs. To the best of our knowledge, such an approach has not been utilized in Cricket hitherto and represents an important advance in state of the art. Second, DPPI shows that by combining K-Means Clustering results (unsupervised learning) with Random Forest classifier (supervised learning), feature importance can be obtained [63]. Researchers have combined unsupervised learning and supervised learning methods to generate sports highlights in the audio domain [68] and distinguish between the on-ball screen and non-on-ball screen in Baseball [69]. Further, this combination has also been used in other decision analytics tasks, e.g., for credit card fraud detection [70], peak load forecasting [71] and dynamic security assessment [72]. Taking the idea forward to design a comprehensive metric for cricket performance evaluation is the significant contribution of this work.

5.2. Implications for practice

DPPI serves multiple purposes. First, it enables the comparison of T20 players playing similar roles in different teams. Our comparison of DPPI with existing methods shows that DPPI captures a player's performance better than existing performance evaluation indexes. Second, for predictive modeling, the DPPI values of these T20 players playing at different positions can predict the match outcome. Thus, practitioners can use DPPI for both predictive analytics and prescriptive analytics. For example, Table 18 shows the batting and bowling strength of IPL 2019 season finalist franchises. For both CSK and MI, the eleven players that played the final are shown in Table 18, along with their batting and bowling DPPI scores. As per the IPL laws, an IPL franchise can play a maximum of only four overseas players, and they are indicated by using (*) in front of their names. Average batting strength is calculated as the summation of the batting DPPIs of the top seven batters (position wise) divided by seven. Further, average bowling strength is calculated as the summation of bowling DPPIs of the five best bowlers who are supposed to bowl four overs each divided by five. Thus, the teams' average batting and bowling strength reflect the overall team strength. Interestingly, the top two teams in the league are the ones that are predicted by aggregated DPPI values of the playing XI of respective teams. Table 19 shows why the top two teams were better than other teams according to the aggregated DPPIs. Further, for predictive modeling, different heuristic-based models [6] can be developed using the DPPI scores to predict the match result. Thus, DPPI enables sports enthusiasts, analysts, fantasy Cricket users, fans, coaches, and managers to understand and evaluate the players' true potential individually and the team overall. Franchises invest heavy sums of money in acquiring players to play for their teams. The index provided can prove to be an excellent guide on where they could put their money with advantage. The index can also help the team management in finalizing the playing eleven out of the squad available because it reflects not only the potential but also the current form of the player.

5.3. Limitations and future research directions

The limitations of DPPI are as follows. First, it does not attempt to capture the player's on-field performance in real-time. It tries to approximate a player's on-field performance based on his current form and cluster-based role in the team. Second, DPPI only considers match statistics-related KPIs while calculating the in-form and role-based DPPI values and does not consider other parameters such as the weather and pitch conditions. Third, DPPI's performance depends on our KPIs to evaluate a batter or a bowler. A different set of KPIs might result in different batting and bowling strengths. Fourth, instead of a position-based approach of evaluating batters, we use a cluster-based approach to identify their roles. A better approach to identify player roles than the cluster-based approach will further improve the player's strength capturing capability of DPPI. Fifth, for domestic players without IPL and T20 internationals experience, their Syed Mushtaq Ali Tournament (SMAT) statistics are considered. However, we acknowledge a difference in the level of play between IPL/T20 internationals and SMAT. Sixth, we capture a player's form by considering his performance statistics from the last four IPL seasons and then assigning higher importance to recent performance. However, a player's performance can change drastically within a year which our model does not capture because IPL happens once every year. Seventh, while capturing a player's current form, we make a trade-off between the simplicity and complexity of our approach. Ideally, the weights for a player who played in 2016 and 2017 vis-à-vis a player who played in 2017 and 2018 should be different. However, in our approach, we assign the same weights to both the subcases and club them under players who played two seasons. Similarly, we follow the logic for players who played three seasons and one season. In this manner, we reduce the number of combinations to four cases. Eighth, DPPI shows Sandeep Lamichhane from Nepal as the top bowler. However, we acknowledge that his T20 international statistics (although quite good) are from matches played against fellow Associate Nations who are not among the top eight or ten among the International Cricket Council (ICC) rankings. Therefore, replicating the same statistics in IPL will be difficult but not impossible.

Some of the extensions are described here to build upon our proposed DPPI. First, since DPPI is a KPI-based methodology, in the future, researchers can explore other KPIs and evaluate the performance of their method to see if the inclusion of new KPIs serves their purpose. Second, DPPI can be extended to simulate a cricket match, i.e., given the context of the match (the progress pattern of last 'x' overs of the match), how the next 'y' overs are likely to take place. Third, DPPI can also be extended to find the right batting (when and whom to bat) and bowling (when and whom to bowl) positions in an ongoing match. Fourth, DPPI can be used to simulate the match in all possible bowling pairs and find the optimal bowler–batter pair to maximize winning. Fifth, a better approach can be developed to capture the performance of players playing in IPL for the first time. Thus, since DPPI captures a player's strength better than other existing approaches, DPPI has much potential in the sports analytics field in the coming years.

6. Conclusion

In this work, we adapted the existing FIFA player performance evaluation metric development guidelines to provide improved metrics in Cricket. The role-based and form-based player performance evaluation KPIs are obtained by combining the benefits of unsupervised learning and supervised learning algorithms. Both domestic statistics and international statistics are utilized with appropriate relative weights. A K-Means Clustering algorithm approach is used to classify players into different clusters and identify their roles based on the KPI values. A Random Forest Algorithm-based supervised classifier is used to obtain feature importance for different KPIs. Using empirical results, we demonstrate that the cluster-based role identification approach captures a player's potential in a better way than the position-based role identification approach. Comparison of our empirical results with extant methodologies on actual IPL 2019 season data clearly shows that our index, DPPI, captures a player's batting and bowling performances better as compared to other existing indexes.

In conclusion, DPPI improves understanding of a player's performance capability in this prescriptive analytics study. It serves as a helpful index for fantasy Cricket users, Cricket fans, coaches and managers to gain better insights into a player's batting and bowling performances. Cricket followers worldwide are very much interested in such data and the work would be of great interest to them. The index provided here is more comprehensive and illustrative than other extant popular indices.

Table 18
Team strength evaluation of top two teams in the IPL 2019.

Points table rank in IPL 2019	Franchise	Team (Player name, batting DPPI, bowling DPPI)	Batting strength and bowling strength
1	MI	Rohit Sharma © [0.57, 0.02] Quinton de Kock* (WK) [0.62, 0] Suryakumar Yadav [0.21, 0] Ishan Kishan [0.17, 0] Krunal Pandya [0.45, 0.64] Kieron Pollard* [0.37, 0.24] Hardik Pandya [0.35, 0.59] Jasprit Bumrah [0.17, 0.71] Mitchell McCleneghan* [0.15, 0.74] Lasith Malinga* [0.06, 0.75] Rahul Chahar [0, 0]	Average Batting Strength = sum (0.57, 0.62, 0.21, 0.17, 0.45, 0.37, 0.35)/7 = 0.39 Average Bowling Strength = sum (0.75, 0.74, 0.71, 0.59, 0.64)/5 = 0.69
2	CSK	Faf du Plessis* [0.53, 0] Shane Watson* [0.47, 0.58] [Did not bowl] Suresh Raina [0.52, 0.32] Ambati Rayudu [0.34, 0] MS Dhoni (WK, C) [0.54, 0] DJ Bravo* [0.41, 0.7] Ravindra Jadeja [0.3, 0.58] Shardul Thakur [0, 0.67] Deepak Chahar [0.16, 0.32] Harbhajan Singh [0.29, 0.64] Imran Tahir* [0, 0.75]	Average Batting Strength = sum (0.53, 0.47, 0.52, 0.34, 0.54, 0.41, 0.30)/7 = 0.46 Average Bowling Strength = sum (0.75, 0.64, 0.67, 0.58, 0.7)/5 = 0.67

Table 19
Evaluation of DPPI's predictive ability at the team level.

Franchise	Average batting strength	Average bowling strength	Aggregated strength (Sum of average batting and bowling strength)
Chennai Super Kings	0.46	0.67	1.13
Mumbai Indians	0.39	0.69	1.08
SunRisers Hyderabad	0.42	0.56	0.98
Kings XI Punjab	0.32	0.65	0.97
Royal Challengers Bangalore	0.38	0.56	0.94
Kolkata Knight Riders	0.29	0.58	0.87
Delhi Capitals	0.31	0.52	0.83
Rajasthan Royals	0.19	0.39	0.58

References

[1] J. Sanderson, K. Brown, COVID-19 and youth sports: Psychological, developmental, and economic impacts, Int. J. Sport Commun. 13 (3) (2020) 313–323, http://dx.doi.org/10.1123/ijsc.2020-0236.

[2] M.D. Hughes, R.M. Bartlett, The use of performance indicators in performance analysis./ utilisation des indicateurs de performance pour l ' analyse de cette performance, J. Sports Sci. 20 (10) (2002) 739–754, http://www.tandf.co.uk/journals%0A, http://articles.sirc.ca/search.cfm?id=S-847006%5Cnhttp://www.tandf.co.uk/journals%5Cnhttp://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=SPHS-847006&site=ehost-live.

[3] P. Steenkiste, Finding the optimal fantasy football team, 2015, http://cs229.stanford.edu/proj2015/105_report.pdf.

[4] Y. Yang, T. Qin, Y.H. Lei, Real-time esports match result prediction, 2016, arXiv, Nips, 1–9.

[5] M. Gowda, A. Dhekne, S. Shen, R.R. Choudhury, X. Yang, L. Yang, S. Golwalkar, A. Essanian, Bringing IoT to sports analytics, in: Proceedings of the 14th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2017, pp. 499–513.

[6] C. Deep, C. Patvardhan, C. Vasantha, Data analytics based deep mayo predictor for IPL-9, Int. J. Comput. Appl. 152 (6) (2016) 6–11, http://dx.doi.org/10.5120/ijca2016911875.

[7] L.M. Hvattum, H. Arntzen, Using ELO ratings for match result prediction in association football, Int. J. Forecast. 26 (3) (2010) 460–470, http://dx.doi.org/10.1016/j.ijforecast.2009.10.002.

[8] T. Singh, V. Singla, P. Bhatia, Score and winning prediction in cricket through data mining, in: International Conference on Soft Computing Techniques and Implementations, ICSCTI 2015, 2016, pp. 60–66, http://dx.doi.org/10.1109/ICSCTI.2015.7489605.

[9] C. Deep, C. Patvardhan, S. Singh, A new machine learning based deep performance index for ranking IPL T20 cricketers, Int. J. Comput. Appl. 137 (10) (2016) 42–49, http://dx.doi.org/10.5120/ijca2016908903.

[10] The laws of cricket | MCC. (n.d.), 2021, https://www.lords.org/mcc/the-laws-of-cricket. (Retrieved 25 December 2021).

[11] M. Imtiaz, S. Akif, M.Imtiaz Subhani, A. Osman, S.Akif Hasan, B. Hussain, Will T20 Clean Sweep Other Formats of Cricket in Future? Munich Personal RePEc Archive, 45144.

[12] D. Sturm, Smash and bash cricket? Affective technological innovations in the big bash, Media Int. Australia 155 (2015) 80–88, http://dx.doi.org/10.1177/1329878x1515500110.

[13] J. Davis, H. Perera, T.B. Swartz, A simulator for twenty20 cricket, Aust. N. Z. J. Stat. 57 (1) (2015) 55–71, http://dx.doi.org/10.1111/anzs.12109.

[14] A.B.W. Manage, S.M. Scariano, C.R. Hallum, Performance analysis of T20-world cup cricket 2012, Sri Lankan J. Appl. Stat. 14 (1) (2013) 1, http://dx.doi.org/10.4038/sljastats.v14i1.5873.

[15] C. Petersen, D.B. Pyne, M.R. Portus, B. Dawson, Quantifying positional movement patterns in twenty20 cricket, Int. J. Perform. Anal. Sport 9 (2) (2009) 165–170, http://dx.doi.org/10.1080/24748668.2009.11868474.

[16] C. Petersen, D.B. Pyne, M.J. Portus, B. Dawson, Analysis of twenty/20 cricket performance during the 2008 Indian premier league, Int. J. Perform. Anal. Sport 8 (3) (2008) 63–69, http://dx.doi.org/10.1080/24748668.2008.11868448.

[17] R. Abrahm, A $150bn market: How fantasy sports transformed cricket fans from being spectators to stakeholders, 2019, The Economic Times, https://economictimes.indiatimes.com/magazines/panache/a-150bn-market-how-fantasy-sports-transformed-cricket-fans-from-being-spectators-to-stakeholders/articleshow/70185196.cms?from=mdr.

[18] L. Jajo, Sports industry development in India: Opportunities, constraints and way forward, Imp. J. Interdiscip. Res. 2 (12) (2016) 700–709.

[19] I.G. McHale, P.A. Scarf, D.E. Folker, On the development of a soccer player performance rating system for the english premier league, Interfaces 42 (4) (2012) 339–351, http://dx.doi.org/10.1287/inte.1110.0589.

[20] A. Rasul, J.M. Proffitt, Bollywood and the Indian premier league (IPL): The political economy of bollywood's new blockbuster, Asian J. Commun. 21 (4) (2011) 373–388, http://dx.doi.org/10.1080/01292986.2011.580851.

[21] D. Steinley, K-means clustering: A half-century synthesis, Br. J. Math. Stat. Psychol. 59 (1) (2006) 1–34, http://dx.doi.org/10.1348/000711005X48266.

[22] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[23] T. Basevi, G. Binoy, The world's best twenty20 players, 2007, Espncricinfo. https://www.espncricinfo.com/story/the-world-s-best-twenty20-players-311962.

[24] M. Ramakrishnan, Gayle, Pieterson steal the show, 2012, ESPN Cricket. https://www.espn.in/cricket/story/_/id/22317940/chris-gayle-kevin-pietersen-steal-show.

[25] A.B.W. Manage, S.M. Scariano, An introductory application of principal components to cricket data, J. Stat. Educ. 21 (3) (2013) http://dx.doi.org/10.1080/10691898.2013.11889689.

[26] S. Bhogle, P. Maji, Gayle leads MVP race, de Villiers not far behind, 2015, Rediff.Com. https://www.rediff.com/cricket/report/icc-world-cup-gayle-leads-mvp-race-de-villiers-not-far-behind/20150301.htm.

[27] TheCricketCouch, A simple metric to understand batting efficiency in the IPL, 2012, Thecricketcouch.Com. http://thecricketcouch.com/blog/2012/06/15/a-simple-metric-to-understand-batting-efficiency-in-the-ipl/.

[28] J.S. Croucher, Player ratings in one-day cricket, in: Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport, 2000, pp. 95–106, http://www.mathsportinternational.com/anziam/Mathsport5.pdf#page=104.

[29] G.D.I. Barr, B.S. Kantor, A criterion for comparing and selecting batsmen in limited overs cricket, J. Oper. Res. Soc. 55 (12) (2004) 1266–1274, http://dx.doi.org/10.1057/palgrave.jors.2601800.

[30] C. Deep, C. Patvardhan, S. Singh, A new category base deep performance index using machine learning for ranking IPL cricketers, Int. J. Electr. Comput. Syst. 5 (2) (2016) 37–47, https://www.researchgate.net/profile/Chellapilla-Deep-Prakash/publication/310240177_A_new_Category_Based_Deep_Performance_Index_using_Machine_Learning_for_ranking_IPL_Cricketers/links/582e5d8708ae102f072db8bc/A-new-Category-Based-Deep-Performance-Index-us.

[31] A. Zimmermann-Niefield, R.B. Shapiro, S. Kane, Sports and machine learning: How young people can use data from their own bodies to learn about machine learning. XRDS: Crossroads, ACM Mag. Stud. 25 (4) (2019) 44–49.

[32] R.Kyle Martin, A. Pareek, A.J. Krych, H.M. Kremers, L. Engebretsen, Machine learning in sports medicine: Need for improvement, J. ISAKOS 6 (1) (2021) 1–2, http://dx.doi.org/10.1136/jisakos-2020-000572.

[33] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, F. Giannotti, PlayeRank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach, ACM Trans. Intell. Syst. Technol. (TIST) 10 (5) (2019) 1–27, http://dx.doi.org/10.1145/3343172.

[34] R.P. Bunker, F. Thabtah, A machine learning framework for sport result prediction, Appl. Comput. Inform. 15 (1) (2019) 27–33, http://dx.doi.org/10.1016/j.aci.2017.09.005.

[35] B.C. Luu, A.L. Wright, H.S. Haeberle, J.M. Karnuta, M.S. Schickendantz, E.C. Makhni, B.U. Nwachukwu, III, P.N. Ramkumar, Machine learning outperforms logistic regression analysis to predict next-season NHL player injury: An analysis of 2322 players from 2007 to 2017, Orthop. J. Sports Med. 8 (9) (2020) 2325967120953404, http://dx.doi.org/10.1177/2325967120953404.

[36] T. Reichherzer, M. Timm, N. Earley, N. Reyes, V. Kumar, Using machine learning techniques to track individuals & their fitness activities, in: Proceedings of the 32nd International Conference on Computers and Their Applications, CATA, 2017, pp. 119–124.

[37] G. Lee, V. Bulitko, E. Ludvig, Sports commentary recommendation system (SCoReS): Machine learning for automated narrative, in: Proceedings of the 8th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2012, 2012, pp. 32–37.

[38] C.Deep Prakash, A. Majumdar, Analyzing the role of national culture on content creation and user engagement on Twitter: The case of Indian premier league cricket franchises, Int. J. Inf. Manage. 57 (2020) (2021) 102268, http://dx.doi.org/10.1016/j.ijinfomgt.2020.102268.

[39] R. Al-Shboul, T. Syed, J. Memon, F. Khan, Automated player selection for sports team using competitive neural networks, Int. J. Adv. Comput. Sci. Appl. 8 (8) (2017) 457–460, http://dx.doi.org/10.14569/ijacsa.2017.080859.

[40] R. Beal, T.J. Norman, S.D. Ramchurn, Optimising daily fantasy sports teams with artificial intelligence, Int. J. Comput. Sci. Sport 19 (2) (2020) 21–35, http://dx.doi.org/10.2478/ijcss-2020-0008.

[41] S. Kampakis, W. Thomas, Using machine learning to predict the outcome of english county twenty over cricket matches, 2015, arXiv Preprint arXiv:1511.05837, 1–17.

[42] K. Passi, N. Pandey, Increased prediction accuracy in the game of cricket using machine learning, Int. J. Data Min. Knowl. Manag. Process 8 (2) (2018) 19–36, http://dx.doi.org/10.5121/ijdkp.2018.8203.

[43] K. Passi, N. Pandey, Predicting players' performance in one day international cricket matches using machine learning, Comput. Sci. Inf. Technol. 11 (2018) 1–126, http://dx.doi.org/10.5121/csit.2018.80310.

[44] M.S. Oughali, M. Bahloul, S.A. El Rahman, Analysis of NBA players and shot prediction using random forest and XGBoost models, in: 2019 International Conference on Computer and Information Sciences, ICCIS 2019, 2019, http://dx.doi.org/10.1109/ICCISci.2019.8716412.

[45] V.C. Pantzalis, C. Tjortjis, Sports analytics for football league table and player performance prediction, in: 11th International Conference on Information, Intelligence, Systems and Applications, IISA 2020, 2020, http://dx.doi.org/10.1109/IISA50023.2020.9284352.

[46] G. Soliman, A. El-Nabawy, A. Misbah, S. Eldawlatly, Predicting all star player in the national basketball association using random forest, in: 2017 Intelligent Systems Conference, IntelliSys 2017, 2018-Janua, September, 2018, pp. 706–713, http://dx.doi.org/10.1109/IntelliSys.2017.8324371.

[47] P. Pugsee, P. Pattawong, Football match result prediction using the random forest classifier, PervasiveHealth Pervasive Comput. Technol. Healthc. 15 (2019) 4–158, http://dx.doi.org/10.1145/3358528.3358593.

[48] R. Jia, C. Wong, D. Zeng, Predicting the major league baseball season, 2013, http://cs229.stanford.edu/proj2013/JiaWongZeng-PredictingTheMajorLeagueBaseballSeason.pdf.

[49] A. Basit, M.B. Alvi, F.H. Jaskani, M. Alvi, K.H. Memon, R.A. Shah, ICC T20 cricket world cup 2020 winner prediction using machine learning techniques, in: Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020, 2020, http://dx.doi.org/10.1109/INMIC50486.2020.9318077.

[50] Z. Shelly, R.F.V. Burch, W. Tian, L. Strawderman, A. Piroli, C. Bichey, Using k-means clustering to create training groups for elite american football student-athletes based on game demands, Int. J. Kinesiol. Sports Sci. 8 (2) (2020) 47–63, http://dx.doi.org/10.7575//aiac.ijkss.v.8n.2p.47.

[51] B. Spencer, S. Morgan, J. Zeleznikow, S. Robertson, W. Bulldogs, F. Club, 2016. Clustering team profiles in the australian football league using performance indicators, in: The 13th Australasian Conference on Mathematics and Computers in Sport, June.

[52] D.F. Gucciardi, M.I. Jones, Beyond optimal performance: Mental toughness profiles and developmental success in adolescent cricketers, J. Sport Exerc. Psychol. 34 (1) (2012) 16–36, http://dx.doi.org/10.1123/jsep.34.1.16.

[53] P.J. Rani, A.V. Kamath, A. Menon, P. Dhatwalia, D. Rishabh, A. Kulkarni, Selection of players and team for an indian premier league cricket match using ensembles of classifiers, in: 2020 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCT, 2020, pp. 1–6, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9198371&casa_token=RGczrMRniPQAAAAA:BjthyvP0hx712cYhA-PYB4YhV8dnUDDZqjM20pZ5_bGWCi_oLfaqClc6u9CPdMAIx3BDhS2xlZQ.

[54] L. Hadley, M. Poitras, J. Ruggiero, S. Knowles, Performance evaluation of national football league teams, Manag. Decis. Econ. 21 (2) (2000) 63–70, http://dx.doi.org/10.1002/1099-1468(200003)21:2<63::AID-MDE964>3.0.CO;2-O.

[55] M.J. Mazur, Evaluating the relative efficiency of baseball players, Data Envel. Anal. Theory Methodol. Appl. 36 (1994) 9–391, http://dx.doi.org/10.1007/978-94-011-0637-5_19.

[56] V. Sarlis, V. Chatziilias, C. Tjortjis, D. Mandalidis, A data science approach analysing the impact of injuries on basketball player and team performance, Inf. Syst. 99 (2021) 101750, http://dx.doi.org/10.1016/j.is.2021.101750.

[57] D. Rösch, R. Hodgson, L. Peterson, T. Graf-Baumann, A. Junge, J. Chomiak, J. Dvorak, Assessment and evaluation of football performance, Am. J. Sports Med. 28 (5 SUPPL.) (2000) 29–39, http://dx.doi.org/10.1177/28.suppl_5.s-29.

[58] M. Ramakrishnan, Steyn, Gayle take pole position, 2012, Espncricinfo. https://www.espncricinfo.com/story/ipl-2012-performance-analysis-dale-steyn-chris-gayle-take-pole-position-566523.

[59] G.D.I. Barr, C.G. Holdsworth, B.S. Kantor, Evaluating performances at the 2007 cricket world cup, South African Statist. J. 42 (2) (2008) 125–142.

[60] H.H. Lemmer, The combined bowling rate as a measure of bowling performance in cricket, South Afr. J. Res. Sport Phys. Educ. Recreat. 24 (2) (2002) 37–44, https://journals.co.za/doi/pdf/10.10520/EJC108746.

[61] M.A. Syakur, B.K. Khotimah, E.M.S. Rochman, B.D. Satoto, Integration k-means clustering method and elbow method for identification of the best customer profile cluster, IOP Conf. Ser. Mater. Sci. Eng. 336 (1) (2018) http://dx.doi.org/10.1088/1757-899X/336/1/012017.

[62] O.A. Abbas, Comparisons between data clustering algorithms, Int. Arab J. Inf. Technol. 5 (3) (2008).

[63] D. Zakrzewska, On integrating unsupervised and supervised classification for credit risk evaluation, Inf. Technol. Control 36 (1A) (2007) 98–102.

[64] R.L. Marchese Robinson, A. Palczewska, J. Palczewski, N. Kidley, Comparison of the predictive performance and interpretability of random forest and linear models on benchmark data sets, J. Chem. Inf. Model. 57 (8) (2017) http://dx.doi.org/10.1021/acs.jcim.6b00753.

[65] S.H. Al-Harbi, V.J. Rayward-Smith, Adapting k-means for supervised clustering, Appl. Intell. 24 (3) (2006) 219–226, http://dx.doi.org/10.1007/s10489-006-8513-8.

[66] R. Rifkin, A. Klautau, In defense of one-vs-all classification, J. Mach. Learn. Res. 5 (2004) 2–6.

[67] R.I. Lerman, S. Yitzhaki, A note on the calculation and interpretation of the Gini index, Econom. Lett. 15 (3–4) (1984) 363–368, http://dx.doi.org/10.1016/0165-1765(84)90126-5.

[68] R. Radhakrishan, Z. Xiong, A. Divakaran, Y. Ishikawa, Generation of sports highlights using a combination of supervised & unsupervised learning in audio domain, in: ICICS-PCM 2003 - Proceedings of the 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing and 4th Pacific-Rim Conference on Multimedia, 2, 935–939, 2021, http://dx.doi.org/10.1109/ICICS.2003.1292595.

[69] N. Seward, Building an on-Ball Screen Dataset using Supervised and Unsupervised Learning (Issue September), University of Ontario Institute of Technology, 2018, https://ir.library.dc-uoit.ca/xmlui/bitstream/handle/10155/984/Seward_Neil.pdf?sequence=1.

[70] F. Carcillo, Y.A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, G. Bontempi, Combining unsupervised and supervised learning in credit card fraud detection, Inform. Sci. 557 (2021) 317–331, http://dx.doi.org/10.1016/J.INS.2019.05.042.

[71] M.R. Amin-Naseri, A.R. Soroush, Combined use of unsupervised and supervised learning for daily peak load forecasting, Energy Convers. Manage. 49 (6) (2008) 1302–1308, http://dx.doi.org/10.1016/j.enconman.2008.01.016.

[72] Y.H. Pao, D.J. Sobajic, Combined use of unsupervised and supervised learning for dynamic security assessment, IEEE Trans. Power Syst. 7 (2) (1992) 878–884, http://dx.doi.org/10.1109/59.141799.

Update

Decision Analytics Journal
Volume 7, Issue , June 2023, Page

DOI: https://doi.org/10.1016/j.dajour.2023.100195

Decision Analytics Journal 7 (2023) 100195

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

Erratum to "A new in-form and role-based Deep Player Performance Index for player evaluation in T20 Cricket" [Decis. Anal. J. 2 (2022) 100025]

C. Deep Prakash *, Sanjay Verma

Information Systems Department, Indian Institute of Management Ahmedabad, Vastrapur, Ahmedabad 380015, India

The publisher regrets to inform that Declaration of Competing Interest statement was not included in the published version of this article. The publisher would like to apologise for any inconvenience caused.

The appropriate Declaration/Competing Interest statements, provided by the Authors, is included below.

I hereby state that this paper is not published in any form in any outlet and it is not under review in any other journal. If you have any questions regarding this submission, it would be my pleasant duty to address them.

DOI of original article: https://doi.org/10.1016/j.dajour.2022.100025.

* Corresponding author.

E-mail address: phd18cdeepp@iima.ac.in (C. Deep Prakash).

https://doi.org/10.1016/j.dajour.2023.100195

Available online 16 March 2023