Smart Stats

Indian Institute of Technology Madras, Gyan Data Private Limited, ESPNcricinfo

July 2021

1 Abstract

In this paper, we discuss Smart Stats - a suite of advanced metrics for context-informed evaluation of players' batting and bowling performances in T20 cricket. Smart Stats comprises of two primary metrics - "smart runs" and "smart wickets", which are adjusted values of runs scored/conceded and wickets taken by a player respectively. The values are adjusted upon analysing the runs and wickets in the context of the game. Smart runs account for the level of pressure on the player's team, and the rate at which the player scores/concedes runs relative to the expected run rate. Smart wickets factor in the quality of the batsman dismissed and the pressure on the bowling team. We also introduce a metric called "impact score" that quantifies the overall performance of a player based on the number of smart runs he scored/conceded and smart wickets he scalped. The Smart Stats metrics are developed on top of the Forecaster - a tool that predicts the end-of-innings score and win probabilities of the teams at any given state of a match. Smart Stats and Forecaster are two of the three modules of ESPNcricinfo Superstats, a cricket analytics platform jointly developed by Indian Institute of Technology Madras, Gyan Data Private Limited, and ESPNcricinfo. The predictions and analyses from Superstats were aimed at increasing fan engagement on the ESPNcricinfo website.

2 Introduction [1, 2]

Cricket is the second most popular sport in the world after football/soccer with an estimated fan base of 2.5 billion, mostly concentrated in the Indian subcontinent [3]. It is a bat-and-ball game similar to baseball and is played between two teams of 11 players each. Every ball bowled is an event and the outcomes for both the batsman and bowler off a ball are easily quantifiable in terms of runs scored/conceded, balls faced/bowled, dismissals, and so on, thereby making the sport readily amenable to data mining. Cricket is primarily played in three formats - Test, One-day International (ODI) and Twenty20 (T20). A game of Test cricket can last up to five days and each team is allowed a maximum of two batting innings. There is however no upper bound on the number of deliveries a team is allowed to bat in an innings in Test cricket. ODI and T20 formats are together referred to as limited overs cricket and allow only one batting innings for each team. An innings in limited overs cricket has an upper bound on the number of deliveries that can be bowled. The limit in ODI cricket is 50 overs (each over consists of six deliveries) and T20 cricket is 20 overs. In Test cricket, an innings is completed when either 10 batsmen are dismissed or the batting team voluntarily "declares" to not bat anymore and in limited overs cricket, an innings is completed when either 10 batsmen are dismissed or the limit on the number of deliveries is reached. A team has to score more runs in total than their opposition to win a match.

Nuanced evaluation of players is perhaps the most important analytical problem that concerns any professional sports team. Identifying and selecting the best players for the different roles in the team goes a long way in improving the team's chances of winning. All the past performances of any player are summarized by a set of metrics like average, strike rate, and economy rate, which often form the basis of selection decisions. Such metrics also grab the attention of the fans as they too are interested in knowing who are the best players and how well their favourite players are performing relative to others. The two most widely used metrics for summarizing all the batting performances of a player are batting average and batting strike rate. Batting average of a player is defined as the ratio of the number of runs scored by the player to the number of times the player was dismissed. It tells us how many runs the player typically scores before he is dismissed. Since it is also important to understand how quickly a player scores his runs, especially in limited overs cricket, the batting strike rate metric was introduced. Batting strike rate is 100 times the ratio of the number of runs scored by the player to the number of balls faced by the player. The two most widely used metrics for summarizing all the bowling performances of a player are bowling strike rate and bowling economy rate. Bowling strike rate is the ratio of the number of balls bowled by the player to the number of wickets he took and bowling economy rate refers to the number of runs conceded by the player per over. Though these metrics are highly popular in the cricketing community, they do not make a fair assessment of players' performances. These metrics treat every run and every wicket the same and therefore, do not reveal the true value of the runs and wickets in the context of the game. Hence there is a need for a re-calibration for measuring batting and bowling performances.

Given the vast differences in the playing rules among the formats, it is necessary to devise format-specific metrics to evaluate batting and bowling performances. The focus of this paper is on T20 cricket, the game's newest and arguably the most popular format owing to its relatively short duration of four hours and more frequent occurrence of boundaries and wickets. Due to the popularity of the format, T20 leagues are organized in every major cricketing nation where both domestic and international players participate. These leagues include the Indian Premier League (IPL), Pakistan Super League (PSL), Big Bash League (BBL), and others. The increasing number of T20 matches and interest surrounding the format has given rise to a deluge in data and hence, scope for advanced analytics applications in the space of T20 cricket.

Let us motivate the need for advanced metrics in T20 cricket through some examples. Consider two knocks - Devdutt Padikkal's 35 runs off 37 balls in the 33^{rd} match of IPL 2020 [4] and David Warner's 35 runs off 20 balls in the 43^{rd} match of IPL 2020 [5]. Though both the knocks are worth the same number of runs, a keen cricket follower would know that the two knocks were starkly different in terms of impact on the team's chances of winning. Warner provided a brisk start to the innings by scoring a sizeable number of runs at well over the match run rate and reduced the required run rate for his team considerably. When Warner got out, the ask for his team was to score only 71 runs in 82 balls with nine wickets in hand. Though his teammates failed to seal the chase, it is fair to say that Warner had done his job. On the other hand, Padikkal played an over-cautious knock by scoring at much lower than the match run rate in spite of having enough batsmen to follow him. His knock kept increasing the required run rate throughout the innings and mounted pressure on his teammates. The required run rate increased from 8.9 at the start of the innings to 10.86 by the time Padikkal got out. If not for his teammate AB de Villiers' outstanding knock of $55^{*}(22)$ at the end, his team would have lost the match. In spite of the difference in impact, the effect of these knocks on the players' batting average is the same as both the knocks are valued at 35 runs.

The same problem exists with batting strike rate. Consider another pair of knocks - MS Dhoni's 29 runs off 17 balls in the 4^{th} match of IPL 2020 [6] and Corey Anderson's 41 off 24 balls in the 40^{th} match of IPL 2017 [7]. The strike rate of both the knocks is around 170. However, similar to the Padikkal-Warner example discussed previously, the impact of these two knocks on the team's chances of winning were starkly different. When Anderson came out to bat, his team required 77 from 50 balls against a strong bowling attack comprising the likes of Bhuvneshwar Kumar and Rashid Khan. Anderson's knock ensured that his team chased down the target and won the match. Dhoni's innings, on the other hand, made a difficult yet possible run chase impossible by scoring only 10 runs in the first 13 balls he faced. He hit three sixes in the last over and increased his strike rate to 170 but the match was already lost by that time. It is clear that Anderson's scoring rate was more meaningful than that of Dhoni though their strike rates are the same.

The argument that runs scored by batsmen should not be taken at face value holds even for wickets taken and runs conceded by bowlers. Consider the three-wicket hauls from two bowlers - Imran Tahir in the 34^{th} match of IPL 2017 [8] and Sandeep Sharma in the 43^{rd} match of IPL 2017 [9]. Imran Tahir's three wickets included Pawan Negi, Adam Milne, and Samuel Badree. These three players are primarily bowlers and not known for their batting. Also, Tahir's wickets came about after the match was already in his team's favour - the opposition needed 97 from 45 balls with just 5 wickets in hand. So his wickets did not have a significant impact on the outcome of the game. To the contrary, Sharma's three wickets were of Chris Gayle, AB de Villiers, and Virat Kohli - three of the greatest T20 batsmen in the world. Moreover, Sharma was bowling under high pressure as his team was defending a small total of 138 against a strong batting lineup. Sharma's wickets turned out to be the most telling contribution towards his team's win. It is therefore obvious that his wickets should be valued much more than those of Tahir.

Finally, let us look at an example to understand how economy rates can also be misleading. Consider the spells from two bowlers - Ben Stokes in the 50^{th} game of IPL 2020 [10] and Adam Zampa in the final match of IPL 2017 [11]. Both spells were four overs long and were delivered at an economy rate of 8 each. If we take a more detailed look at the data, we can see that Stokes bowled at an economy rate of 8 in a match where the run rate of the batsmen across both innings was 9.46 per over. Furthermore, one of his overs was bowled in the death where batsmen typically put a lesser price on their wicket and tend to score as many boundaries as possible. Zampa's spell was in a match where the batsmen run rate across both innings was 6 per over. Zampa also bowled all his four overs in the 7^{th}-13^{th} over period where batsmen generally play conservatively and yet he conceded at more than the overall run rate in the match. Though the conventional economy rate values both the spells as the same, Stokes did a better job in containing the opposition batsmen than Zampa.

In this paper, we discuss Smart Stats - a suite of advanced metrics for context-informed evaluation of players' batting and bowling performances in T20 cricket. Smart Stats comprises of two primary metrics - "smart runs" and "smart wickets", which are adjusted values of runs scored/conceded and wickets taken by a player respectively. The values are adjusted upon analysing the runs and wickets in the context of the game. Smart runs account for the level of pressure on the player's team, and the rate at which the player scores/concedes runs relative to the expected run rate. "Smart strike rate" (SSR) and "smart economy rate" (SER) are two metrics that will be derived from smart runs. SSR is the number of smart runs scored by a player per 100 balls and SER is the number of smart runs conceded by a player per over. Smart wickets factor in the quality of the batsman dismissed and the pressure on the bowling team. Fantasy points is another metric that is gaining traction among cricket fans owing to the popularity of the fantasy leagues. There are a number of fantasy league providers that assign points to players based on their batting and bowling performances to rank them and determine who are the most valuable players. Each of these providers uses their own formula to convert the runs and wickets into fantasy points. As we have seen from the above examples, this can be flawed as conventional runs and wickets do not reveal the real value of a performance. We therefore introduce a metric called "impact score" that quantifies the overall performance of a player in a match based on the number of smart runs he scored/conceded and smart wickets he scalped. The Smart Stats metrics are developed on top of the Forecaster - a tool that predicts the final scores and the win probabilities of the teams at any given state of a match. Smart Stats and Forecaster are two of the three modules of ESPNcricinfo Superstats, a cricket analytics platform jointly developed by Indian Institute of Technology Madras, Gyan Data Private Limited, and ESPNcricinfo. The predictions and analyses from Superstats were aimed at increasing the fan engagement on the ESPNcricinfo website.

3 Related Work

The International Cricket Council, the governing body of international cricket, regularly releases the official rankings of batsmen, bowlers, and all-rounders in all formats including international T20s [12, 13]. However, their ranking methodology has not been not revealed to the public probably because it is proprietary. Swartz provides a summary of the published literature on analysing batting and bowling performances apart from using traditional metrics like averages and strike rates [14]. Examples include combining multiple traditional metrics into one using Principal Component Analysis [15] and ways of analysing the form [16] and consistency [17, 18] of a batsman. Davis et al. propose a metric called "expected run differential" (ERD) [19]. ERD is the number of additional runs a player would contribute in a typical playing XI compared to an average player having the same role. Their calculation of ERD makes use of a T20 match simulator they had developed earlier for estimating from any state of an innings, the number of runs that will be scored by a team in the remainder of the innings. Thomson et al. analyse how batsmen and bowlers perform in the second innings with respect to the match situation, i.e., number of overs bowled, run required for the chasing team, and wickets fallen [20]. They devised metrics called "clutch batting" and "clutch bowling" for evaluating the batting and bowling performances based on match situation respectively. Shah and Shah [21], and Bhatacharjee and Lemmer [22] proposed two versions of a metric called "pressure index" to quantify pressure on a batsman in the second innings and evaluate how he performs under pressure. However, as pointed out by Thomson et al. [20], the proposed pressure indices assume a maximum value of 1 at the beginning of any run chase irrespective of the size of the target.

Jarrod Kimber, in his article for ESPNcricinfo, introduces "true strike rate" (TSR) and "true economy rate" (TER) [23]. He defines TSR of a player as the weighted average difference between the batting strike rate of the player and the typical batting strike rate of all other players in the overs he batted. Similarly, TER of a player is the weighted average difference between economy rate of the player and the typical economy rate of all other players in the overs he bowled. These metrics consider the phases in which a player usually bats/bowls and tell us the rate at which he scores/concedes runs relative to other players in those phases. So a player who bowls at an economy of 8 in a phase where the typical economy is 7.5 will have a lesser TER than a player who bowls at an economy of 8.5 in a phase where the typical economy rate is 9. This is a good starting point and there is scope for improving these metrics further. Our corresponding metrics - SSR and SER, will also account for the level of difficulty of the match situation for the player. So if there are two players who bowl in the same phases at similar economy rates, but one bowls in high-scoring and batsman-friendly conditions and the other bowls in low-scoring and bowler-friendly conditions, the former player will have a lower SER than the latter. Moreover, SSR and SER will reward players who perform well in difficult situations. So if there are two players who bowl in the same phases at the same economy rates in similar conditions, the one who bowls in more high pressure situations will have a better SER than the other.

4 Pressure

We first develop a metric that quantifies the pressure on the batsman and bowler at the time of facing/delivering a ball in a match. Pressure essentially denotes the level of difficulty of the match situation for the batsman and bowler. We shall discuss the desired properties of the pressure metric and see how our formulation satisfies those properties. The value of pressure may be different depending on whether it is calculated for the batsman or the bowler. But the formulation of pressure for both the batsman and bowler is the same. Pressure has been devised as a quantity that is bounded between 0 and 1. Pressure will be maximum (1) when the player's team is expected to lose the match or the match is expected to be a tie. This means that the pressure on a player will be 1 when the win probability for the player's team is lesser than or equal to 0.50. The win probability is obtained from the Forecaster. When the player's team is expected to win the match, i.e. his team's win probability is greater than 0.50, the pressure on the player will decrease linearly from 1 to 0 as the win probability increases. The pressure on a player will be minimum (0) when the win probability of his team is 1.00. Therefore, the pressure values on a player will be 0.75, 0.5, and 0.25 when the win probability values of his team are 0.625, 0.75, and 0.875 respectively.

There is another necessary characteristic for the pressure metric. In a significant number of matches, there will be situations where it would be certain which team will lose the match. Both the teams just play out the remaining overs from such situations as a mere formality. For examples, a team requiring 100 runs to win from just 2 overs or a team defending 10 runs but still having 10 more overs left to bowl. Though a team is expected to lose from such situations, a player from that team would not really feel any pressure to make an impact. There would not be any expectations from a player batting/bowling in such situations as his performance would have absolutely zero impact on the eventual result of the game. Thus, the pressure should be 0 even on a player whose team is certain to lose the match. In order to incorporate this characteristic in our pressure metric, we can set the pressure value on a player from the losing team as 0 if the win probability for that team is 0. But we do not want the pressure metric to instantaneously shift from 1 to 0 for minute win probability transitions such as, from 0.01 to 0. We rather prefer a steady linear decline in pressure from 1 to 0 as the win probability reduces from, say 0.10, to 0.

We will now summarize the above discussion as an equation. Let $\hat{p}$ be the win probability of a team. The pressure P on a player who plays for the team is calculated based on the following equation.

$$P = \begin{cases} \dfrac{\hat{p}}{0.10} & 0 \leq \hat{p} \leq 0.10 \\[2ex] 1 & 0.10 < \hat{p} \leq 0.50 \\[2ex] \dfrac{1-\hat{p}}{0.50} & 0.50 < \hat{p} \leq 1.00 \end{cases} \tag{1}$$

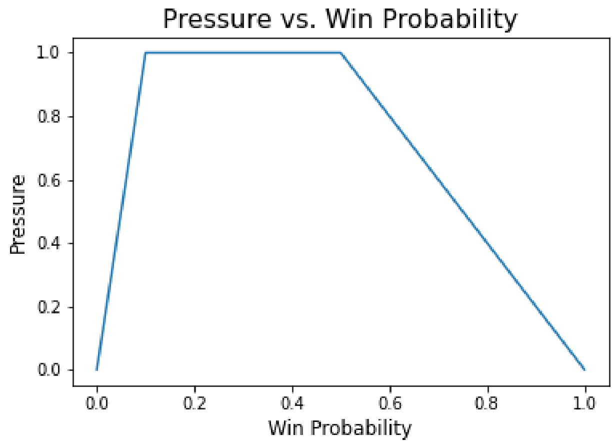

Figure 1: Pressure versus Win Probability Plot

5 Smart Runs

We have often heard commentators and experts describing a knock from a player as, "His 50 runs were probably worth 75 given the situation of the match when he scored those runs.". Similar statements can be extended to bowling spells of players too. A four-over spell wherein 24 runs were conceded might probably be worth only 12 runs overall. This is exactly what we aim to formulate in a structured manner. We will develop an algorithmic framework to convert the runs scored/conceded off any ball (r) in any given match to equivalent adjusted values for the batsman and bowler upon analysing those runs in the context of the game. This quantity will be referred to as smart runs and it will have a different formula depending on whether it is calculated for the batsman ($r_{\mathsf{s.bat}}$) or the bowler ($r_{\mathsf{s.bowl}}$).

$$r_{\mathsf{s.bat}} = r + \delta_{\mathsf{bat}} \tag{2}$$

$$r_{\mathsf{s.bowl}} = r + \delta_{\mathsf{bowl}} \tag{3}$$

For example, if 6 runs were scored/conceded off a particular ball in a match, the reward for the batsman and penalty for the bowler for that ball can be 4.5 and 4 runs respectively when analysed in the context of the game. This would be expressed as 10.5 smart runs for the batsman and 10 smart runs for the bowler. $r = 6$, $\delta_{\mathsf{bat}} = 4.5$, $\delta_{\mathsf{bowl}} = 4$, $r_{\mathsf{s.bat}} = 10.5$, and $r_{\mathsf{s.bowl}} = 10$ for this example. As another example, the penalty for the batsman and reward for the bowler for a dot ball (0 runs scored/conceded) can be 1.25 and 1.5 runs respectively when analysed in the context of the game. This would be expressed as -1.25 smart runs for the batsman and -1.5 smart runs for the bowler. $r = 0$, $\delta_{\mathsf{bat}} = -1.25$, $\delta_{\mathsf{bowl}} = -1.5$, $r_{\mathsf{s.bat}} = -1.25$, and $r_{\mathsf{s.bowl}} = -1.5$ for this example. In essence, a positive value of δ_{bat} would be a reward for the batsman and a positive value of δ_{bowl} would be a penalty for the bowler, and vice versa.

δ_{bat} and δ_{bowl} are the quantities of interest here as they are the reward/penalty values for the runs scored by a batsman and conceded by a bowler respectively. In order to determine whether to reward/penalise the runs scored/conceded off a ball, we need a reference or baseline value (r_0), i.e., how many runs are "par" for that particular ball. r_0 is the same for both the batsman and bowler. To obtain this value, we will use historical ball-by-ball data to find out how many runs are typically scored/conceded off a ball delivered in a similar match situation. A situation is characterized by two input variables - total number of balls remaining to be bowled in the innings at the state (BR) and the combined run rate of all batsmen in the entire match (RR). Intuitively, we expect r_0 to vary as per the following trends with respect to each of the two input variables (provided the other input variable is held constant).

1. r_0 should increase monotonically as BR decreases. When fewer balls are left in an innings, batsmen typically look to score as quickly as possible thereby resulting in a higher scoring rate.

2. r_0 should increase monotonically as RR increases. When the overall scoring rate in the match is higher, it means that the conditions are relatively more batting friendly and so the expected scoring rate is higher.

We shall now see how to model r_0. Based on ball-by-ball data in the three-year period prior to a match of interest, we prepare a data set with BR as the input variable and the number of runs scored off a ball as the output variable. We fit a random forest regression model $\hat{f}(.)$ to this data set. This model essentially predicts the average number of runs scored off a ball given the number of balls left in the innings as shown in Figure 2. We then scale this prediction using the RR value of the match we are interested in for obtaining r_0.

$$r_0 = \hat{f}(\mathsf{BR}) \times \frac{\mathsf{RR}}{1.30} \tag{4}$$

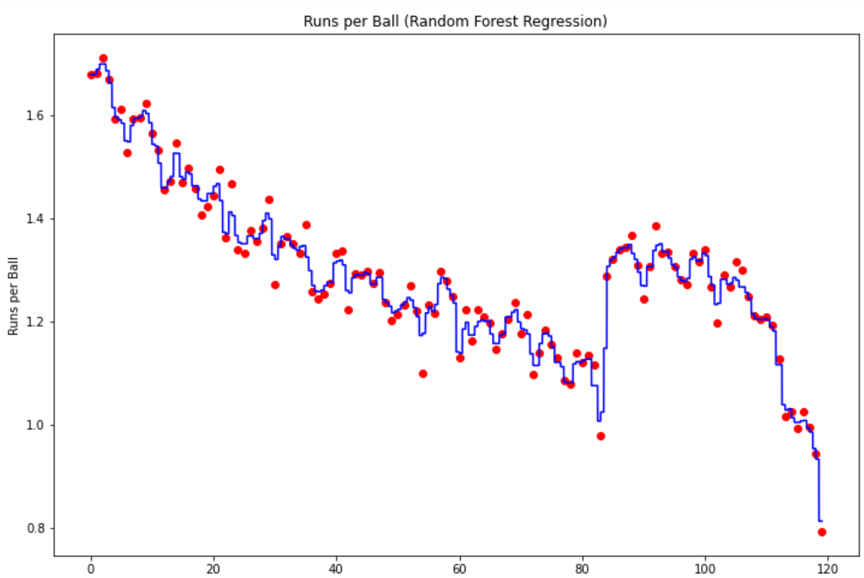

Figure 2: Random Forest Model Plot Runs per Ball vs BB

As shown in Figure 2, we see jump in the Runs per Balls value when Ball Remaining (BR) is 84 which is the end of power play (i.e. end of 6th over). As BR decreases the Runs per Ball increases as batsmen tends to take more risk when they have less balls remaining.

In equation 4, $\hat{f}(.)$ is the prediction of the random forest model. 1.30 is the overall average number of runs scored per ball in the data set we have considered. Hence, we scale $\hat{f}(.)$ by multiplying it with the ratio of RR to 1.30. δ_{bat} and δ_{bowl} are directly proportional to the difference between r and r_0. However, it is not enough to just analyse how a batsman/bowler scores/concedes with respect to the par value. Even if a batsman scores well above par or a bowler concedes well below par, impact would be insignificant if there is little or no pressure on the player or his team. On the other hand, a batsman scoring well above par or a bowler conceding well below par consistently in high pressure situations would be highly valuable. Therefore, it is clear that δ_{bat} and δ_{bowl} should also be proportional to pressure.

$$\delta_{\mathsf{bat}} = P^*_{\mathsf{bat}} \times (r - r_0) \tag{5}$$

$$\delta_{\mathsf{bowl}} = P^*_{\mathsf{bowl}} \times (r - r_0) \tag{6}$$

P^*_{bat} in equation 5 will be the maximum among the pressure values on the batsman or any of his teammates starting from the ball for which we are computing smart runs till the end of the match. The reason why we are not just using the pressure value on the batsman when the ball was delivered can be understood from the following example. Consider a team chasing a modest total of 120. The pressure on the opening batsmen would be low at the start of the innings due to the small target. The openers however get out after batting slowly. They scored well below par in aggregate and their team is in a tricky situation wherein the required run rate is say, 9 per over with 10 overs still left. The pressure on the player coming out to bat in such a situation would be much higher than what it was on the openers at the start of the innings. The openers caused the rise in pressure and so, the penalty on them should be proportional to the maximum value of pressure that got built up.

Similarly, P^*_{bowl} in equation 6 will be the maximum among the pressure values on the bowler or any of his teammates starting from the ball for which we are computing smart runs till the end of the match. The following example will illustrate the reason behind the formulation. Consider a team defending a large total of 240. The pressure on the bowling team would be low at the start of the innings. However, pressure on the bowling team will increase to 1 if the opening bowlers concede runs excessively, for example, they let the opposition score 100 runs in the first five overs itself without taking any wickets. The opening bowlers should therefore be penalised in proportion to the maximum pressure value that got built up. Thus it is now clear why we should not use just the pressure values on the batsman and bowler during the ball for which we are computing smart runs.

6 Smart Wickets

As we have discussed in section 2, we need a new methodology for measuring the value of a wicket. Conventionally, any wicket is valued at 1 irrespective of the quality of the batsman dismissed and significance of the wicket in the match. This does not represent the true worth of a wicket as we have seen from the Tahir-Sharma example in section 2. Therefore, analogous to smart runs, we propose a novel metric called smart wickets for quantifying the value of a dismissal better. The smart wicket value of a dismissal is formulated such that it is always bounded between 0 and 2. Intuitively, the value of a wicket should increase with the quality of the batsman dismissed. However, we cannot just formulate the smart wicket value as just a monotonically increasing function with respect to batsman quality. This is because, even wickets of poorly rated batsmen would be valuable for the bowling team if they are taken in a high pressure situation. For example, consider a chasing team that is 8 wickets down and requiring only 10 runs to win with enough number of overs left. In such a situation, even the wickets of two tailenders would be of great value as they would win the match for the bowling team. Therefore, pressure on the bowling team should also be a component of smart wickets. Let us first see how to quantify the quality of a batsman.

6.1 Batsman Rating

To value a wicket based on the quality of the batsman dismissed, we need a metric to evaluate a batsman based on his past performances. We call this metric as "batsman rating". We will be considering data from all T20 matches three years prior to the match for which we want to compute smart wicket values. So all the batting performances of a player relative to the other players in the three-year period prior to the match would determine his rating. It is important for a player to do well in both batting average and strike rate metrics, especially in T20s. Commentators have recently started to use a metric referred to as "BASRA" (Batting Average Strike Rate Aggregate) for evaluating batsmen in T20s. BASRA is just the sum of batting average and strike rate of a batsman. However, we are not allowed to simply add those two quantities as their units are different. The unit for batting average is runs per dismissal and the unit for batting strike rate is runs per 100 balls. Therefore, we have developed a metric that combines both batting average and strike rate in a mathematically sound manner. Also, we will be using the average and strike rate based on smart runs as we know it is a better representation of a player's batting performance than traditional runs. It is important to note that the rating value of a batsman not just depends on his own smart average and strike rate, but also the smart averages and strike rates of all the players who have batted in the three-year period prior to the match. The following is the outline of the rating computation for a batsman.

1. Retrieve the number of innings batted and the total number of balls faced by each player.

2. Compute the smart batting average of each of the players if the player has batted in at least 15 innings. For those who have not batted in at least 15 innings yet, we will not be considering their actual smart average as their sample size is small. For example, a player might have batted in just 1 innings and scored 100 smart runs before getting out. His actual smart average would be 100. Such a high smart average value is unlikely to be maintained over a significant number of games. We will modify the smart average of such players before proceeding further. The modification logic will be discussed later in this section.

3. Compute the smart batting strike rate of each of the players using the standard formula if the player has faced at least 200 balls. Similar to smart average, we will modify the smart strike rate of those who have batted lesser than 200 balls before proceeding further to counter the problem of small sample size.

4. Once we have the smart averages of all players, find the maximum and minimum smart average values among them. Use these values to scale the smart average of every player between 0 and 1.

$$\mathsf{SA}_{\mathsf{sc}} = \frac{\mathsf{SA} - \mathsf{SA}_{\mathsf{min}}}{\mathsf{SA}_{\mathsf{max}} - \mathsf{SA}_{\mathsf{min}}} \tag{7}$$

5. Similarly scale the smart strike rate value of all players between 0 and 1.

$$\mathsf{SSR}_{\mathsf{sc}} = \frac{\mathsf{SSR} - \mathsf{SSR}_{\mathsf{min}}}{\mathsf{SSR}_{\mathsf{max}} - \mathsf{SSR}_{\mathsf{min}}} \tag{8}$$

6. Now that we have normalized smart average and strike rate, we are allowed to add them. We then rescale the sum for each player between 0 and 1 to obtain the final batsman rating.

$$R_0 = \mathsf{SA}_{\mathsf{sc}} + \mathsf{SSR}_{\mathsf{sc}} \tag{9}$$

$$R = \frac{R_0 - R_{0.\mathsf{min}}}{R_{0.\mathsf{max}} - R_{0.\mathsf{min}}} \tag{10}$$

Before we look at the modification logic for smart average and smart strike rate for those who do not clear the corresponding thresholds, we define the three batting roles - top-order, middle-order, and lower-order as follows.

1. Top-order: Batting positions 1 to 4
2. Middle-order: Batting positions 5 to 7
3. Lower-order: Batting positions 8 to 11

6.1.1 Smart Average Modification

The following quantities of a player who has batted in lesser than 15 innings are first retrieved.

- SR : total number of smart runs scored
- ND : total number of dismissals
- NI_{T} : total number of innings played as a top-order batsman
- NI_{M} : total number of innings played as a middle-order batsman
- NI_{L} : total number of innings played as a lower-order batsman

Let NI be the total number of innings played, i.e., $\mathsf{NI}_{\mathsf{T}} + \mathsf{NI}_{\mathsf{M}} + \mathsf{NI}_{\mathsf{L}}$. The following "default" quantities are computed for the batting roles.

- **Default top-order smart average**

$$\mathsf{SA}^*_{\mathsf{T}} = \frac{\text{Total smart runs scored by all top-order batsmen}}{\text{Total dismissals of all top-order batsmen}} \tag{11}$$

- **Default middle-order smart average**

$$\mathsf{SA}^*_{\mathsf{M}} = \frac{\text{Total smart runs scored by all middle-order batsmen}}{\text{Total dismissals of all middle-order batsmen}} \tag{12}$$

- **Default lower-order smart average**

$$\mathsf{SA}^*_{\mathsf{L}} = \frac{\text{Total smart runs scored by all lower-order batsmen}}{\text{Total dismissals of all lower-order batsmen}} \tag{13}$$

- **Default top-order dismissal rate**

$$\mathsf{DI}^*_{\mathsf{T}} = \frac{\text{Total dismissals of all top-order batsmen}}{\text{Total innings of all top-order batsmen}} \tag{14}$$

- **Default middle-order dismissal rate**

$$\mathsf{DI}_\mathsf{M}^* = \frac{\text{Total dismissals of all middle-order batsmen}}{\text{Total innings of all middle-order batsmen}} \tag{15}$$

- **Default lower-order dismissal rate**

$$\mathsf{DI}_\mathsf{L}^* = \frac{\text{Total dismissals of all lower-order batsmen}}{\text{Total innings of all lower-order batsmen}} \tag{16}$$

We want smart average of a player to be a quantity that is computed over 15 innings at least. So for those who have batted lesser than 15 innings, we will impute a weighted default smart average and dismissal rate for the deficit number of innings (15 - NI). Let us denote the deficit number of innings as $\mathsf{NI_d}$. The weighted default smart average ($\mathsf{SA_0}$) and dismissal rate ($\mathsf{DI_0}$) is based on the batting role(s) assumed by the player in the innings he has batted so far.

$$\mathsf{SA_0} = \frac{\mathsf{NI_T} \times \mathsf{SA_T^*} + \mathsf{NI_M} \times \mathsf{SA_M^*} + \mathsf{NI_L} \times \mathsf{SA_L^*}}{\mathsf{NI}} \tag{17}$$

$$\mathsf{DI_0} = \frac{\mathsf{NI_T} \times \mathsf{DI_T^*} + \mathsf{NI_M} \times \mathsf{DI_M^*} + \mathsf{NI_L} \times \mathsf{DI_L^*}}{\mathsf{NI}} \tag{18}$$

$$\mathsf{SA} = \frac{\mathsf{SR} + \mathsf{NI_d} \times \mathsf{SA_0}}{\mathsf{ND} + \mathsf{NI_d} \times \mathsf{DI_0}} \tag{19}$$

6.1.2 Smart Strike Rate Modification

A similar approach would be used to modify smart strike rate which depends on the number of balls faced rather than innings batted. The following quantities of the player are first retrieved.

- SR : total number of smart runs scored
- $\mathsf{BF_T}$: total number of balls faced as a top-order batsman
- $\mathsf{BF_M}$: total number of balls faced as a middle-order batsman
- $\mathsf{BF_L}$: total number of balls faced as a lower-order batsman

Let BF be the total number of balls faced, i.e., $\mathsf{BF_T} + \mathsf{BF_M} + \mathsf{BF_L}$. The following "default" quantities are computed for the batting roles.

- **Default top-order smart strike rate** (per ball)

$$\mathsf{SSR_T^*} = \frac{\text{Total smart runs scored by all top-order batsmen}}{\text{Total balls faced by all top-order batsmen}} \tag{20}$$

- **Default middle-order smart strike rate** (per ball)

$$\mathsf{SSR_M^*} = \frac{\text{Total smart runs scored by all middle-order batsmen}}{\text{Total balls faced by all middle-order batsmen}} \tag{21}$$

- **Default lower-order smart strike rate** (per ball)

$$\mathsf{SSR_L^*} = \frac{\text{Total smart runs scored by all lower-order batsmen}}{\text{Total balls faced by all lower-order batsmen}} \tag{22}$$

Similar to section 6.1.1, we will use a weighted default smart strike rate value to impute for the deficit number of balls faced (200 - BF). Let us denote the deficit number of balls faced as $\mathsf{BF_d}$. The weighted default smart strike rate ($\mathsf{SSR_0}$) is based on the batting role(s) assumed by the player in the innings he has batted so far.

$$\mathsf{SSR_0} = \frac{\mathsf{BF_T} \times \mathsf{SSR^*_T} + \mathsf{BF_M} \times \mathsf{SSR^*_M} + \mathsf{BF_L} \times \mathsf{SSR^*_L}}{\mathsf{BF}} \tag{23}$$

$$\mathsf{SSR} = 100 \times \frac{\mathsf{SR} + \mathsf{BF_d} \times \mathsf{SSR_0}}{200} \tag{24}$$

6.2 Smart Wickets Formula

As mentioned earlier, the smart wicket value of a dismissal depends on two factors - the quality of the batsman dismissed and the pressure on the bowling team. The quality of the batsman dismissed is represented by the batsman rating metric (R) we discussed in section 6.1 and pressure (P) was discussed in section 4. Both quantities are bounded between 0 and 1, and therefore, smart wicket value of a wicket will be bounded between 0 and 2. We will be adjusting the rating value based on the timing of the wicket. Wickets taken when there is a significant number of balls left (at least 3 overs, i.e., 18 balls) in an innings would be a lot more valuable than those taken towards the end of the innings. The potential runs saved by a dismissal in the last over of an innings would be insignificant. Therefore, we will "dampen" the batsman rating value depending on the number of balls remaining (BR) in the innings when the batsman was dismissed. The adjusted batsman rating R' of a batsman having a rating value of R is obtained using the following equation.

$$R' = \begin{cases} R \times \dfrac{\mathsf{BR}}{18} & 0 \leq \mathsf{BR} < 18 \\ \\ R & 18 \leq \mathsf{BR} < 120 \end{cases} \tag{25}$$

To account for pressure, we will use the maximum among the pressure values on the bowler or any of his teammates starting from the ball of dismissal till the end of the match. We will denote this pressure value as P^*. This is similar to the way we incorporated pressure for computing smart runs. The reason why we use the maximum pressure starting from the state till the end of the match can be understood from an example scenario. Consider a team defending a high total of 240. The pressure on the bowling team would be extremely low at the start of the innings because they seem to have enough runs to defend. Let us say a bowler picked up a couple of wickets in the first over itself when the pressure was low. But the chasing team eventually caught up with the required run rate and the match was back in the balance. The pressure on the bowling team would have risen to a relatively higher value by this time. Therefore, in hindsight, the wickets that were taken in the first over should be valued highly as the match would have been tougher for the bowling team had those wickets not been taken. The formula for computing the smart wicket value (w_s) of a dismissal is shown below.

$$w_s = R' + P^* \tag{26}$$

7 Impact Score

Impact score is a metric that we have developed for quantifying both batting and bowling performances together. This metric is proposed as an alternative to fantasy points which also intends to evaluate batting and bowling performances on the same scale. However, all the existing fantasy points scoring methodologies rely on conventional runs and wickets. Now that we have better representation of runs and wickets in the form of smart runs and smart wickets respectively, we can use them to create an improved fantasy points scoring methodology. Impact score (IS) is the sum of two components - batting impact score ($\mathsf{IS}_{\mathsf{bat}}$) and bowling impact score ($\mathsf{IS}_{\mathsf{bowl}}$).

$$\mathsf{IS} = \mathsf{IS}_{\mathsf{bat}} + \mathsf{IS}_{\mathsf{bowl}} \tag{27}$$

7.1 Batting Impact Score ($\mathsf{IS}_{\mathsf{bat}}$)

We will use the total smart runs scored by a player in a match (SR) as his batting impact score for that match.

$$\mathsf{IS}_{\mathsf{bat}} = \mathsf{SR} \tag{28}$$

7.2 Bowling Impact Score ($\mathsf{IS}_{\mathsf{bowl}}$)

Bowling impact score is a bit more involved compared to batting impact score. A bowler can gain points both by bowling economically as well as taking wickets. If the bowler has conceded fewer smart runs in the match with respect to par in aggregate, he should get more points. Bowling impact score should also increase with the total number of smart wickets a bowler scalped in a match. The following is the equation for bowling impact score.

$$\mathsf{IS}_{\mathsf{bowl}} = \left(\sum r_0 - \mathsf{SR}\right) + 25 \times \mathsf{SW} \tag{29}$$

In the above equation, $\sum r_0$ is the sum total of the par value of runs for every ball bowled, SR is the total number of smart runs conceded, and SW is the total number of smart wickets taken by the player in the match. A multiplier of 25 has been chosen for smart wickets such that the aggregate of positive batting impact scores would be equal to the aggregate of positive bowling impact scores of all players across all matches.

8 Smart Stats of IPL

8.1 Rating Table

8.1.1 Top 20 Batsman by R value Table for batsmen who played at least 30 innings

R value is relative rating score of a player with respect to others, the more the better, but some players have played very less number of matches and runs and had very high Smart Strike Rate and Smart Batting Average, hence for calculating the top 20 batsmen list we have applied a filter of number of innings greater than or equal to 30.

	Rank	batsmanName	key	dismissals	balls	innings	Rnominal	R
0	1	AB de Villiers	21	67	1828	84	0.648358862	0.746289099
1	2	Devon Conway	637	28	1079	34	0.56978684	0.735107291
2	3	Lokesh Rahul	177	62	2099	71	0.569985707	0.695758714
3	4	Cam Fletcher	3069	15	549	32	0.565391621	0.669464851
4	5	Andre Russell	90	70	1267	95	0.724230466	0.661386896
5	6	Laurie Evans	924	46	1518	65	0.554413702	0.641921139
6	7	Luke Ronchi	532	50	966	54	0.637267081	0.636642164
7	8	Babar Azam	426	74	2603	88	0.529389166	0.615572001
8	9	Aaron Finch	219	76	1972	92	0.598174442	0.615357477
9	10	Matthew Wade	390	35	844	37	0.593364929	0.60796958
10	11	Moeen Ali	12	38	768	48	0.651041667	0.606739661
11	12	MS Dhoni	23	29	1061	54	0.545523091	0.596890562
12	13	Rahmanullah Gurbaz	357	32	544	31	0.638970588	0.59091686
13	14	Joe Clarke	868	30	649	33	0.597842835	0.590835272
14	15	D'Arcy Short	282	62	1932	68	0.546376812	0.590003242
15	16	Jos Buttler	188	56	1493	62	0.582987274	0.589499388
16	17	Jonny Bairstow	54	38	941	41	0.592136026	0.585528285
17	18	David Warner	72	56	1738	69	0.535558113	0.580631326
18	19	Kieron Pollard	117	84	1840	118	0.622173913	0.580535546
19	20	Hardik Pandya	118	36	729	57	0.662825789	0.57778608

8.1.2 Top 20 Bowler by Smart Bowling Average Table for bowlers who played at least 30 innings

AVG: Bowling Average
innings: No of innings the bowller got to bowl at least bowl.

	Rank	Name	Key	SmartWicket	wicket	innings	smartAVG	AVG
0	1	Lungi Ngidi	49	82.87741053	65	36	11.49477655	15.10769231
1	2	Jasprit Bumrah	119	110.664412	91	65	11.66206903	17.9010989
2	3	Lockie Ferguson	96	60.5040188	47	32	11.97776514	18
3	4	Rashid Khan	63	231.4816131	187	134	12.38035107	17.61497326
4	5	Jofra Archer	178	167.0413688	131	91	13.82038071	19.23664122
5	6	Imran Tahir	36	198.2749758	166	116	13.94908564	17.46987952
6	7	Peter Siddle	687	64.51075459	52	41	14.05866253	18.96153846
7	8	Mustafizur Rahman	248	90.87265769	84	55	14.06177332	16.54761905
8	9	Blair Tickner	499	77.99906139	61	38	14.288034	18.52459016
9	10	Shaheen Afridi	427	114.6570012	92	61	14.43292792	18.18478261
10	11	Harry Gurney	55	96.87253992	83	55	14.48581373	17.80722892
11	12	Tabraiz Shamsi	397	100.811206	75	59	14.85975361	20.77333333
12	13	Dale Steyn	34	82.21909314	60	45	14.95182379	20.33333333
13	14	Fawad Ahmed	681	152.0129969	123	95	15.13448696	19.13821138
14	15	Andrew Tye	179	149.0733338	127	80	15.16294803	18.78740157
15	16	Wahab Riaz	567	138.5350128	130	93	15.31481337	18.6
16	17	Sandeep Lamichhane	120	108.5793248	89	71	15.39044646	19.48314607
17	18	Mujeeb Ur Rahman	193	122.3445674	101	88	15.78290385	21.10891089
18	19	Haris Rauf	415	78.46923455	72	43	15.89984962	17.40277778
19	20	Rahul Chahar	123	50.79462443	37	31	16.0593618	21.64864865

8.1.3 Top 20 Bowler by Economy Rate Table for bowlers who played at least 30 innings

	Rank	Name	Key	innings	smartRuns	Runs	SmartER	ER
0	1	Jasprit Bumrah	119	65	1290.576012	1629	5.189983961	6.550938338
1	2	Rashid Khan	63	134	2865.823636	3294	5.50590516	6.328530259
2	3	Mujeeb Ur Rahman	193	88	1930.952544	2132	5.684845567	6.276741904
3	4	Sunil Narine	93	101	2218.752804	2518	5.760500574	6.537429684
4	5	Lockie Ferguson	96	32	724.7029273	846	5.948314041	6.943912449
5	6	Wahab Riaz	567	93	2121.637866	2418	6.243171748	7.115252575
6	7	Keshav Maharaj	510	31	726.8077463	728	6.283640458	6.293948127
7	8	Imran Tahir	36	116	2765.754617	2900	6.331372644	6.638687524
8	9	Chris Green	238	76	1613.575384	1646	6.394618431	6.523117569
9	10	Mustafizur Rahman	248	55	1277.830714	1390	6.426642315	6.990779547
10	11	Max Waller	1057	37	865.4442097	911	6.466581891	6.806973848
11	12	Shahid Afridi	533	57	1238.15601	1273	6.499506616	6.682414698
12	13	Mohammad Amir	569	71	1720.554674	1865	6.5008363	7.046599496
13	14	Fawad Ahmed	681	95	2300.638718	2354	6.529722001	6.681173132
14	15	T Natarajan	59	31	786.293668	828	6.561560512	6.909596662
15	16	Jofra Archer	178	91	2308.57531	2520	6.564669128	7.165876777
16	17	Sandeep Lamichhane	120	71	1671.084286	1734	6.574757845	6.822295082
17	18	Peter Siddle	687	41	906.9349286	986	6.59589039	7.170909091
18	19	Imad Wasim	422	76	1804.972827	1812	6.67272764	6.6987061
19	20	Murugan Ashwin	189	40	976.8236585	996	6.713564663	6.845360825

ER: Economy Rate
SmartRuns: Smart Runs Conceded
Runs: Runs conceded
innings: No of innings the bowller got to bowl at least bowl.

8.2 Results of Selected Players

8.2.1 R -value (Rating Score) of Selected Batsmen

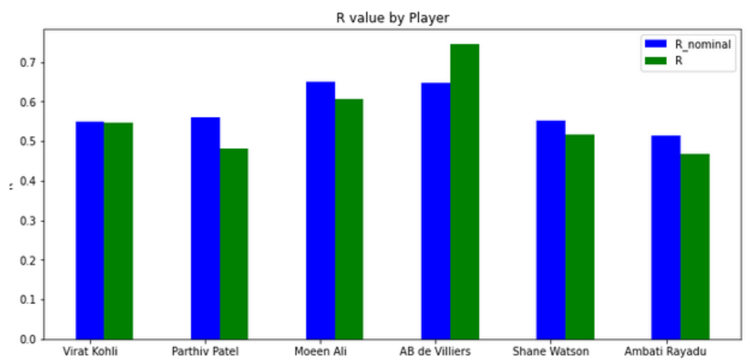

Figure 3: R value comparison 1

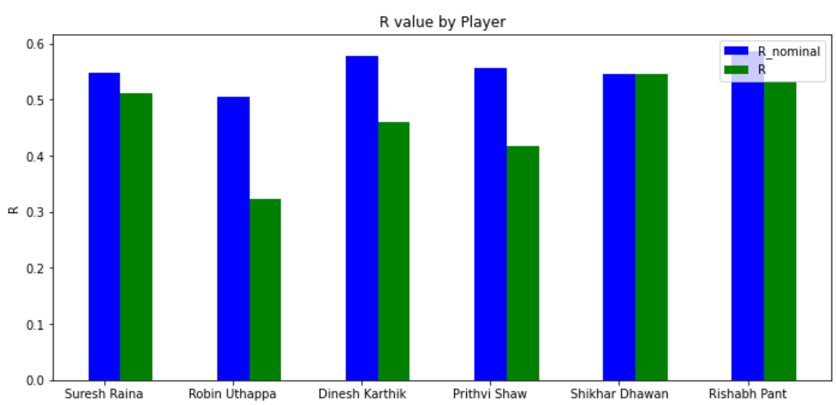

Figure 4: R value comparison 2

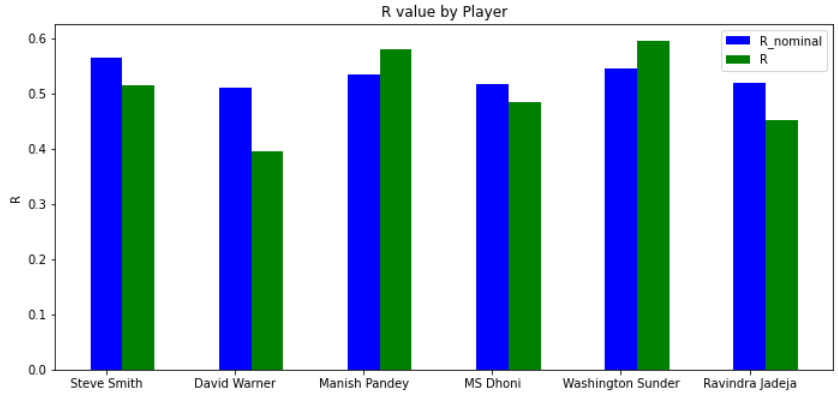

Figure 5: R value comparison 3

8.2.2 Batting Average Smart vs Nominal comparison of Selected Batsmen

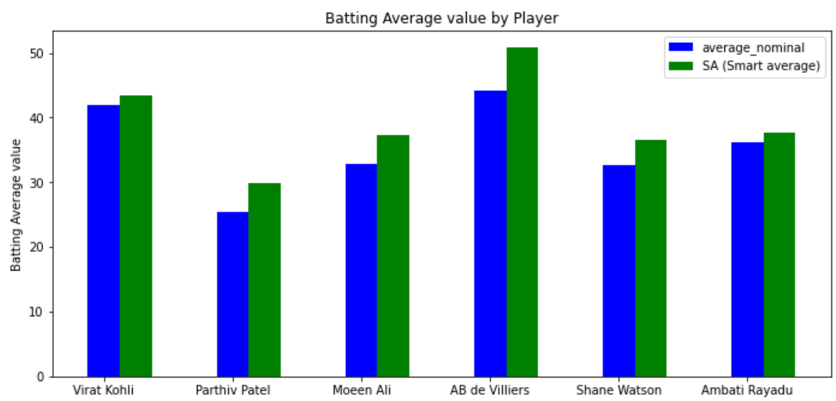

Figure 6: Batting Average Smart vs Nominal comparison 1

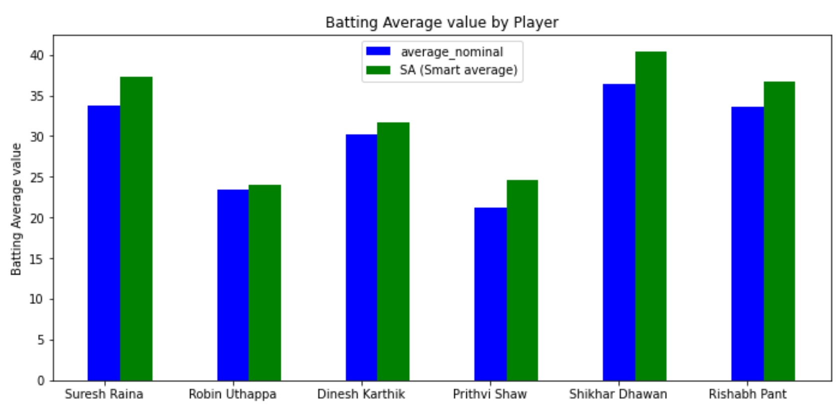

Figure 7: Batting Average Smart vs Nominal comparison 2

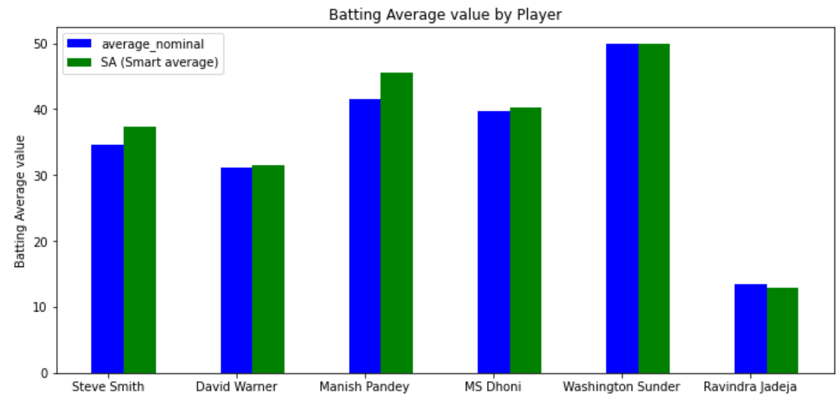

Figure 8: Batting Average Smart vs Nominal comparison 3

8.2.3 Batting Strike Rate Smart vs Nominal comparison of Selected Batsmen

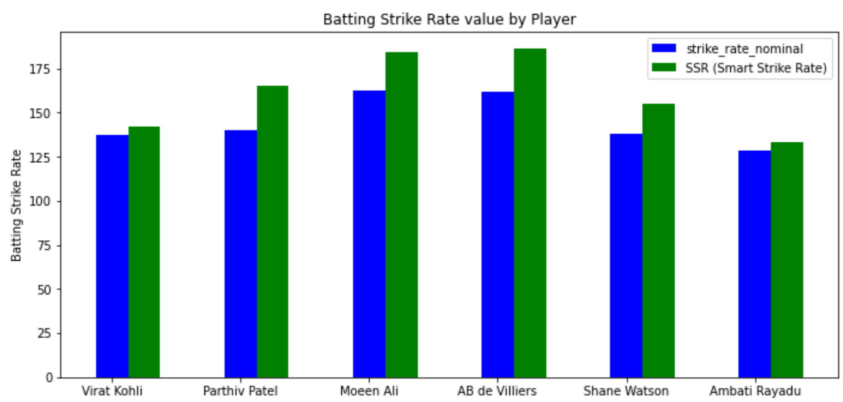

Figure 9: Batting Strike Rate Smart vs Nominal comparison 1

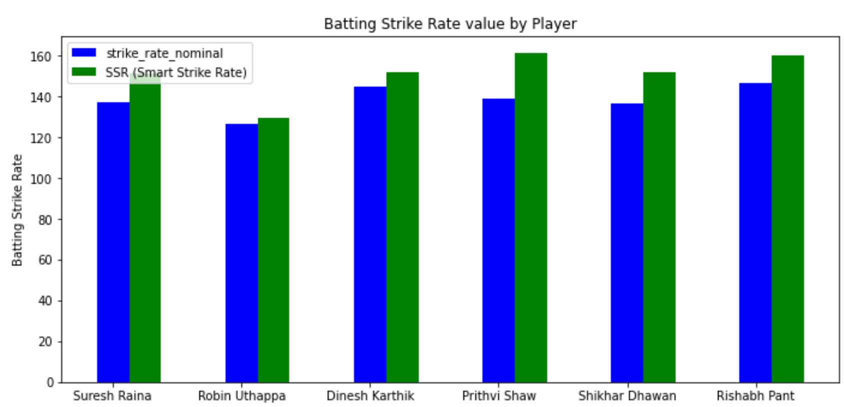

Figure 10: Batting Strike Rate Smart vs Nominal comparison 2

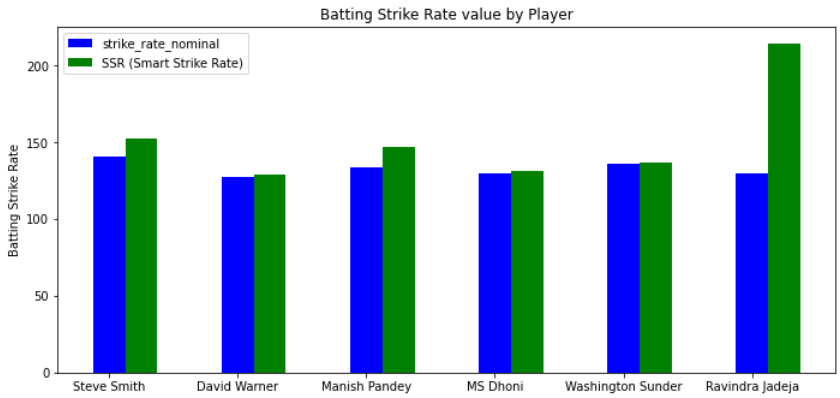

Figure 11: Batting Strike Rate Smart vs Nominal comparison 3

8.2.4 Smart Economy Rate vs Nominal Economy Rate comparison of Selected Bowlers

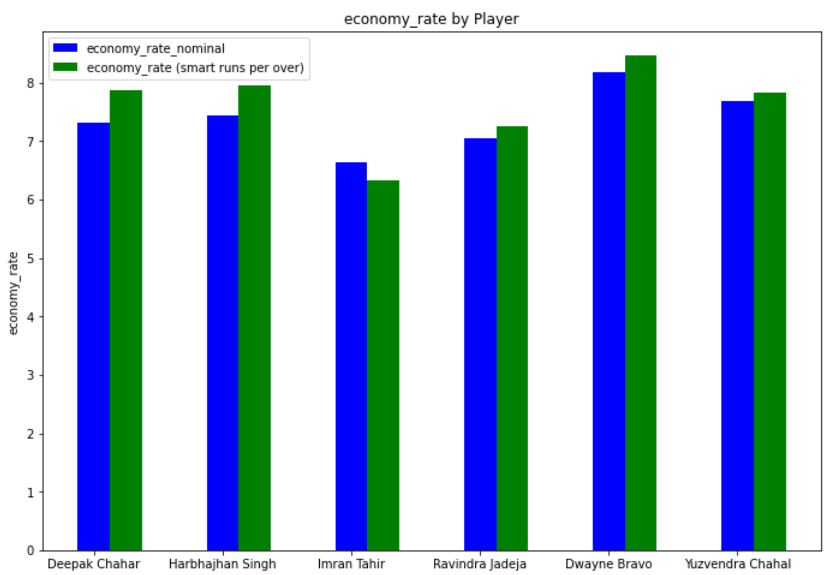

Figure 12: economy rate 1

8.2.5 Smart Bowling Average vs Nominal Bowling Average comparison of Selected Bowlers

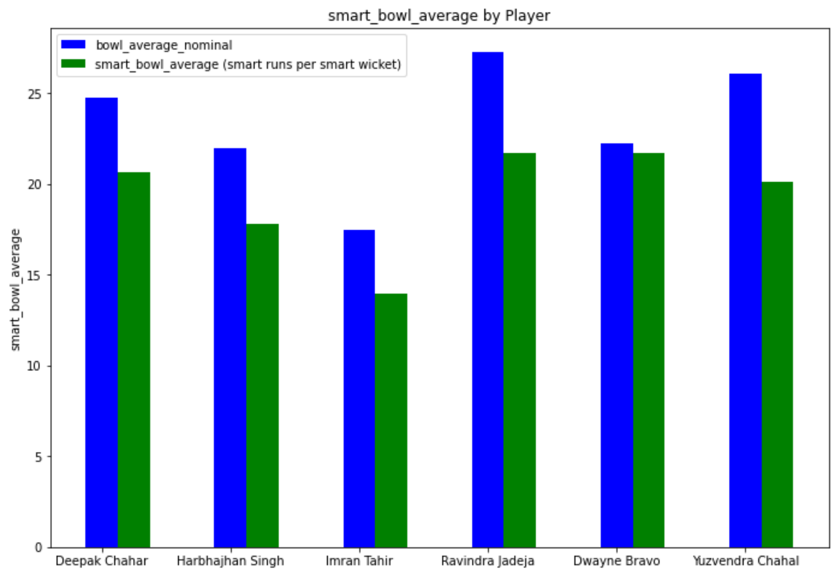

Figure 13: Bowler average 2

8.3 Match Ball by Ball Events

Match Ball by Ball Events for 1st Innings:-

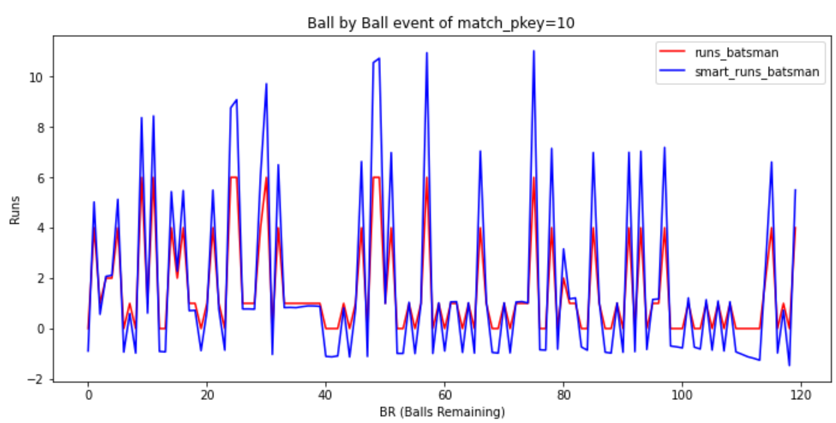

Figure 14: Match key= 10

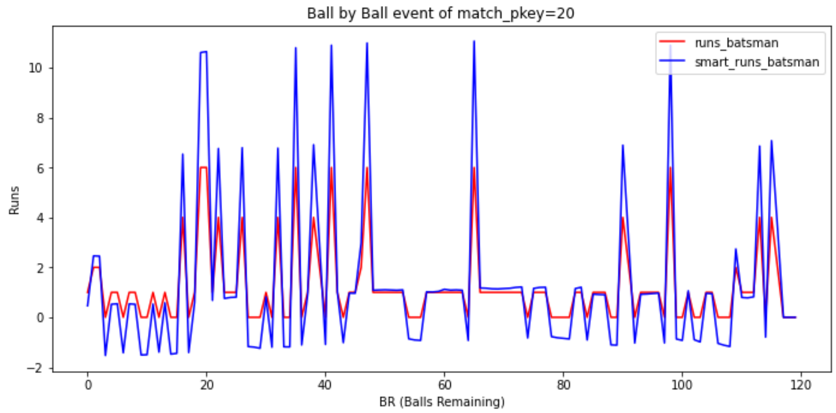

Figure 15: Match key=20

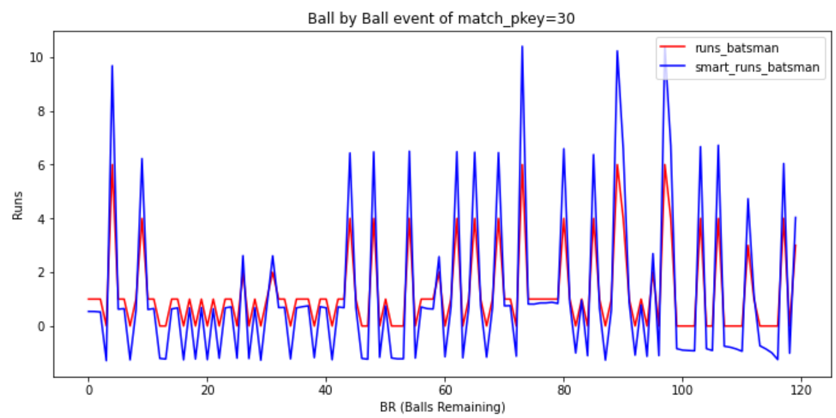

Figure 16: Match key=30

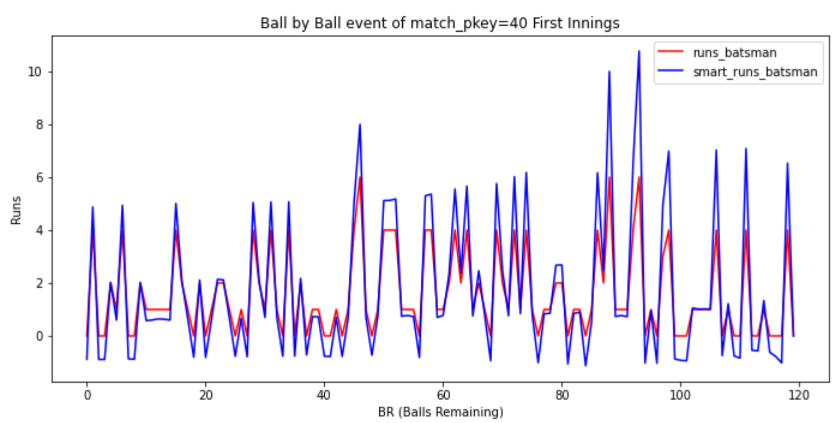

Figure 17: Match key=40

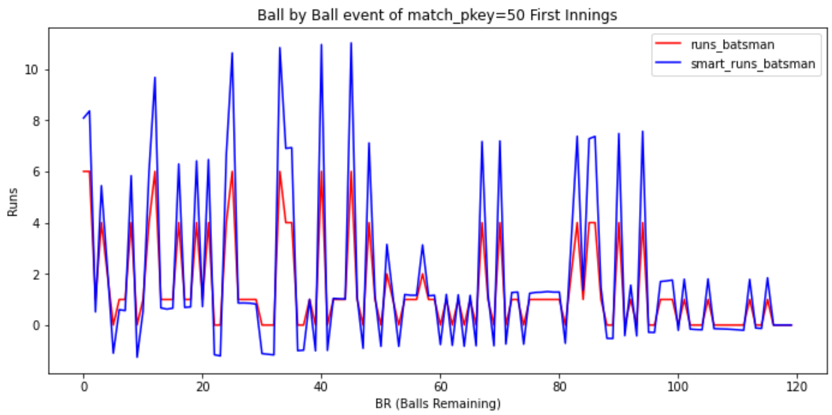

Figure 18: Match key=50

9 Conclusion

New Method for evaluating player performance taking into account the context (based on the idea of pressure) is proposed for Cricket. Smart Strike Rate, Smart Batting Average, Smart bowling Average and Smart Economy Rate is calculated for each player.

References

[1] https://en.wikipedia.org/wiki/Cricket

[2] https://en.wikipedia.org/wiki/Cricket_statistics

[3] https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html

[4] https://www.espncricinfo.com/series/ipl-2020-21-1210595/rajasthan-royals-vs-royal-challengers-bangalore-33rd-match-1216522/full-scorecard

[5] https://www.espncricinfo.com/series/ipl-2020-21-1210595/kings-xi-punjab-vs-sunrisers-hyderabad-43rd-match-1216498/full-scorecard

[6] https://www.espncricinfo.com/series/ipl-2020-21-1210595/rajasthan-royals-vs-chennai-super-kings-4th-match-1216496/full-scorecard

[7] https://www.espncricinfo.com/series/ipl-2017-1078425/delhi-daredevils-vs-sunrisers-hyderabad-40th-match-1082630/full-scorecard

[8] https://www.espncricinfo.com/series/ipl-2017-1078425/rising-pune-supergiant-vs-royal-challengers-bangalore-34th-match-1082624/full-scorecard

[9] https://www.espncricinfo.com/series/ipl-2017-1078425/royal-challengers-bangalore-vs-kings-xi-punjab-43rd-match-1082633/full-scorecard

[10] https://www.espncricinfo.com/series/ipl-2020-21-1210595/kings-xi-punjab-vs-rajasthan-royals-50th-match-1216537/full-scorecard

[11] https://www.espncricinfo.com/series/ipl-2017-1078425/mumbai-indians-vs-rising-pune-supergiant-final-1082650/full-scorecard

[12] https://www.icc-cricket.com/rankings/mens/player-rankings/t20i

[13] https://www.icc-cricket.com/rankings/womens/player-rankings/t20i

[14] Swartz, Tim B. , "Research Directions in Cricket" , in Handbook of Statistical Methods and Analyses in Sports ed. Jim Albert , Mark E. Glickman , Tim B. Swartz and Ruud H. Koning (Boca Raton: CRC Press, 21 Dec 2016), accessed 23 Jul 2021 , Routledge Handbooks Online.

[15] Ananda B. W. Manage & Stephen M. Scariano (2013) An Introductory Application of Principal Components to Cricket Data, Journal of Statistics Education, 21:3, DOI: 10.1080/10691898.2013.11889689

[16] Koulis, Theodoro, Muthukumarana, Saman and Briercliffe, Creagh Dyson. "A Bayesian stochastic model for batting performance evaluation in one-day cricket" Journal of Quantitative Analysis in Sports, vol. 10, no. 1, 2014, pp. 1-13. https://doi.org/10.1515/jqas-2013-0057

[17] Lemmer, H.H. (2004). A measure for the batting performance of cricket players. South African Journal for Research in Sport, Physical Education and Recreation, 26, 55-64.

[18] Borooah, Vani K and Mangan, John E. "The "Bradman Class": An Exploration of Some Issues in the Evaluation of Batsmen for Test Matches, 1877-2006" Journal of Quantitative Analysis in Sports, vol. 6, no. 3, 2010. https://doi.org/10.2202/1559-0410.1201

[19] Davis, J., Perera, H. and Swartz, T.B. (2015b). Player evaluation in Twenty20 cricket. Journal of Sports Analytics, 1, 19-31.

[20] https://www.sfu.ca/ tswartz/papers/contextual.pdf

[21] Shah, P. and Shah, M. (2014). Pressure index in cricket. Journal of Sports and Physical Education, 1, 9-11.

[22] Bhattacharjee, D. and Lemmer, H.H. (2016). Quantifying the pressure on the teams batting or bowling in the second innings of limited overs cricket matches. International Journal of Sports Science & Coaching, 11, 683-692.

[23] https://www.espncricinfo.com/story/jarrod-kimber-the-new-vocabulary-of-t20-1133997