Did Batsmen perform the best in the latest edition of IPLT20

A bottom-up rating approach to analyse the players and rate them for IPLT20 – 2018

Saurabh Mallik | S3623575

Table of Contents

ABSTRACT 3
LITERATURE REVIEW 3
INTRODUCTION 4
OBJECTIVES 4
DATASET 5
METHODOLOGY 6
ASSUMPTIONS 10
ANALYSIS AND INFERENCES 10
CONCLUSIONS 15
DISCUSSION AND FURTHER STUDY 16
REFERENCES 17

Abstract

The purpose of this report is to understand whether or not batsman performed and contributed the most to their team in the IPLT20 – 2018. Being a batsman friendly format of the game, this was required to understand overall player contributions. With the help of data collected from the official IPLT20 website (https://www.iplt20.com/stats/2018), bottom-up approach of using KPIs from linear regression analysis was used to rate players. KPIs used were a mix of batting and bowling attributes – Fours, Sixes, Batting Average, Non-Boundary Runs, Wickets, Dot Balls, runs given per over etc. These KPI coefficients were quantified using regression to generate a model to rate the players.

From the analysis undertaken, it was clear that pure bowlers and all-rounders were at the most valuable players in IPLT20 – 2018, with Sunil Narine from KKR being the MVP for the year. Interestingly enough, he is primarily a spin bowler from West Indies, who can knock the ball out of the park with the bat.

Some further data collection is recommended to improve the model and fine-tune the rating system – these include identifying when a bowler is bowling during power play restrictions, when a weak player is facing a strong opponent and quantifying fielder contributions.

Literature Review

In literature, if cricket scores need to be compared for performance rating, it is important to convert scores to a more meaningful total using Duckworth-Lewis-Stern method [1] in order to convert wickets in hand and over remaining into runs [2]. This helps measure team and player performance in a more accurate manner across the board. This is important for the project in question, as in this format of the game, there have been a lot of occurrences where team batting second has won by wickets.

When it comes to analysing, quantifying and assessing player performances [3], in cricket being a sport of discrete nature, players perform in in short-bursts of well-defined and measurable tasks – like wickets taken, boundaries hit – and can hence be used as key performance indicators (KPIs) for the sport [4].

As the research aims to identify the MVP in the series, as the format depends heavily on player auctions, this kind of player rating is important in order to be able to predict auction prices [5] and player value for teams. It has been researched that combining bowling attributes [6] can help in creating a variable that better rates a bowler's performance in a match. Keeping all of this in mind, generating a consistent model for rating prediction [7] would require attributing performance in all aspects of the game – bowling, batting and fielding.

Introduction

The Indian Premier League (IPL), is a professional Twenty20 format cricket league in India, with teams representing Indian cities and some states. In the 2018 IPLT20, there were 8 represented teams – Chennai Super Kings (CSK), Sunrisers Hyderabad (SRH), Kolkata Knight Riders (KKR), Royal Challengers Bangalore (RCB), Kings XI Punjab (KXIP), Rajasthan Royals (RR), Mumbai Indians (MI) and Delhi Daredevils (DD).

Over the years, the league has garnered huge fan following globally, and thus, talented sportsmen from all international cricket playing nations find this competition attractive. Each team is allowing to have 4 international players in their playing eleven, and hence prior to the league, there is an auction in which teams bid to get or retain their key players as well as try to get the best international player on their team.

This format of 20 over game has been seen to hugely favour the batsman – so much so that the IPL logo itself has a batsman swinging away. As players don't really stroke play to get in to form, they usually start on a war footing, which can be compared to calculated pinch hitting. Thus, with such unfavourable conditions for bowlers, it is crucial to pay importance to their contributions as well.

The purpose of this project is to formulate a rating method for this current season of IPLT20, to understand which kind of player contributes the most to his team – A batsman, bowler or an all-rounder. It is important to understand, whether bowlers are just throwing deliveries to make the batsman look good or are they crucial game changers as well.

To undertake this analysis, the bottom-up method will be used with linear regression playing a key part. The following sections with methodology and analysis will delve deeper into how this was carried out.

The dataset was manually scraped and recreated for this research from the official IPLT20 website (https://www.iplt20.com/stats/2018).

Objectives

The purpose of this report is to utilise the dataset derived from the official IPLT20 website (https://www.iplt20.com/stats/2018), and analyse the data using regression methods to rate the following:

- Best Batsmen in the league
- Best Bowler in the league
- MVP in the league

The key criteria of the research are to identify whether or not batsmen were the most crucial players for the IPLT20 2018 edition, and what players (Batsmen/Bowler/All-rounder) made the top 10 of the MVP list.

Dataset

The data was sourced from the official IPLT20 website (https://www.iplt20.com/stats/2018). The variables derived for this analysis are as follows:

Batsman Data

- Team – Which shows which team the players represent.
- Player – The names of the player for that row.
- M – Total number of matches played in IPLT20 2018
- R – The total number of runs scored in the season
- BP – The total number of deliveries faced (Balls Played).
- NO – The number of not outs in the season
- F – Number of fours hit by the player
- SX – Number of Sixes hit by the player
- NBR – Total non- boundary runs scored by the player
- BA – The Batting average of the player in the season.

Bowler Data

- Team – Which shows which team the players represent
- Player – Player Name
- M - Matches
- BB – Total Balls bowled
- O – Total Overs Bowled (BB/6)
- W – Number of Wickets taken
- D – Number of dot balls bowled
- RGPO – Runs given per over bowled (Economy)

Player Data

- Team – Which shows which team the players represent.
- Player – The names of the player for that row.
- M – Total number of matches played in IPLT20 2018
- R – The total number of runs scored in the season
- BP – The total number of deliveries faced (Balls Played).
- NO – The number of not outs in the season
- F – Number of fours hit by the player
- SX – Number of Sixes hit by the player
- NBR – Total non- boundary runs scored by the player
- BA – The Batting average of the player in the season.
- W – Wickets taken
- BPW – Balls per wicket
- RPHB – Runs per hundred balls (Strike rate)

Methodology

For the purpose of this project, the aim was to use bottom-up approach to rate players according to certain KPIs. These KPIs differed for batsmen and bowler and were combined to get the overall MVP for the season.

Linear regression approach on MS Excel was used to put a value on these particular KPI/coefficients, which were multiplied with the player data to give them ratings. The details of each of the three ratings methods are described in detail below.

Batting Rating

For batsmen ratings, I used F, SX, NBR and BA as KPI's. To get values for this, the match data for batsmen during the entire season was differenced based on the above KPIs. Below is a sample illustration of the spreadsheets used in match data and differencing.

Table 1. Showing first 10 entries of the matches of IPLT20 -2018

Match	Team A	Team B	R	BF	BA	RPHB	DLS	F	SX	NBR
1	MI	CSK	165	120	41.3	137.5	165	18	5	63
1	CSK	MI	169	119	18.8	142.0	174	11	10	65
2	KXIP	DD	167	113	41.8	147.8	184	16	7	61
2	DD	KXIP	166	120	23.7	138.3	166	11	5	92
3	KKR	RCB	177	113	29.5	156.6	194	15	9	63
3	RCB	KKR	176	120	25.1	146.7	176	14	10	60
4	SRH	RR	127	95	127.0	133.7	168	17	2	47
4	RR	SRH	125	120	13.9	104.2	125	12	0	77
5	CSK	KKR	205	119	41.0	172.3	208	9	14	85
5	KKR	CSK	202	120	33.7	168.3	202	11	17	56
6	RR	DD	153	107	30.6	143.0	153	11	6	73
6	DD	RR	60	36	15.0	166.7	177	7	2	20
7	SRH	MI	151	120	16.8	125.8	151	18	1	73
7	MI	SRH	147	120	18.4	122.5	147	15	6	51
8	RCB	KXIP	159	117	26.5	135.9	166	16	5	65
8	KXIP	RCB	155	116	15.5	133.6	155	12	6	71
9	MI	DD	194	120	27.7	161.7	194	19	7	76
9	DD	MI	195	120	65.0	162.5	195	18	10	63
10	KKR	SRH	138	120	17.3	115.0	138	12	4	66
10	SRH	KKR	139	114	27.8	121.9	151	14	3	65

The spreadsheet above shows how each match was taken twice, from the viewpoint of each team. In the occurrence of an outcome where the team won by wickets, or when the match overs were reduced due to rain or play stoppage, the runs were converted into duckworth-lewis-stern (DLS) score, to be able to give a more accurate comparison for differencing, which is shown in the next table below.

Table 2. Showing first 10 entries of the differenced data for batsman regression variables.

Match	Team A	F	SX	NBR	BA DIFF	RPHB DIFF	DLS Margin
1	MI	7	-5	-2	22.5	-4.5	-9
1	CSK	-7	5	2	-22.5	4.5	9
2	KXIP	5	2	-31	18.0	9.5	18
2	DD	-5	-2	31	-18.0	-9.5	-18
3	KKR	1	-1	3	4.4	10.0	18
3	RCB	-1	1	-3	-4.4	-10.0	-18
4	SRH	5	2	-30	113.1	29.5	43
4	RR	-5	-2	30	-113.1	-29.5	-43
5	CSK	-2	-3	29	7.3	3.9	6
5	KKR	2	3	-29	-7.3	-3.9	-6
6	RR	4	4	53	15.6	-23.7	-24
6	DD	-4	-4	-53	-15.6	23.7	24
7	SRH	3	-5	22	-1.6	3.3	4
7	MI	-3	5	-22	1.6	-3.3	-4
8	RCB	4	-1	-6	11.0	2.3	11
8	KXIP	-4	1	6	-11.0	-2.3	-11
9	MI	1	-3	13	-37.3	-0.8	-1
9	DD	-1	3	-13	37.3	0.8	1
10	KKR	-2	1	1	-10.6	-6.9	-13
10	SRH	2	-1	-1	10.6	6.9	13

The above table shows the differenced batsman data, which was used to run linear regression in order to predict the DLS Margin.
The DLS scores were calculated using the following formula.

Team 2's par score = Team 1's score x (Team 2's resources/Team 1's resources)

Where resources are the wickets in hand and overs left.

Once the regression is carried out, certain coefficient values are derived which are then used in the batsman dataset to get the ratings.

Bowler Ratings

For bowler ratings - W, D and runs given per over, were used as KPIs. Similar to the batsmen rating, to get values for this, the match data for bowlers during the entire season was differenced based on the above KPIs.

Below is a sample illustration of the spreadsheets used in match data and differencing.

Table 3. Showing first 10 entries of the matches using bowlers data of IPLT20 -2018

`	Team A	Team B	R	DLS	BB	W	D
1	MI	CSK	165	165	119	9	46
1	CSK	MI	169	174	120	4	49
2	KXIP	DD	167	184	120	7	28
2	DD	KXIP	166	166	113	4	41
3	KKR	RCB	177	194	120	7	46
3	RCB	KKR	176	176	113	6	44
4	SRH	RR	127	168	120	9	48
4	RR	SRH	125	125	95	1	44
5	CSK	KKR	205	208	120	6	45
5	KKR	CSK	202	202	119	5	31
6	RR	DD	153	153	36	4	14
6	DD	RR	60	177	107	5	34
7	SRH	MI	151	151	120	8	62
7	MI	SRH	147	147	120	9	39
8	RCB	KXIP	159	166	116	10	42
8	KXIP	RCB	155	155	117	6	42
9	MI	DD	194	194	120	3	38
9	DD	MI	195	195	120	7	39
10	KKR	SRH	138	138	114	5	43
10	SRH	KKR	139	151	120	8	54

Table 4. Showing first 10 entries of the differenced data for bowler regression variables.

Match	Team A	W	D	BPW	RG	DLS Margin
1	MI	5	-3	-16.8	-4	-9
1	CSK	-5	3	16.8	4	9
2	KXIP	3	-13	-11.1	1	18
2	DD	-3	13	11.1	-1	-18
3	KKR	1	2	-1.7	1	18
3	RCB	-1	-2	1.7	-1	-18
4	SRH	8	4	-81.7	2	43
4	RR	-8	-4	81.7	-2	-43
5	CSK	1	14	-3.8	3	6
5	KKR	-1	-14	3.8	-3	-6
6	RR	-1	-20	-12.4	93	-24
6	DD	1	20	12.4	-93	24
7	SRH	-1	23	1.7	4	4
7	MI	1	-23	-1.7	-4	-4
8	RCB	4	0	-7.9	4	11
8	KXIP	-4	0	7.9	-4	-11
9	MI	-4	-1	22.9	-1	-1
9	DD	4	1	-22.9	1	1

10	KKR	-3	-11	7.8	-1	-13
10	SRH	3	11	-7.8	1	13

The above table shows the differenced bowler data, which was used to run linear regression in order to predict the DLS Margin.

Once the regression is carried out, certain coefficient values are derived which are then used in the bowler dataset to get the ratings.

MVP Ratings

For MVP ratings, a mix of batsman and bowler ratings were used to generate a linear regression model and valuate the coefficients. A sampled of the differenced data used for the MVP rating is given in the table below.

Table 5. Showing first 10 entries of the differenced data for MVP regression variables.

Match	Team A	F	SX	NBR	BA DIFF	BPW	W	RPHB DIFF	D	RG	DLS Margin
1	MI	7	-5	-2	22.5	-16.8	5	-4.5	-3	-4	-9
1	CSK	-7	5	2	-22.5	16.8	-5	4.5	3	4	9
2	KXIP	5	2	-31	18.0	-11.1	3	9.5	-13	1	18
2	DD	-5	-2	31	-18.0	11.1	-3	-9.5	13	-1	-18
3	KKR	1	-1	3	4.4	-1.7	1	10.0	2	1	18
3	RCB	-1	1	-3	-4.4	1.7	-1	-10.0	-2	-1	-18
4	SRH	5	2	-30	113.1	-81.7	8	29.5	4	2	43
4	RR	-5	-2	30	-113.1	81.7	-8	-29.5	-4	-2	-43
5	CSK	-2	-3	29	7.3	-3.8	1	3.9	14	3	6
5	KKR	2	3	-29	-7.3	3.8	-1	-3.9	-14	-3	-6
6	RR	4	4	53	15.6	-12.4	-1	-23.7	-20	93	-24
6	DD	-4	-4	-53	-15.6	12.4	1	23.7	20	-93	24
7	SRH	3	-5	22	-1.6	1.7	-1	3.3	23	4	4
7	MI	-3	5	-22	1.6	-1.7	1	-3.3	-23	-4	-4
8	RCB	4	-1	-6	11.0	-7.9	4	2.3	0	4	11
8	KXIP	-4	1	6	-11.0	7.9	-4	-2.3	0	-4	-11
9	MI	1	-3	13	-37.3	22.9	-4	-0.8	-1	-1	-1
9	DD	-1	3	-13	37.3	-22.9	4	0.8	1	1	1
10	KKR	-2	1	1	-10.6	7.8	-3	-6.9	-11	-1	-13
10	SRH	2	-1	-1	10.6	-7.8	3	6.9	11	1	13

Once the regression is carried out, certain coefficient values are derived which are then used in the MVP dataset to get the ratings.

Certain assumptions were used in the project, which have biased the ratings slightly, these have been detailed in the next section.

Assumptions

Rating players can be done in multiple ways, as players contribute not only as a bowler and batsman, but as a fielder as well. However, as it is hard to quantify the number of runs saved by an individual on the field, the ratings cannot be accurate. Therefore, the following assumptions have been used in carrying out this project.

1. Catches – The number of catches caught by players have been excluded. The reasoning behind this is, catches may be attributed to the bowler delivering a good delivery, hence forcing the batsman to mishit the ball to a field location where a fielder has been placed. Hence, using it purely as an attribute to increase rating for the catch taker is not ideal.

2. Stumping – As above, stumpings are carried out only by the wicket keepers of each team. Hence, using this as a criterion to rate players would not be ideal, as it is the bowler who has bowled an unplayable delivery to beat the batsman.

3. Run outs – Run outs are again a combination of bowler bowling a good delivery, building pressure and forcing the batsman to miss-hit and risk a run.

Based on the above assumptions, the overall ratings may be slightly biased, but will provide a more accurate representation of player performances based on batting and bowling.

Analysis and Inferences

This section is divided into three parts which cater Batting Ratings, Bowler Ratings and MVP Ratings. This has been done in order to understand the most valuable players and their contributions in this season of IPLT20 – 2018.

Batsman Ratings

After using the regression variables to predict the DLS margins, the linear regression model generated the following results.

Table 6. Regression Statistics for Batsman Variables

Multiple R	0.79156459
R Square	0.6265745
Adjusted R Square	0.61358579
Standard Error	18.1210968
Observations	120

From table 6, we can see that the model is a good fit, with an R squared value of 62.65%, and a standard error of approximately 1.6%. This goes to show that the model resulted in variables, which are able to produce satisfactory ratings for the batsmen. Table 7 below shows the coefficient values from this linear regression model.

Table 7. Regression coefficient values and significance for Batsmen.

	Coefficients	*Standard Error*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0	1.654222243	1	-3.2766959	3.27669589	-3.2766959	3.27669589
F	2.94916636	0.388026434	8.6624E-12	2.18056068	3.71777205	2.18056068	3.71777205
SX	4.11589402	0.470930163	2.2127E-14	3.183072	5.04871604	3.183072	5.04871604
NBR	0.07351617	0.100137695	0.46435106	-0.1248373	0.27186967	-0.1248373	0.27186967
BA DIFF	0.2899413	0.050353919	7.2037E-08	0.19019988	0.38968273	0.19019988	0.38968273

From the above table, we can see that all the coefficients, except for NBR are significant at a 5% significance level, hence showing a good model.

For batsmen, the number of sixes hit, has been the highest rated coefficient with a 4.11 valuation, followed by number of fours hit at 2.949. The batting average and NBR have not been rated too high, which is a good indication of the aggressive nature of batting in this format of the game.

Based on these KPI coefficient values, a batting rating for all the players was generated, which yielded the batsmen rating table. Table 8 below, shows the top 10 batsmen for this season of IPLT20.

Table 8. Batsman Rating for IPLT20 -2018

Rank	Team	Payer	Batsman Rating	Fours	Sixes	NBR	RPHB
1	DD	Rishabh Pant	382.05	68	37	190	173.60
2	KXIP	Lokesh Rahul	357.20	66	32	203	158.41
3	SRH	Kane Williamson	342.08	64	28	311	186.55
4	CSK	Ambati Rayudu	322.39	53	34	186	149.75
5	CSK	Shane Watson	297.74	44	35	169	154.60
6	RR	Jos Buttler	271.41	52	21	214	155.24
7	MI	Suryakumar Yadav	269.00	61	16	172	133.33
8	RCB	AB de Villiers	263.00	39	30	144	174.55
9	KKR	Chris Lynn	260.42	56	18	159	130.24
10	RCB	Virat Kohli	257.15	52	18	214	139.11

According to these ratings, Rishabh Pant (DD) is the most valuable batsman for this season of IPLT20, having 68 fours, 37 sixes and a strike rate of 173.6. Even though Kane Williamson (SRH) had the highest number of runs scored, as he played more matches, the number of boundaries scored by him was lower albeit having a slightly better strike rate.

For IPL 2019, these players would be valued as the most valuable batsman and should have higher auction rates than their peers.

<u>Bowler Ratings</u>

After using the regression variables to predict the DLS margins, the linear regression model generated the following results.

Table 9. Regression Statistics for Bowler Variables	
Multiple R	0.82180126
R Square	0.67535731
Adjusted R Square	0.66696137
Standard Error	16.8230694
Observations	120

From table 9 above, we can see that the model is a good fit, with an R squared value of 67.53%, and a standard error of approximately 1.5%. This goes to show that the model resulted in variables, which are able to produce satisfactory ratings for the bowlers. Table 10. below, show the coefficient values from this linear regression model.

Table 10. Regression coefficient values and significance for Bowlers.

	Coefficients	*Standard Error*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0	1.5357291	1	-3.041705	3.04170499	-3.041705	3.04170499
W	3.52148178	0.51152195	3.161E-10	2.5083481	4.53461545	2.5083481	4.53461545
D	0.72692955	0.16896115	3.5447E-05	0.3922807	1.06157839	0.3922807	1.06157839
RG	0.52596671	0.05680297	1.29E-15	0.41346127	0.63847215	0.41346127	0.63847215

From the above table, we can see that all the coefficients are significant at a 5% significance level, hence showing a good model.

For bowlers, the number of wickets taken, has been the highest rated coefficient with a 3.52 valuation, followed by number of dot balls bowled at 0.72. The economy of bowlers has not been rated very highly, which yet again reflects on this format being highly favourable for batsman.

Based on these KPI coefficient values, a bowler rating for all the players was generated, which yielded the bowler rating table. Table 11 below, shows the top 10 bowlers for this season of IPLT20.

Table 11. Bowler Rating for IPLT20 -2018

Rank	Team	PLAYER	Bowler Rating	W	D	RGPO
1	SRH	Rashid Khan	191.81	21	167	6.74
2	RCB	Umesh Yadav	173.88	20	148	7.86
3	SRH	Siddarth Kaul	164.82	21	131	8.29
4	KXIP	Andrew Tye	164.63	24	116	8.00
5	KKR	Sunil Narine	155.43	17	137	7.66
6	MI	Jasprit Bumrah	152.92	17	133	6.89
7	DD	Trent Boult	144.51	18	118	8.85
8	CSK	Shardul Thakur	133.63	16	113	9.24
9	RCB	Yuzvendra Chahal	131.49	12	128	7.26
10	MI	Hardik Pandya	129.93	18	98	8.93

According to these ratings, Rashid Khan (SRH) is the most valuable bowler for this season of IPLT20, having taken 21 wickets, delivered 167 dot balls and having an economy of 6.74 runs per over. Even though Andrew Tye (KXIP) had the highest number of wickets, the number of dot balls bowled by him was lower as well as having a lower economy.

For IPL 2019, these players would be valued as the most valuable bowler and should have higher auction rates than their peers.

MVP Ratings

Now that we have seen that the regression models are working accurately, a combination of variables for both batsmen and bowler have been used to generate the MVP list. This part was the most crucial, as it was necessary to understand what kind of players were contributing most in this format and league.

After using linear regression on the combined player variables to predict DLS margin, the results obtained are reported in the table below.

Table 12. Regression Statistics for MVP Variables	
Multiple R	0.89583877
R Square	0.80252711
Adjusted R Square	0.79204182
Standard Error	13.2937019
Observations	120

The results of the regression were indeed promising. With an R squared value of 80.2%, this model shows it's a good fit. The standard error is 1.3%. This goes to show that the model resulted in variables, which are able to produce satisfactory ratings for the MVP. Table 13. below, show the coefficient values from this linear regression model.

Table 13. Regression coefficient values and significance for MVPs.

	Coefficients	*Standard Error*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0	1.213543399	1	-2.4042483	2.40424835	-2.4042483	2.40424835
F	2.08085264	0.299471403	2.503E-10	1.48754577	2.6741595	1.48754577	2.6741595
SX	3.01170165	0.365929132	3.607E-13	2.28673004	3.73667325	2.28673004	3.73667325
NBR	0.14914342	0.073859563	0.04582458	0.0028143	0.29547254	0.0028143	0.29547254
BA DIFF	1.95360632	0.255590703	7.5021E-12	1.44723504	2.4599776	1.44723504	2.4599776
W	4.41962711	0.533610873	2.7463E-13	3.36244773	5.47680649	3.36244773	5.47680649
BPW	3.20785681	0.424611052	1.1805E-11	2.36662573	4.04908789	2.36662573	4.04908789

From the above table, we can see that all the coefficients are significant at a 5% significance level, hence showing a good model.

For MVPs, it is seen that the weightage given to both batting and bowling coefficients are well balanced, which is quite an interesting result. This shows that a good balanced individual should be rated higher than one that is strong in just one aspect.

Based on these KPI coefficient values, an MVP rating for all the players was generated, which yielded the MVP rating table. Table 14 below, shows the MVPs for this season of IPLT20.

Table 14. MVP Ratings for IPLT20 - 2018

Rank	Team	Player	Overall Rating	Fours	Sixes	BA	Wickets	RGPO
1	KKR	Sunil Narine	719.50	40	23	22.31	17	7.66
2	SRH	Rashid Khan	662.16	3	6	3.93	21	6.74
3	RCB	Umesh Yadav	564.05	0	0	0.23	20	7.86
4	MI	Hardik Pandya	542.11	20	11	28.89	18	8.93
5	SRH	Siddarth Kaul	513.86	0	0	0.19	21	8.29
6	CSK	Shane Watson	508.98	44	35	39.64	6	8.96
7	MI	Jasprit Bumrah	506.38	1	0	0.83	17	6.89
8	KXIP	Andrew Tye	493.25	2	1	2.67	24	8.00
9	SRH	Shakib Al Hasan	492.19	26	5	15.93	14	8.00
10	KKR	Andre Russell	486.72	17	31	24.31	13	9.38

According to these ratings, Sunil Narine (KKR) is the most valuable player for this season of IPLT20, for having not only taken 17 wickets, but having hit 40 fours, 23 sixes and having a batting average of 22.31%, and boasting an economy of 7.66 runs per over.

It is clear from the above table, that it is primarily all-round players and bowlers who consist of the MVP top ten.

The reason for this may be the fact that a good bowling spell and economy can help restrict runs by the batsmen, which would impact the overall outcome of a match. Being such a batsman friendly format, having a good run with the bat is necessary for every individual.

So, while it is clear that skilled bowlers who can bat are forming this list, like Andre Russell, Shane Watson, Hardik Pandya, there are also pure top bowlers like Umesh Yadav and Jasprit Bumrah who make the top 10 this year.

For IPL 2019, these players would be valued as the most valuable players of 2018 and should have higher auction rates than their peers.

Conclusions

From the linear regression models generated it is seen that the model for MVP was the best fit, in terms of R squared value and rating the players. Sunil Narine was the MVP for IPL 2018 in actuality as well. Certain tweaks are recommended to improve the model, so as to be able to generate a more accurate rating system, and these recommendations have been mentioned in the next section.

From the above sections, it is clear, the KPIs used to understand a player's performance is accurate.

The main question that needed to be answered was whether or not batsmen were the most contribution and best performers in this format of the game, and since IPLT20 seem to favour batsmen, the common thought would point in that direction. However, results from the analysis above are indeed contradictory to that, and they show all-rounders and pure bowlers making the top of the most valuable players list for 2018. This can be seen from the table below.

MVP List with Rating for IPLT20 – 2018 (The top 20 this year)

Rank	Team	Player	Overall Rating	Fours	Sixes	BA	Wickets	RGPO
1	KKR	Sunil Narine	719.50	40	23	22.31	17	7.66
2	SRH	Rashid Khan	662.16	3	6	3.93	21	6.74
3	RCB	Umesh Yadav	564.05	0	0	0.23	20	7.86
4	MI	Hardik Pandya	542.11	20	11	28.89	18	8.93
5	SRH	Siddarth Kaul	513.86	0	0	0.19	21	8.29
6	CSK	Shane Watson	508.98	44	35	39.64	6	8.96
7	MI	Jasprit Bumrah	506.38	1	0	0.83	17	6.89
8	KXIP	Andrew Tye	493.25	2	1	2.67	24	8.00
9	SRH	Shakib Al Hasan	492.19	26	5	15.93	14	8.00
10	KKR	Andre Russell	486.72	17	31	24.31	13	9.38
11	RCB	Yuzvendra Chahal	463.64	0	0	0.00	12	7.26
12	DD	Trent Boult	458.08	0	0	0.00	18	8.85
13	MI	Krunal Pandya	454.02	22	10	20.73	12	7.07
14	CSK	Deepak Chahar	445.80	1	4	4.55	10	7.28
15	SRH	Bhuvneshwar Kumar	443.51	1	0	1.30	9	7.67
16	CSK	Shardul Thakur	442.33	3	0	1.25	16	9.24
17	CSK	Dwayne Bravo	432.20	8	10	14.10	14	9.96
18	KKR	Piyush Chawla	426.76	1	1	2.25	14	8.41
19	SRH	Sandeep Sharma	425.15	0	0	0.00	12	7.57
20	RR	Krishnappa Gowtham	410.90	9	9	11.45	11	7.77

Thus, it is safe to say that this year, the bowlers and all-rounders were the top contributors in IPLT20 by being able to not only take wickets and reduce runs scored by opposition, but by contributing with the bat consistently as well.

Discussion and Further Study

The results and analysis from the above sections, show that the project is headed in the right direction, however there is a huge scope to finetune the dataset collected and introducing certain variables that will help predict player ratings in a more accurate manner. Some of the areas that may help improve the model are explained below.

- <u>Introducing a variable that explains whether a bowler was bowling during power plays</u>
  As power plays are crucial periods in the game with field restrictions active, it is easier for batsmen to hit over the inner field and score runs. Fluke shots from edges and mishits tend to result in more boundaries during power plays because of this reason. Adding this as a variable will help understand whether a bowler took wickets and had a good economy rate during the power plays versus no-restriction times will help fine-tune the model further.

- <u>An attribute that measures when a weak player is facing a strong opponent on either side of the innings</u>
  A weaker bowler taking the wicket of an in-form batsman is of more importance and weightage, than a strong bowler bowling out a tail ender. Similarly, for a weaker batsman scoring against a strong bowler and vice versa. Identifying this attribute will help further improve the rating system.

- <u>Measuring the number of runs saved by a player on field</u>
  Being able to quantify runs saved by a fielder in the form of boundaries and good fielding would further help this rating system, as then the fielding aspect of the sport would also be able to help predict the final DLS margin in the linear regression.

A detailed collection and pre-processing of the ball-by-ball data may help in understanding the significance of the first two points detailed above. For the fielding attribute, data will need to be collected and attributed to the players during the game itself, which can they be used to analyse and improve the rating system further. This would make the game more competitive and valuate players on a whole new level.

References

1. Stern, S.E., 2016. The Duckworth-Lewis-Stern method: extending the Duckworth-Lewis methodology to deal with modern scoring rates. *J Oper Res Soc*, 67(12), pp.1469–1480.
2. Perera, H.P. & Swartz, T.B., 2013. Resource estimation in T20 cricket. *IMA Journal of Management Mathematics*, 24(3), pp.337–347.
3. Davis, J., Perera, H. & Swartz, T.B., 2015. Player evaluation in Twenty20 cricket. *Journal of Sports Analytics*, 1(1), pp.19–31.
4. Jhawar, M.G.G. & Pudi, V., 2017. Honest mirror: Quantitative assessment of player performances in an ODI cricket match. *CEUR Workshop Proceedings*, 1971, pp.62–72.
5. Siddhartha K. Rastogi, 2017. What's a Cricketer's Worth? Predicting Bid Prices for Indian Premier League Auctions. *Annals of Dunarea de Jos University. Fascicle I : Economics and Applied Informatics*, 23(1), pp.127–133.
6. Lemmer, H.H., 2014. Perspectives on the Use of the Combined Bowling Rate in Cricket. *International Journal of Sports Science & Coaching*, 9(3), pp.513–523
7. Lemmer, H.H., Bhattacharjee, D. & Saikia, H., 2014. A Consistency Adjusted Measure for the Success of Prediction Methods in Cricket. *International Journal of Sports Science & Coaching*, 9(3), pp.497–512.
8. Khondker, Habibul Haque et al., 2018. Glocalization, consumption, and cricket: The Indian Premier League. *Journal of Consumer Culture*, 18(2), pp.279–297.
9. Data sourced from - Iplt20.com. (2018). *IPLT20.com - Indian Premier League Official Website - Stats*. (https://www.iplt20.com/stats/2018)