

MACHINE LEARNING TO PREDICT CRICKET SCORECARD

Abstract:

Cricket, over 2.5 Billion people watch this and 105 countries which plays game all over the world. The statistics of professional sports, including players and teams, provide numerous opportunities for research. The aim of this study is to investigate to what degree it is possible to predict the outcome the cricket matches by predicting individual performances in the team. The target competition is T20 cricket matches. The original features alongside engineered features gave rise to more than 8000 matches' statistics. In this study I will develop machine learning model using **neural networks** in order to predict the score card of a given cricket match.

Motivation:

Data science has recently become a popular term in the industry and providing plenty of opportunities for both professionals and organizations in growth. The two main reasons doing this project are **cricket** and **data science**.

Literature review:

An extensive online search produced very few articles related to players' performance prediction in the game of cricket. A very small number of researchers have studied the performance of cricket players. Muthuswamy and Lam predicted the performance of Indian bowlers against seven international teams against which the Indian cricket team plays most frequently. They used back propagation network and radial basis network function to predict how many runs a bowler is likely to concede and how many wickets a bowler is likely to take in a given ODI match. Wikramasinghe predicted the performance of batsmen in a test series using a hierarchical linear model. Barr and Kantor defined a criterion for comparing and selecting batsmen in limited overs cricket. Iyer and Sharda used neural networks to predict the performance of players where they classify batsmen and bowlers separately in three categories – performer, moderate and failure.

Although there are very few researches done in the past. Most of them are limited to a particular category either bowling or batting with very limited data. And aiming to predict while match is in progress. I am aiming to use each ball that has been bowled till the date, with an extensive usage of data, the accuracy of prediction will increase.

Objective:

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The objective of this project is to identify the best features that affects player performance and build an optimized model to predict each player individual contribution to the final score of the match. To calculate the players features I have consider all the historical matches played by a player till date.

Modules of the project:

- Web scraping data from cricinfo
- Modelling Data (identifying features/calculating new features and modelling)
- Tuning the model with optimal hyper parameters.
- Prediction, furcating and visualization

Methodology:

Since the outcome of our problem is numerical data we will be using regression algorithms for modelling.

Till now I was under impression that the traditional features like batting average, centuries, form, bowling average bowling strike rate etc. are the deciding factors of a batsmen performance when facing a bowler. But in my study found the head to head performance is effecting a batsmen performance more than anything else.

Features selection:

After doing statistical analysis on data finally I came up with the fallowing features that has more correlation towards a batsmen getting out.

Feature	Explanation
POSITION	Batting Position
OVER_STARTED	The first ball batsmen faced in a match
GROUND_AVG	The average over that batsmen get out in ground
AVG_OUT	The average over that batsmen get out in career
BOL_AVG_OUT	The average over that opponent team will get batsmen out
TEAM_EFF	The head to head effort on the batsmen from the opponent team
AVG	The average of all the features
OUT	The actual over batsmen got out

Predicting the batsmen runs will be based on when the batsmen will start playing i.e the starting over and the ending over.

The below table help us to understand correlation between the features with actual out

AVG==> 1.0

[7]:

	PLAYER_NAME	POSITION	OVER_STARTED	GROUND_AVG	AVG_OUT	BOL_AVG_OUT	TEAM_EFF	AVG
0	SR Watson	0.570499	0.630865	0.604646	0.623107	7.804301e-01	0.713924	8.364736e-01
1	F du Plessis	0.446521	0.498424	0.538206	0.576487	6.525376e-01	0.755470	7.824509e-01
2	SK Raina	0.340074	0.491799	0.437161	0.421492	7.557349e-01	0.605368	7.705938e-01
3	MS Dhoni	0.043807	0.550215	0.347945	0.452416	5.911428e-01	0.645165	7.239499e-01
4	AT Rayudu	0.596852	0.699395	0.600947	0.681598	7.630061e-01	0.812487	8.618313e-01
5	KM Jadhav	0.493946	0.748265	0.560255	0.673188	7.271079e-01	0.695502	7.953082e-01
6	SC Kuggeleijn	0.217571	0.267112	0.113228	0.679366	5.070725e-01	0.730769	4.961389e-01
7	RA Jadeja	0.226754	0.600710	0.420369	0.489488	6.483416e-01	0.549259	7.075238e-01
8	Harbhajan Singh	0.738016	0.840454	0.771662	0.867449	8.446283e-01	0.735925	8.522755e-01
9	DL Chahar	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	Imran Tahir	-0.400320	0.544581	NaN	0.144338	2.314677e-17	0.365148	-1.259018e-17
11	KL Rahul	0.632997	0.610501	0.659062	0.776772	8.141830e-01	0.815337	8.872970e-01
12	CH Gayle	0.141508	0.138678	0.305174	0.327733	6.455053e-01	0.662874	7.736229e-01
13	MA Agarwal	0.785906	0.832810	0.777462	0.883407	8.759426e-01	0.810622	9.201792e-01
14	DA Miller	0.330431	0.662709	0.379690	0.491105	6.872401e-01	0.755478	7.884241e-01
15	SN Khan	0.511334	0.643465	0.497403	0.914893	5.276587e-01	0.492922	6.503011e-01
16	Mandeep Singh	0.769098	0.783108	0.737971	0.865894	8.910221e-01	0.829362	9.266558e-01
17	SM Curran	0.619419	0.827545	0.608244	0.839289	8.649975e-01	0.946726	9.560547e-01
18	R Ashwin	0.033097	0.744751	0.419363	0.463631	5.917081e-01	0.786012	7.367552e-01
19	Mohammed Shami	-0.798584	0.989160	NaN	0.989160	6.142951e-01	0.828315	8.411582e-01
20	AJ Tye	0.478502	0.810950	0.729832	0.742385	5.583634e-01	0.914453	7.445831e-01
21	M Ashwin	1.000000	1.000000	1.000000	1.000000	1.000000e+00	1.000000	1.000000e+00

Since we are using neural networks, the hyper parameter tuning is very critical aspect of modelling. For hyper parameters tuning we have used grid search to identify the optimal neurons and hidden layers

The below table contains the hyper parameter tuning details.

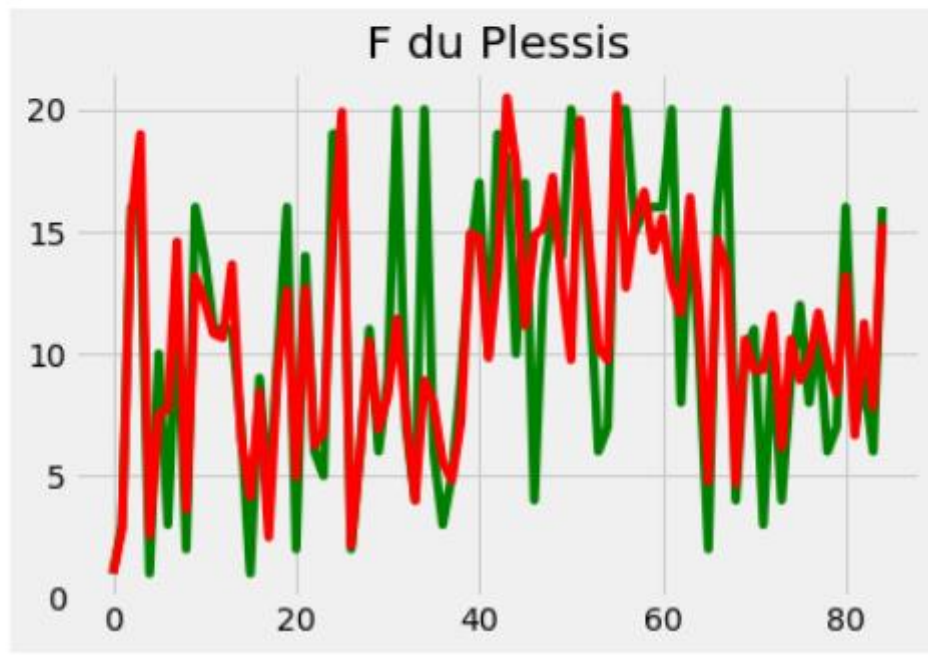
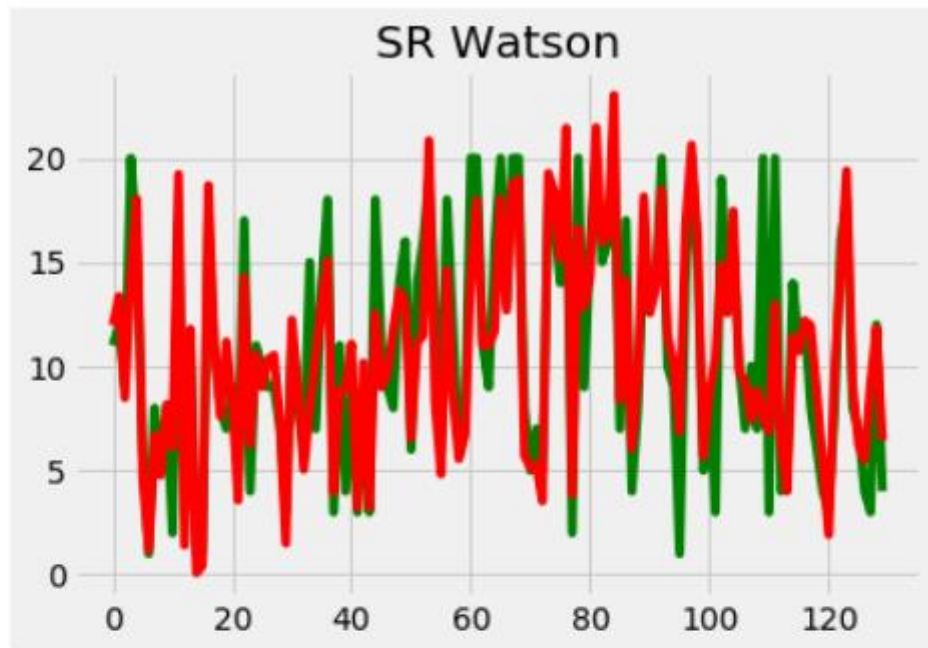
1] :

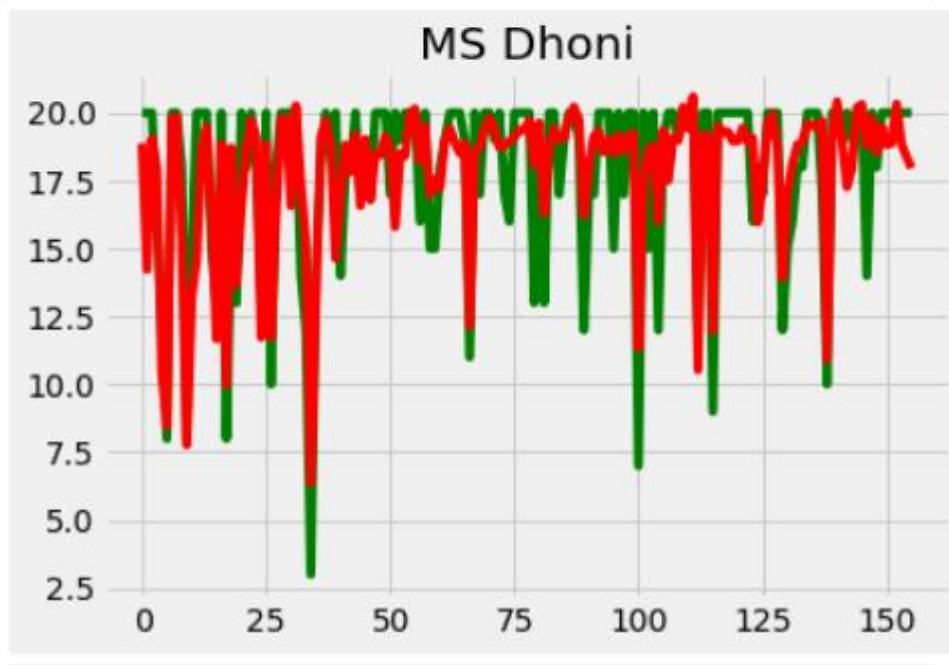
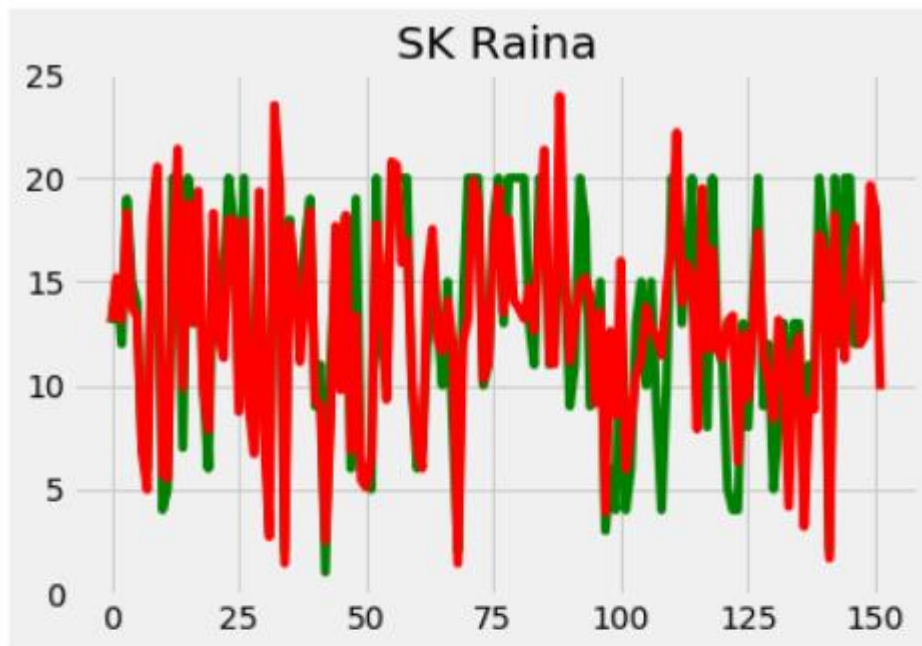
	PLAYER_ID	PLAYER_NAME	HIDDEN_LAYER	SCORE	INN
0	8180	SR Watson	(3, 8, 7)	7.305210e-01	1
1	44828	F du Plessis	(1, 4, 4)	6.353966e-01	1
2	33335	SK Raina	(3, 5, 8)	6.489154e-01	1
3	28081	MS Dhoni	(1, 6, 2)	6.948763e-01	1
4	33141	AT Rayudu	(5, 6, 3)	8.342122e-01	1
5	290716	KM Jadhav	(7, 8, 5)	7.621364e-01	1
6	539548	SC Kuggeleijn	(7, 1, 7)	-7.867318e-10	1
7	234675	RA Jadeja	(6, 8, 6)	7.001487e-01	1
8	29264	Harbhajan Singh	(4, 4, 7)	8.982150e-01	1
9	447261	DL Chahar	(0, 0, 0)	0.000000e+00	1
10	40618	Imran Tahir	(4, 1, 3)	9.083306e-01	1
11	422108	KL Rahul	(3, 2, 8)	8.543797e-01	1
12	51880	CH Gayle	(1, 5, 8)	6.406560e-01	1
13	398438	MA Agarwal	(4, 4, 2)	9.789275e-01	1
14	321777	DA Miller	(7, 6, 5)	7.992920e-01	1
15	642525	SN Khan	(4, 4, 8)	9.515977e-01	1
16	398506	Mandeep Singh	(8, 2, 5)	9.001676e-01	1
17	662973	SM Curran	(7, 4, 5)	9.399293e-01	1
18	26421	R Ashwin	(2, 5, 6)	8.490632e-01	1
19	481896	Mohammed Shami	(3, 1, 2)	-7.038881e-11	1
20	459508	AJ Tye	(4, 7, 6)	9.563923e-01	1
21	528067	M Ashwin	(0, 0, 0)	0.000000e+00	1

Result

Since we have the optimal parameters now we can fit model for each player and see how the prediction works. The prediction will be performed for the selected match from IPL.

Below we can see the graphs and charts that help us to understand how good the model for each player





The above are examples of modelling.

Now we can see the results of the model in the form of numeric values.

For Batting:

	PLAYER_ID	NAME	TEAM	TYPE	CR	START	END	RUNS
0	8180	SR Watson	CSK	BAT	10.0	1	11.0	47.0
1	44828	F du Plessis	CSK	BAT	9.0	1	10.0	39.0
2	33335	SK Raina	CSK	BAT	9.5	10	15.0	24.0
3	28081	MS Dhoni	CSK	WK	9.0	11	19.0	39.0
4	33141	AT Rayudu	CSK	BAT	9.5	15	15.0	3.0
5	290716	KM Jadhav	CSK	BAT	9.0	15	20.0	30.0
5	539548	SC Kuggeleijn	CSK	BOL	8.0	19	19.0	4.0
7	234675	RA Jadeja	CSK	ALL	8.5	19	19.0	0.0
3	29264	Harbhajan Singh	CSK	BOL	8.0	19	19.0	0.0
9	447261	DL Chahar	CSK	BOL	8.5	19	19.0	0.0
10	40618	Imran Tahir	CSK	BOL	8.5	19	19.0	0.0

	PLAYER_ID	NAME	TEAM	TYPE	CR	START	END	RUNS
0	422108	KL Rahul	KING	WK	10.0	1	20	92.0
1	51880	CH Gayle	KING	BAT	10.5	1	13	62.0
2	398438	MA Agarwal	KING	BAT	8.0	13	13	4.0
3	321777	DA Miller	KING	BAT	9.0	13	13	3.0
4	642525	SN Khan	KING	BAT	8.0	13	15	10.0
5	398506	Mandeep Singh	KING	BAT	8.0	15	20	32.0
6	662973	SM Curran	KING	ALL	8.5	20	20	0.0
7	26421	R Ashwin	KING	ALL	9.0	20	20	0.0
8	481896	Mohammed Shami	KING	BOL	8.5	20	20	0.0
9	459508	AJ Tye	KING	BOL	9.0	20	20	0.0
10	528067	M Ashwin	KING	BOL	8.0	20	20	0.0

For bowling:

	PLAYER_ID	NAME	TEAM	TYPE	CR	RUNS	RNK
2	28081	MS Dhoni	CSK	WK	9.0	16.0	1.0
14	422108	KL Rahul	KING	WK	10.0	0.0	2.0
19	539548	SC Kuggeleijn	CSK	BOL	8.0	14.0	1.0
3	29264	Harbhajan Singh	CSK	BOL	8.0	0.0	4.5
6	40618	Imran Tahir	CSK	BOL	8.5	0.0	4.5
15	447261	DL Chahar	CSK	BOL	8.5	0.0	4.5
16	459508	AJ Tye	KING	BOL	9.0	0.0	4.5
17	481896	Mohammed Shami	KING	BOL	8.5	0.0	4.5
18	528067	M Ashwin	KING	BOL	8.0	0.0	4.5
20	642525	SN Khan	KING	BAT	8.0	18.0	1.0
12	398438	MA Agarwal	KING	BAT	8.0	17.0	2.0
4	33141	AT Rayudu	CSK	BAT	9.5	15.0	4.5
10	290716	KM Jadhav	CSK	BAT	9.0	15.0	4.5
11	321777	DA Miller	KING	BAT	9.0	15.0	4.5
13	398506	Mandeep Singh	KING	BAT	8.0	15.0	4.5
5	33335	SK Raina	CSK	BAT	9.5	13.0	7.0
7	44828	F du Plessis	CSK	BAT	9.0	1.0	8.0
0	8180	SR Watson	CSK	BAT	10.0	0.0	9.5
8	51880	CH Gayle	KING	BAT	10.5	0.0	9.5
21	662973	SM Curran	KING	ALL	8.5	13.0	1.0
1	26421	R Ashwin	KING	ALL	9.0	0.0	2.5
9	234675	RA Jadeja	CSK	ALL	8.5	0.0	2.5

Discussion

Based on the result we can now say that predicting a very dynamic sport is not that easy. There are many factors that will affect the nature of the game and many of the features are not considered in the study like weather information.

However, our model performed quite well that we will have a better understanding of the game than a normal fan of the cricket.

Based on the result now we can make some recommendation in fantasy cricket. Below are the teams that are forecasted before the match.

```
*****: 99.0 427.0
SR Watson[CSK]
KM Jadhav[CSK]
F du Plessis[CSK]
CH Gayle[KING]
Mandeep Singh[KING]
Imran Tahir[CSK]
Harbhajan Singh[CSK]
DL Chahar[CSK]
R Ashwin[KING]
SM Curran[KING]
KL Rahul[KING]
*****: 98.5 425.0
SR Watson[CSK]
KM Jadhav[CSK]
F du Plessis[CSK]
CH Gayle[KING]
Mandeep Singh[KING]
Imran Tahir[CSK]
SC Kuggeleijn[CSK]
Harbhajan Singh[CSK]
DL Chahar[CSK]
R Ashwin[KING]
KL Rahul[KING]
*****: 99.0 424.0
SR Watson[CSK]
KM Jadhav[CSK]
F du Plessis[CSK]
CH Gayle[KING]
Mandeep Singh[KING]
Imran Tahir[CSK]
Harbhajan Singh[CSK]
DL Chahar[CSK]
RA Jadeja[CSK]
R Ashwin[KING]
KL Rahul[KING]
*****: 99.0 422.0
```