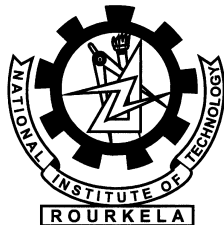


Prediction of team line-up

Project Report, CS670 Data Warehousing and Mining Lab

Sitansu Subudhi
(712CS1154)

(Under the Guidance of)
Dr. Bidyut Kumar Patra



Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela – 769 008, Odisha, India

Abstract

This project would require collecting information and statistics on the performance of different players playing in different fields of sports available on certain sports website like ESPN, StarSports, etc. Here the prediction of a team line-up for Indian cricket team against other teams is done. After collecting various performance metrics like batting average, bowling average, catches, etc. in certain number of matches last played, a prediction is made whether the player might be present for the particular line-up when the actual match is played. Initially the user is asked to input all the possible players that might be present in the team, from which the code selects the best 6 batsmen and best 5 bowlers taking into input the performance parameters.

1 Introduction

In this project, the main technique used to predict the team line-up is polynomial interpolation using Ridge expression which is used to approximate a function with polynomial n -degree. This uses the Vandermonde matrix which can be interpreted as a matrix of pseudo features. The project tries to do the non-linear regression using linear and quadratic model. The error estimate found out for different models uses the mean squared error technique.

2 Methodology

- **Dataset Collection:** Datasets are collected from different sources for different teams. The dataset contains different player combinations and their winning %age against different teams. About 500 rows each having a combination of 6 batsmen and 5 bowlers are stored in the dataset.
- **Creating a Dictionary:** Two separate dictionaries each for batting and bowling are made which contain all the performance metrics of each individual present in the dataset.
- **Splitting the Dataset:** The dataset collected is split into 80% for training set and 20% for test set. 400 records for training set and 100 for test set here.
- **Model creation:** Several models which will be used in predicting the team are created for different performance parameters(avg, centuries, economy, wickets, runs, etc.), each of which take arrays containing the respective performance values of the players.
- **Training:** The training of the dataset is carried out by using the polynomial interpolation technique using Ridge expression using the degrees 1 and 2, i.e, linear and quadratic interpolation. Here the model fitting takes place.

- **Model prediction:** The model prediction is carried out by running the linear and quadratic regression technique on the test dataset.
- **Error Calculation:** The error is calculated using the mean square error where the difference between the actual winning %age of the team against other team and the predicted winning %age is taken into account. Different errors for all the models are found out respectively.
- **Team Prediction:** One by one different models for difference performance metrics are applied and a minimum requirement for each model needs to be satisfied for a particular person to be present in the team.

Here, for a batsman, we take into account the average, runs scored, centuries, half-centuries, notouts and highest score for his inclusion, whereas, for bowlers, we have taken into account the bowling average, wickets, economy rate, wickets per inning ratio, number of 5 wicket and 10 wicket halls. Here, the project assumes that a player might either be a bowler or a batsman, and not both and accordingly his selection takes place.

3 OUTPUT

Here are some of the output screenshots attached for the simulation of the program for the prediction of the Indian team line-up when they play against Australia.

```
team prediction starts
*****
*****
batsmen
('CPS Chauhan', 'MAK Pataudi', 'MS Dhoni', 'S Abid Ali', 'KS More', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'MS Dhoni', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'MS Dhoni', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Ramesh', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MS Dhoni', 'S Ramesh', 'S Abid Ali', 'KS More', 'AL Wadekar')
('CPS Chauhan', 'MS Dhoni', 'S Ramesh', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MS Dhoni', 'S Ramesh', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MS Dhoni', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'MS Dhoni', 'S Ramesh', 'S Abid Ali', 'KS More', 'AL Wadekar')
('MAK Pataudi', 'MS Dhoni', 'S Ramesh', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'MS Dhoni', 'S Ramesh', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'MS Dhoni', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MS Dhoni', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
bowler
('PP Ojha', 'NS Yadav', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth')
('PP Ojha', 'NS Yadav', 'S Venkataraghavan', 'Maninder Singh', 'I Sharma')
('PP Ojha', 'NS Yadav', 'S Venkataraghavan', 'Maninder Singh', 'C Sharma')
('PP Ojha', 'NS Yadav', 'S Venkataraghavan', 'S Sreesanth', 'C Sharma')
('PP Ojha', 'NS Yadav', 'S Venkataraghavan', 'I Sharma', 'C Sharma')
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'S Sreesanth', 'I Sharma')

batsmen
('CPS Chauhan', 'MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Ramesh', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MS Dhoni', 'S Ramesh', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MS Dhoni', 'S Ramesh', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MS Dhoni', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'MS Dhoni', 'S Ramesh', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'MS Dhoni', 'S Ramesh', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'MS Dhoni', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MS Dhoni', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
bowler
('PP Ojha', 'NS Yadav', 'S Venkataraghavan', 'Maninder Singh', 'I Sharma')
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'I Sharma', 'C Sharma')
('PP Ojha', 'NS Yadav', 'S Sreesanth', 'I Sharma', 'C Sharma')
('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth', 'C Sharma')
('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'I Sharma', 'C Sharma')
('PP Ojha', 'S Venkataraghavan', 'S Sreesanth', 'I Sharma', 'C Sharma')
('PP Ojha', 'Maninder Singh', 'S Sreesanth', 'I Sharma', 'C Sharma')
*****
```

```

*****
batsmen
('CPS Chauhan', 'MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Ramesh', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'MAK Pataudi', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('CPS Chauhan', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MS Dhoni', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
bowler
('PP Ojha', 'NS Yadav', 'S Venkataraghavan', 'Maninder Singh', 'I Sharma')
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
*****
batsmen
('CPS Chauhan', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
('MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
bowler
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
*****
batsmen
('MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
bowler
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
*****

('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
*****
batsmen
('MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
bowler
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
('PP Ojha', 'S Venkataraghavan', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
*****
batsmen
('MAK Pataudi', 'S Ramesh', 'S Abid Ali', 'KS More', 'SR Tendulkar', 'AL Wadekar')
bowler
('PP Ojha', 'NS Yadav', 'Maninder Singh', 'S Sreesanth', 'I Sharma')
*****

```

The final team line-up

4 Error Estimation

And here are the attached screenshots of the error calculations of the program simulation done above:

```
Error Estimate Batsmen avg = 0.287385231126
Error Estimate Bowler avg = 0.287141468425
Error Estimate Batsmen runs = 0.289968945251
Error Estimate Bowler wickets = 0.286202052882
Error Estimate Batsmen centuries = 0.289479339324
Error Estimate Bowler economy = 0.290286387045
Error Estimate Batsmen half-centuries = 0.290446300586
Error Estimate Bowler wickets per innings = 0.286360374294
Error Estimate Batsmen not-outs = 0.291118652805
Error Estimate Bowler 5-wicket halls = 0.287213598503
Error Estimate Batsmen highest score = 0.288928498134
Error Estimate Bowler 10-wicket halls = 0.286436157732
Error Estimate Batsmen avg = 0.299043763085
Error Estimate Bowler avg = 0.298373389193
Error Estimate Batsmen runs = 0.303413988849
Error Estimate Bowler wickets = 0.292000355544
Error Estimate Batsmen centuries = 0.298445045596
Error Estimate Bowler economy = 0.290152357883
Error Estimate Batsmen half-centuries = 0.307493183501
Error Estimate Bowler wickets per innings = 0.287187850283
Error Estimate Batsmen not-outs = 0.287484818653
Error Estimate Bowler 5-wicket halls = 0.293702954133
Error Estimate Batsmen highest score = 0.293042112589
Error Estimate Bowler 10-wicket halls = 0.300910590045
```

5 Datasets

Here are some screenshots of the datasets used.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Player	Country	Matches	Innings	Not Out	Runs	Highest Sc	Average	Centuries	Half Centu	Ducks	World	HighNotO	Duckratio	Player ID	
2	A Kumble	India	132	173	32	2506	110	17.77	1	5	17	0	1	0.128788	22	
3	AD Gaekw	India	40	70	4	1985	201	30.07	2	10	4	0	0	0.1	87	
4	AL Wadek	India	37	71	3	2113	143	31.07	1	14	7	0	0	0.189189	191	
5	CA Pujara	India	24	42	4	1872	206	49.26	6	5	1	0	1	0.041667	407	
6	CPS Chau	India	40	68	2	2084	97	31.57	0	16	6	0	0	0.15	525	
7	DB Vengsi	India	116	185	22	6868	166	42.13	17	35	15	0	0	0.12931	597	
8	DN Sardes	India	30	55	4	2001	212	39.23	5	9	4	0	0	0.133333	669	
9	ED Solkar	India	27	48	6	1068	102	25.42	1	6	1	0	0	0.037037	741	
10	FM Engine	India	46	87	3	2611	121	31.08	2	16	7	0	0	0.152174	855	
11	G Gambhi	India	56	100	5	4046	206	42.58	9	21	6	0	0	0.107143	881	
12	GR Viswar	India	91	155	10	6080	222	41.93	14	35	10	0	0	0.10989	1027	
13	Harbhajar	India	101	142	22	2202	115	18.35	2	9	19	0	0	0.188119	1073	
14	IK Pathan	India	29	40	5	1105	102	31.57	1	6	5	0	0	0.172414	1183	
15	J Srinath	India	67	92	21	1009	76	14.21	0	4	13	0	0	0.19403	1248	
16	K Srikkant	India	43	72	3	2062	123	29.88	2	12	7	0	0	0.162791	1461	
17	KS More	India	49	64	14	1285	73	25.7	0	7	7	0	0	0.142857	1537	
18	M Amarna	India	69	113	10	4378	138	42.5	11	24	12	0	0	0.173913	1616	
19	M Azharu	India	99	147	9	6215	199	45.03	22	21	5	0	0	0.050505	1617	
20	M Prabha	India	39	58	9	1600	120	32.65	1	9	3	0	0	0.076923	1635	
21	M Vijay	India	27	47	0	1706	167	36.29	4	6	2	0	0	0.074074	1639	
22	MAK Pata	India	46	83	3	2793	203	34.91	6	16	7	0	1	0.152174	1659	
23	MS Dhoni	India	88	140	15	4808	224	38.46	6	33	9	0	0	0.102273	1810	
24	N Kapil D	India	131	184	15	5248	169	31.05	9	27	16	0	0	0.123123	1817	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Player	Country	Matches	Innings	Balls	Runs	Wickets	Average	Economy	Strike Rat	5 W I	10 W M	world	BestB Inni	BestBinnii	Wickt/Inn	ball/innin	Runs/inni	Player ID	
2	A Kumble	India	132	236	40850	18355	619	29.65	2.69	65.9	35	8	0	0.135135	0.09396	2.622881	173.0932	77.77542	22	
3	AB Agarka	India	26	46	4857	2745	58	47.32	3.39	83.7	1	0	0	0.146341	0.05	1.26087	105.587	59.67391	51	
4	BKV Prasa	India	33	58	7041	3360	96	35	2.86	73.3	7	1	0	0.181818	0.065359	1.655172	121.3966	57.93103	313	
5	BS Bedi	India	67	118	21364	7637	266	28.71	2.14	80.3	14	1	0	0.071429	0.051546	2.254237	181.0508	64.72034	333	
6	BS Chandr	India	58	97	15963	7199	242	29.74	2.7	65.9	16	2	0	0.101266	0.115385	2.494845	164.567	74.21649	334	
7	C Sharma	India	23	39	3470	2163	61	35.45	3.74	56.8	4	1	0	0.103448	0.053191	1.564103	88.97436	55.46154	354	
8	DR Doshi	India	33	55	9322	3502	114	30.71	2.25	81.7	6	0	0	0.058824	0.07767	2.072727	169.4909	63.67273	587	
9	EAS Prasai	India	49	86	14353	5742	189	30.38	2.4	75.9	10	2	0	0.105263	0.078571	2.197674	166.8953	66.76744	644	
10	Harbhajar	India	101	186	28293	13372	413	32.37	2.83	68.5	25	5	0	0.095238	0.069124	2.22043	152.1129	71.89247	933	
11	I Sharma	India	58	101	11715	6542	178	36.75	3.35	65.8	6	1	0	0.094595	0.092593	1.762376	115.9901	64.77228	1009	
12	IK Pathan	India	29	54	5884	3226	100	32.26	3.28	58.8	7	2	0	0.118644	0.095238	1.851852	108.963	59.74074	1034	
13	J Srinath	India	67	121	15104	7196	236	30.49	2.85	64	10	1	0	0.093023	0.098485	1.950413	124.8264	59.47107	1092	
14	KD Ghavri	India	39	69	7036	3656	109	33.54	3.11	64.5	4	0	0	0.151515	0.058333	1.57971	101.971	52.98551	1301	
15	M Prabha	India	39	68	7475	3581	96	37.3	2.87	77.8	3	0	0	0.045455	0.065217	1.411765	109.9265	52.66176	1436	
16	Maninder	India	35	52	8218	3288	88	37.36	2.4	93.3	3	2	0	0.259259	0.093458	1.692308	158.0385	63.23077	1460	
17	N Kapil D	India	131	227	27740	12867	434	29.64	2.78	63.9	23	2	0	0.108434	0.075342	1.911894	122.2026	56.68282	1620	
18	NS Yadav	India	35	61	8360	3580	102	35.09	2.56	81.9	3	0	0	0.065789	0.067797	1.672131	137.0492	58.68852	1679	
19	PP Ojha	India	24	48	7633	3420	113	30.26	2.68	67.5	7	1	0	0.12766	0.11236	2.354167	159.0208	71.25	1771	
20	R Ashwin	India	21	39	6330	3066	107	28.65	2.9	59.1	9	2	0	0.067961	0.141176	2.74359	162.3077	78.61538	1792	
21	RJ Shastri	India	80	125	15751	6185	151	40.96	2.35	104.3	2	0	0	0.066667	0.044693	1.208	126.008	49.48	1902	
22	S Madan L	India	39	63	5997	2846	71	40.08	2.84	84.4	4	0	0	0.217391	0.12766	1.126984	95.19048	45.1746	1993	
23	S Sreesan	India	27	50	5419	3271	87	37.59	3.62	62.2	3	0	0	0.125	0.080808	1.74	108.38	65.42	2005	
24	S Venkate	India	57	86	14873	6524	155	36.88	2.87	85.3	3	1	0	0.131111	0.078047	1.625	154.658	58.6035	2006	

6 Conclusions

The main goal of this project was to introduce a model which would predict the best team line-up to play against a particular team. Several models were developed each for batsmen and bowlers using performance features of the individual players calculated from past games. Ridge regression with polynomial interpolation algorithm is used to predict the team line-up with the best winning possibility. Most of the error estimate was around 28-30%. There may be some other high degree of polynomial interpolation used to decrease the error estimate below 20% for which the simulation of the program will take a longer time to give the output which might not be feasible for a stand-alone system like our laptops.

References

- [1] V.V Sankaranarayanan, J. Sattar, L.V.S Laksmanan, *Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction*
- [2] M. Bailey and S. R. Clarke, *Predicting the match outcome in one-day international cricket matches, while the game is in progress*, Journal of sports Science and Medicine, 5(4):480{487, 2006.