



Project Report

CSCI 6406 - Visualization

Rakshith Shetty

B00782671

Abstract:

The visualization tools have started playing an important role in the field of sports for analyzing statistics of players. Decoding-IPL is a visualization tool that has been implemented in this project. It provides the statistics of the players in the Indian Premier League using three graphs Parallel coordinates, Radial graph, and Map. Parallel co-ordinates graph provides the statistics of each player across all ten seasons from 2008-2017. Multiple players can be selected for comparison which would help the franchisees to bid for the players during the auction. The radial graph provides the aggregate statistics of the selected player for ten seasons. This would prove if the players were consistent in the league. The map provides the ground details. It highlights the city where the maximum number of matches were played by a selected player. This information can be used by the players to improve their game and coaches can use it to analyze a player and provide him some insights. During auction franchisees or a team owner can use this information to bid for the players.

1.Introduction:

Cricket a gentlemen's game was initially adopted by royal families in India. The game was a part of British legacy in India which India has revered as the game of Kings. Initially, it was popular only with the elite in India but later everybody started playing. But the strong push to the popularity was received in 1983 when Kapil Dev won the world cup for India and the introduction of color TV transmission helped the cause. This was one of the strongest factors that made cricket so popular in India. The popularity and glam associated with the game of cricket never looked back again.[1]

Indian Premier League (IPL) is an Indian professional T20 cricket league established in 2008 which is based on a round-robin group and knockout format. It has major teams in major Indian cities and each player is auctioned every four years. IPL is the one of the most attended sports league in the world and is considered to be the brainchild of the Board of Control for Cricket in India (BCCI). It has developed into the most lucrative and most popular outlet for the game of cricket. The matches were played on a home-and-away basis between all the teams. The four teams that are at the top of the tables in the series contest three playoff matches. With the arrival of IPL, almost overnight the world's best cricketers became millionaires. Few of the franchises are the Mumbai Indians, the Chennai Super Kings, and the Royal Challengers Bangalore.[2]

Cricket is a sophisticated game where all the minute details will have a lasting impact on the outcome of a game or even the tournament. These vital details must be studied and understood. This is where Decoding-IPL comes into the picture. It provides the statistics of a player which can be used by either players, coaches or franchisees. Players can use this information to assess his game whereas coaches can provide useful insights to the players. It also helps the team owners to bid a player during the auction.

The rest of the report is organized as follows: Section 2 provides information on related work. Section 3 presents the data pre-processing. Section 4 presents the implementation. Section 5 presents the limitation and future work. Section 6 presents the dependencies. Section 7 is the conclusion.

2.Related Work:

Many cricket visualization systems have been developed to analyze the players and provide the statistics of the cricket players. Pranavan Somaskandhan et al. [3] has done a research which helps to identify an optimal set of attributes. These attributes impose a high impact on the results of a cricket

match. These optimal set of attributes will help the team owners to look for players with these attributes to form a formidable team. It relies on statistical analysis and machine learning while minimizing the use of domain knowledge. The attribute set that has the highest accuracy in the evaluation is identified. These attributes will be the optimal set of attributes that impose a high impact on the end result of a cricket match.[3]

Utsav et al.[4] utilized association rule mining technique on Indian players to make a set of players for the cricket team. Player's statistics are analyzed to identify the best team. The factors such as grounds, venue, strike rate, runs scored, ball-faced etc are considered. The results of this study are useful for the Indian selector to choose a winning team.[4]

Ananthapadmanabha et al.[5] researched about match-fixing in a cricket match using the network concepts like graph mining, link mining, community mining and algorithmic design notation like greedy technique. This paper speaks about match-fixing using network analysis.[5]

Haseeb Ahmad et al.[6] employed machine learning techniques to identify the rising stars in cricket domain. Generative and discriminative machine learning algorithms were used for classification purpose. Two models from each category are evaluated. Cross-validation is employed to demonstrate high accuracy for top ten rising cricketers. [6]

N. Harikrishna et al.[7] presented a novel approach for classification of events in cricket videos and summarize its visual content. It segments a cricket video into shots and the visual content is identified in those segments. Sequential pattern mining and support vector machine are used to classify the sequence of shots into four events, namely RUN, FOUR, SIX and OUT. Based on user-supplied parameters the cricket video is summarized.[7]

For visualization of high-dimensional data, Parallel Coordinates is the most commonly used tool. Li Hui et al.[8] speaks about segmental offset-mapping parallel coordinates. It first counts and calculates the proportion of each integer value in each dimension. To express the distribution of data in each dimension stacked coordinate axis was adopted and to obtain corresponding points in axes offset-mapping technology was used. [8].

To improve the visualization of cluster information directly from the overview, Rodrigo et al.[9] proposed an edge bundling approach which is applied on parallel coordinates. A single curve between the axes has all the lines belonging to that cluster bundled into it. [9]

Keerthika Koka et al.[10] used Radial chart visualization technique to the online review classification into fake and genuine reviews. To explore the best feature selection through visualization for classification the Radial chart and the color overlaps were used. Each text review based on certain attributes were structured by the system.[10]

Yifan Zhang et al.[11] speaks about the critical visual task while using choropleth maps. According to them, identifying spatial clusters in the data is one of the critical visual tasks. Spatial clusters are visually identified when spatial units have the same color and are in the same neighborhood. Elements that lie near the classification boundary are the critical map elements in the classification scheme since these elements could belong to different classes with a slight adjustment of the classification boundary.[11]

Based on Lorenz curve and choropleth map, Hongsheng et al.[12] proposed a new visual data mining and knowledge discovery method. Spatial distribution pattern on the choropleth maps can be found and identified by the map readers using "Gestalt's Law".[12]

3.Data pre-processing:

The dataset has been obtained from the IPL official site and from Kaggle. The dataset obtained from Kaggle had two CSV files, one with a ball by ball details for all matches for all seasons and the other CSV file had IPL match details across all seasons. All these files were pre-processed into two CSV files “data.csv” which has statistics of the players from 2008-2017. and “avgData.csv” which has aggregate statistics of the players for all ten seasons. For the map, I have used the “states.json” which has all the coordinates. Both the CSV files and a JSON file is uploaded on GitHub gist so that it can be accessed by chrome and modern browsers. [13][14][15]

	A	B	C	D	E	F	G	H	I	J	K
	Players	Year	Matches	Total runs	Batting Average	Batting Strike rate	Wickets	Bowling Average	Economy	Bowling Strike rate	
2	Virat Kohli	2008	13	165	15	105.09	2	30.5	7.95	23	
3	Chris Gayle	2008	0	0	0	0	0	0	0	0	
4	Lasith Malinga	2008	0	0	0	0	0	0	0	0	
5	Suresh Raina	2008	16	421	38.27	142.22	1	30	9	20	
6	Sunil Narine	2008	0	0	0	0	0	0	0	0	
7	Sreesanth	2008	15	0	0	0	19	23.26	8.63	16.15	
8	Shane Watson	2008	15	472	47.2	151.76	17	22.52	7.07	19.11	
9	Dwayne Bravo	2008	9	154	30.8	130.5	11	17.81	7.68	13.9	
10	Sachin Tendulkar	2008	7	188	31.33	106.21	0	0	0	0	
11	Sohail Tanvir	2008	11	36	12	124.14	22	12.09	6.49	11.23	
12	Virat Kohli	2009	16	289	33.53	145.45	0	0	7.66	0	
13	Chris Gayle	2009	9	0	0	0	1	144	7.64	113	
14	Lasith Malinga	2009	13	359	81.08	144.81	18	17.33	6.3	16.5	
15	Suresh Raina	2009	14	196	22.7	9	7	23.42	5.92	23.71	
16	Sunil Narine	2009	0	2	0	100	0	0	0	0	
17	Sreesanth	2009	7	523	28.5	142.85	6	32	8.34	23	
18	Shane Watson	2009	0	405	22.12	103.63	0	0	0	0	
19	Dwayne Bravo	2009	11	0	0	128.57	11	23.54	7.8	18.09	
20	Sachin Tendulkar	2009	13	505	30.8	121.08	0	0	9.66	0	
21	Sohail Tanvir	2009	0	491	22.22	183.13	0	0	0	0	
22	Virat Kohli	2010	16	307	27.9	144.81	0	0	9.37	0	
23	Chris Gayle	2010	9	9	9	9	4	38	9.5	24	
24	Lasith Malinga	2010	13	1	1	100	15	22.93	7.02	19.6	
25	Suresh Raina	2010	16	520	47.27	142.85	6	29.66	7.46	23.83	

Fig 1: data.csv has statistics of the players from 2008-2017

	Aplayers	TotalRuns	BattingAvg	BattingStrikeRate	TotalWickets	BowlingAvg	BowlingEconomy	Team	AvgScore
2	Virat Kohli	4449	37.28	129.48	4	92	8.79	Bangalore	188
3	Chris Gayle	3676	41.2	151.2	18	40.22	7.91	Bangalore	188
4	Lasith Malinga	88	5.5	88.88	154	19.01	6.86	Mumbai	180
5	Suresh Raina	4544	33.91	138.96	25	44.48	7.39	Chennai	168
6	Sunil Narine	321	15.28	158.12	96	21.46	6.34	Kolkata	138
7	Sreesanth	34	8.5	61.81	40	29.85	8.14	Punjab	182
8	Shane Watson	2638	31.03	138.47	88	27.95	7.83	Rajasthan	170
9	Dwayne Bravo	1306	23.74	128.92	122	22.48	8.13	Chennai	168
10	Sachin Tendulkar	2334	34.83	119.81	0	0	9.66	Mumbai	180
11	Sohail Tanvir	36	12	124.14	22	12.09	6.49	Rajasthan	170

Fig2: avgData.csv has aggregate statistics of the players for all ten seasons.

Fig 1 is the pre-processed dataset which has statistics of players from 2008-2017 and fig 2 is the pre-processed dataset which has aggregate statistics for all ten seasons. Data pre-processing included removing the null values. For example, when a particular player hasn't played a particular match or a particular season because of injury or lack of form, it had null values and those null values were removed.[13][14]

Few attributes that are included in this dataset are:

- Matches: Number of matches played by a player.
- Total runs: Total runs scored by a player.
- Batting Average: Batting average of a player.
- Wickets: Number of wickets taken by a player.
- Bowling Average: Ratio of number of runs conceded per wickets taken
- Economy: The average number of runs conceded per over. [13][14]

4.Implementation:

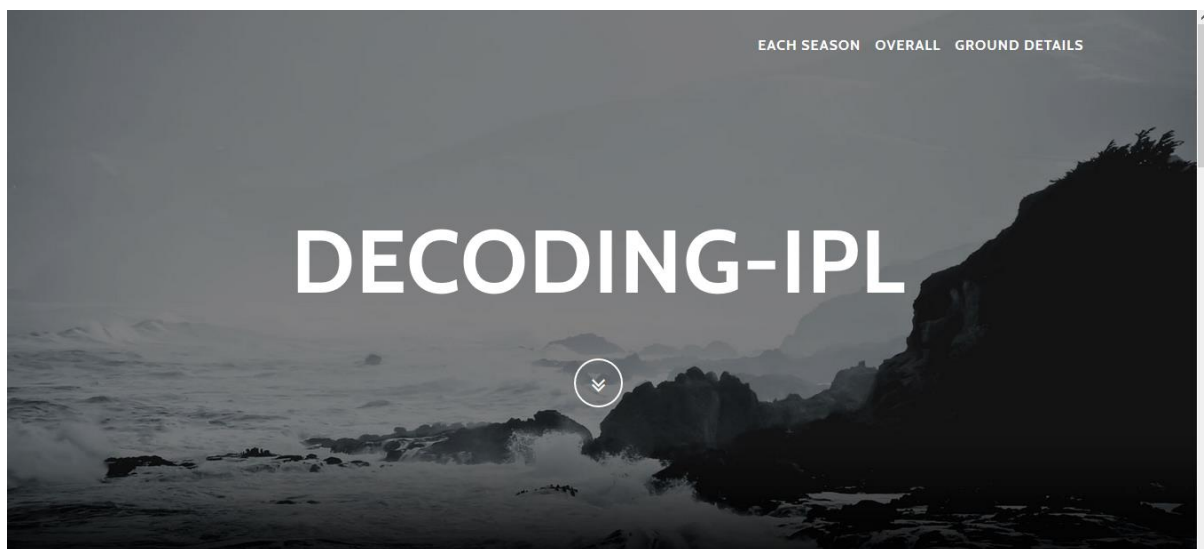


Fig3: Home page of the Decoding-IPL visualization tool

Decoding-IPL visualization tool has three graphs. Fig 3 is the home page of the Decoding-IPL visualization tool and have used the template for this. Parallel coordinates which is used for visualizing high-dimensional geometry and analyzing multivariate data is the first graph. It provides the statistics of players across all seasons from 2008-2017. Fig 4 is the parallel coordinates which display the statistics of Virat Kohli and Sohail Tanvir, Virat Kohli is represented by the color red and Sohail Tanvir is represented by the color aqua. The radial chart is the second graph which is used to display the aggregate statistics of the selected player for all ten seasons. Fig 6 is the radial graph which displays the aggregate statistics of Virat Kohli for all ten seasons. India map is the third graph which displays the ground details and highlights the city where the selected player has played the maximum number of matches as shown in fig 7. Virat Kohli has played all the ten seasons in Bangalore and hence Bangalore is highlighted. When we hover on the city it displays the ground details like ground capacity, team name, stadium name and the average runs scored by a team in that city. [16][17]

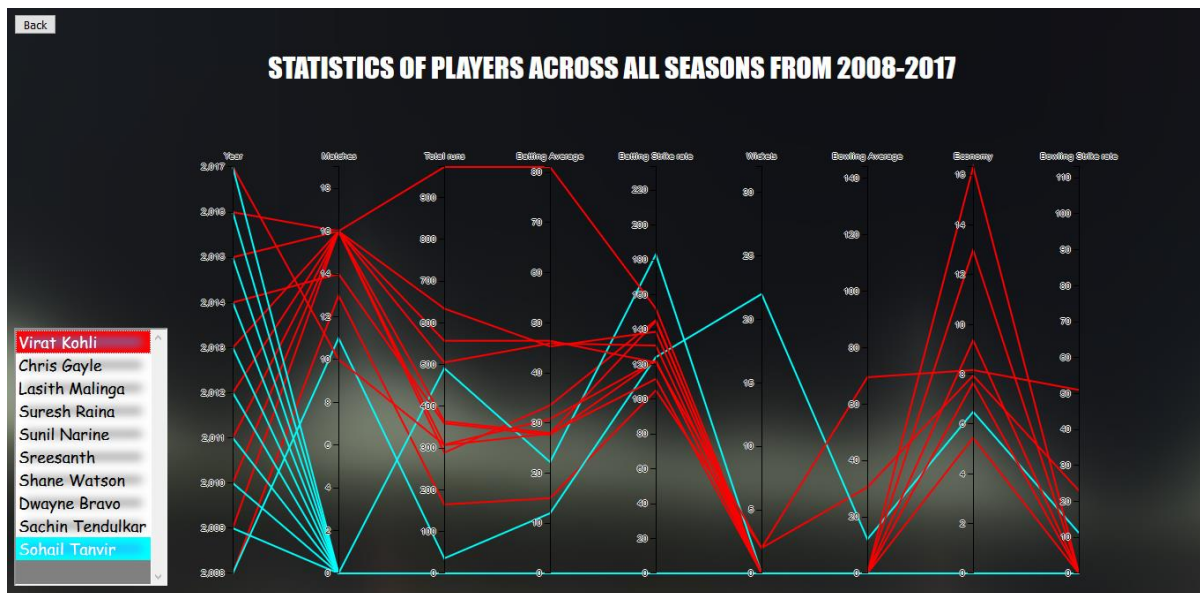


Fig4: Statistics of Virat Kohli and Sohail Tanvir across all seasons from 2008-2017

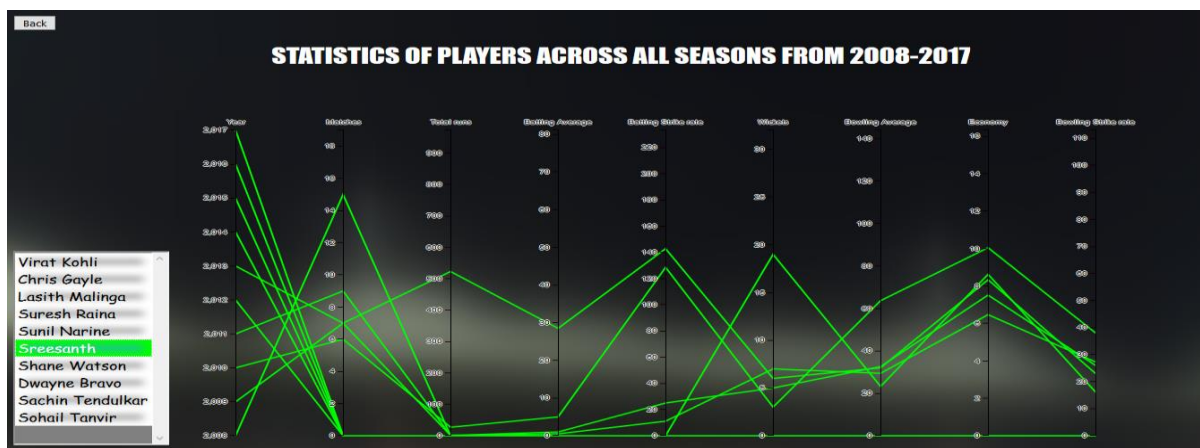


Fig5: Statistics of Sreesanth across all seasons from 2008-2017

Hypothesis:

As you can see from the graph(fig4) for Sohail Tanvir it displays data only for the year 2008 and from 2009-2017 it plots zero for all the remaining attributes since he didn't get an opportunity to play in IPL after 2008. This is because the Pakistani players weren't allowed to play in India due to the terrorist attack and political reasons. And for Sreesanth(fig5) it displays data from 2008 to 2013 and from 2014-2017 it plots zero for all the remaining attributes since he was accused of spot-fixing during 2013 season and was banned from playing IPL thereafter.

Fig 6 is the radial graph which displays the aggregate statistics of Virat Kohli for all ten seasons. Since he is purely a batsman all his batting statistics are very high. But wickets arc is very low. In cricket, lesser the bowling average and bowling economy better the bowler and since Virat Kohli is a batsman and bowls occasionally his bowling average and economy is very high.

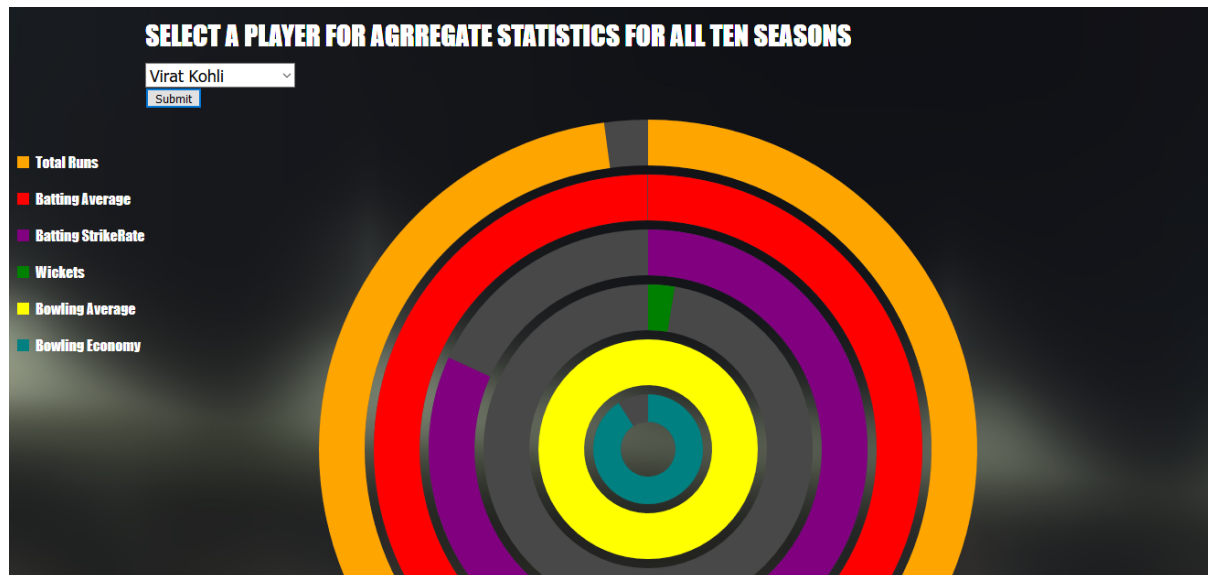


Fig6: Aggregate statistics of Virat Kohli for all ten seasons.



Fig7: India map

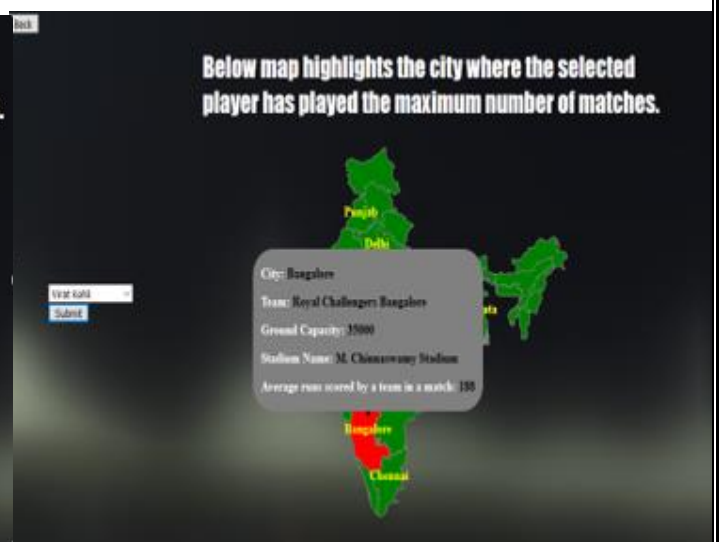


Fig8: India map with ground details.

Fig7 has Bangalore city highlighted when Virat Kohli was selected since he has played in Bangalore most of the time. It also displays the ground details in fig8. This graph provides a very interesting hypothesis. Bangalore ground is known to be a batting paradise and it clearly shows in the ground details. The ground size is really small and has a capacity of only 35000. It has the highest average runs scored by a team in a match(188). This is the reason Virat Kohli is one of the highest run-getters in IPL.

Issues faced while implementing:

Merging two csv files and collecting data from IPL site was a little challenging because not all the required attributes were found at one place. Data pre-processing consumed some time since there were players who hadn't played a match or a complete season. These had null values and had to be removed. Choosing appropriate ten players and coming up with valid hypothesis for those players was one of the challenges since it required a domain knowledge and a good background study of the IPL league.

5. Dependencies:

The various dependencies that have been utilized are as follows:

1. Data has been retrieved from Kaggle and IPL website.
2. For the map, the states.json has been used from GitHub.
3. D3.js javascript library has been used for creation of svg and the graphs.
4. Template for the homepage from startbootstrap site.
5. Default home page is index.html.
6. Internet connection is needed.
7. Libraries used are Bootstrap, jquery and googleapis.

6. Limitations and Future work:

The India map highlights the city where the player has played the maximum number matches. Although it displays the ground details using a tooltip, I would like to visually display the ground details using graphs so that it helps the users to visually analyze the data instead of reading the data in the tooltip. I have used only ten players in this visualization tool. I would like to add all the players who have played for IPL to analyze their statistics in the future work.

7.Conclusion:

Decoding-IPL is a visualization tool that is used to display the statistics of players and the ground details for the Indian premier league using three different graphs. It displays statistics of a player across all seasons from 2008-2010 using parallel coordinates. We can also choose multiple players and compare their statistics in this graph. It also highlights the city where the player has played the maximum number matches using India map and hovering on these cities provides us the ground details. The radial chart is used to display the aggregate statistics of a selected player for all ten seasons. This information can help a lot of people from team owner to coach. Team owners can use this information to bid for the player during the auction. Coaches can use this information to analyze and help players improve their game.

References:

- [1] Why Indians are so Crazy about Cricket? (2014, October 15). Retrieved April 11, 2018, from <http://greatlifezone.com/why-indians-are-so-crazy-about-cricket/>
- [2] Longmore, Andrew. "Indian Premier League." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 13 Nov. 2017, www.britannica.com/topic/Indian-Premier-League.
- [3] P. Somaskandhan, G. Wijesinghe, L. B. Wijegunawardana, A. Bandaranayake and S. Deegalla, "Identifying the optimal set of attributes that impose high impact on the end results of a cricket match using machine learning," *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, Peradeniya, 2017, pp. 1-6.
doi: 10.1109/ICIINFS.2017.8300399

- [4] U. J. Solanki and J. Vala, "Selection for balanced cricket team fourth coming ICC Championship 2017," *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, coimbatore, India, 2017, pp. 794-797.
doi: 10.1109/CESYS.2017.8321193
- [5] K. Ananthapadmanabha and K. Udayakumar, "Match fixing network analysis to verify nearness among internal participants of a cricket match," *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, 2017, pp. 1043-1048.
doi: 10.1109/RTEICT.2017.8256758
- [6] H. Ahmad, A. Daud, L. Wang, H. Hong, H. Dawood and Y. Yang, "Prediction of Rising Stars in the Game of Cricket," in *IEEE Access*, vol. 5, pp. 4104-4124, 2017.
doi: 10.1109/ACCESS.2017.2682162
- [7] N. Harikrishna, S. Satheesh, S. D. Sriram and K. S. Easwarakumar, "Temporal classification of events in cricket videos," *2011 National Conference on Communications (NCC)*, Bangalore, 2011, pp. 1-5.
doi: 10.1109/NCC.2011.5734784
- [8] H. Li, Q. He, H. Bao, N. Ma and G. Pang, "Segmental Offset-Mapping Parallel Coordinates for Multidimensional Integer Dataset," *2017 13th International Conference on Computational Intelligence and Security (CIS)*, Hong Kong, 2017, pp. 419-422.
doi: 10.1109/CIS.2017.00097
- [9] R. S. D. A. D. Lima, C. G. R. D. Santos and B. S. Meiguins, "A Visual Representation of Clusters Characteristics Using Edge Bundling for Parallel Coordinates," *2017 21st International Conference Information Visualisation (IV)*, London, 2017, pp. 90-95.
doi: 10.1109/iv.2017.29
- [10] K. Koka and S. Fang, "Online Review Analysis by Visual Feature Selection," *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, Orlando, FL, USA, 2017, pp. 1084-1091.
doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.176
- [11] Y. Zhang and R. Maciejewski, "Quantifying the Visual Impact of Classification Boundaries in Choropleth Maps," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 371-380, Jan. 2017.
doi: 10.1109/TVCG.2016.2598541
- [12] Hongsheng Li and Jiafu Han, "Discovery of population distribution knowledge visually through Lorenz curve and choropleth map," *The 2nd International Conference on Information Science and Engineering*, Hangzhou, China, 2010, pp. 3657-3660.
doi: 10.1109/ICISE.2010.5690467
- [13] Manas. Indian Premier League | Kaggle. November 23, 2017. Accessed March 14, 2018.
<https://www.kaggle.com/manasgarg/ipl/data/>
- [14] "Indian Premier League Official Website." IPLT20.com - Indian Premier League Official Website. Accessed March 14, 2018. <http://www.iplt20.com/>.

[15] J. (n.d.). India Map in D3. Retrieved April 11, 2018, from <https://gist.github.com/JohnCoogan/1531818>

[16] Bostock, M. (n.d.). Data-Driven Documents. Retrieved April 11, 2018, from <https://d3js.org/>

[17] B. (n.d.). Grayscale - Free One Page Bootstrap Theme. Retrieved April 11, 2018, from <https://startbootstrap.com/template-overviews/grayscale/>

[18] "Free Image on Pixabay - Cricket, Sport, New, Zealand." Free photo: Cricket, Sport, New, Zealand - Free Image on Pixabay - 2511043. Accessed March 14, 2018. <https://pixabay.com/en/cricket-sport-new-zealand-2511043/>