Centre for Analysis, Research and Visualisation (C-MANAV)

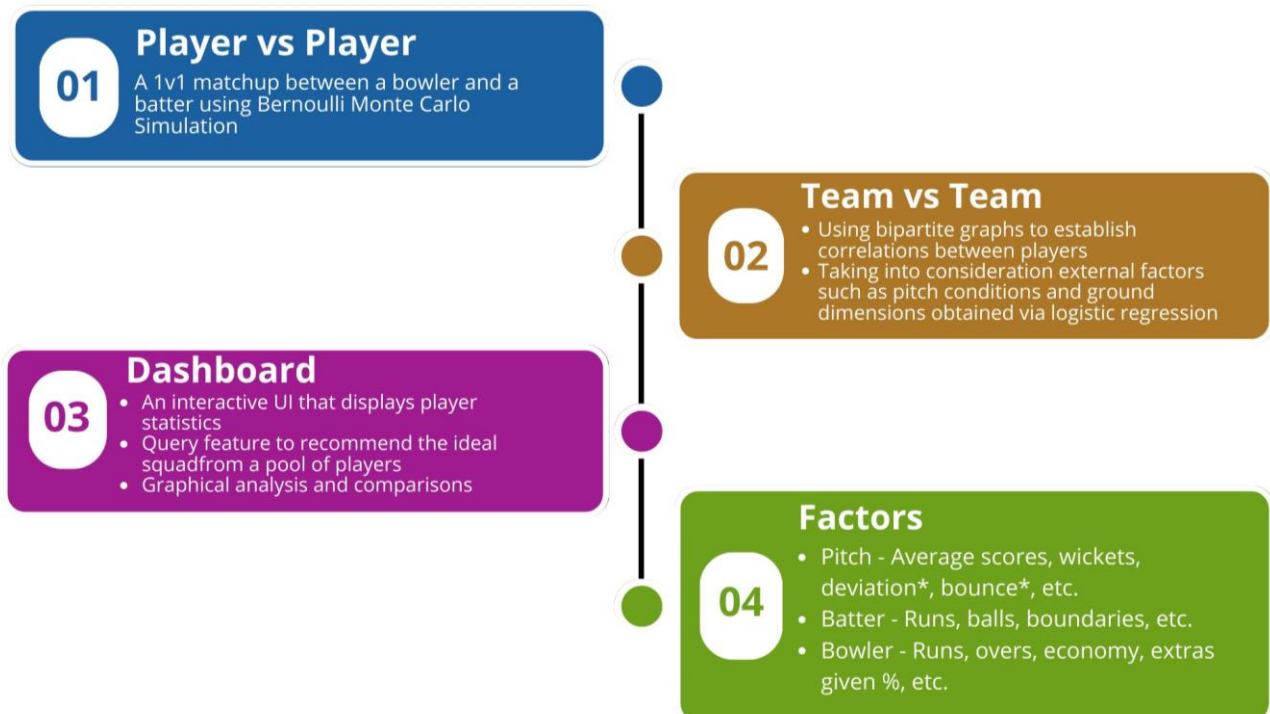## STARC -Summer Program 2023

Spectoclava – Cricket Matchup Dashboard with Recommender and Predictor

Abstract:

Cricket, a captivating sport with its intricate dynamics, is shaped by a multitude of factors, both known and unknown such as pitch conditions, player performance (recent form), team dynamics, and historical performance data, that play a significant role in determining match outcomes. We aim to explore the various conditions that affect the game of cricket and propose a framework using a dashboard to predict matchups between players and teams.

Framework
Spectoclava – Matchup and Recommender Dashboard

Player vs Player

01 A 1v1 matchup between a bowler and a batter using Bernoulli Monte Carlo Simulation

Team vs Team

02
- Using bipartite graphs to establish correlations between players
- Taking into consideration external factors such as pitch conditions and ground dimensions obtained via logistic regression

Dashboard

03
- An interactive UI that displays player statistics
- Query feature to recommend the ideal squad from a pool of players
- Graphical analysis and comparisons

Factors

04
- Pitch - Average scores, wickets, deviation*, bounce*, etc.
- Batter - Runs, balls, boundaries, etc.
- Bowler - Runs, overs, economy, extras given %, etc.

This framework will aid in making informed decisions to form the most efficient and competitive 'Playing Eleven' for tournaments, considering the diverse factors mentioned above.

The model utilizes a comprehensive approach that incorporates statistical analysis, data mining, and machine learning techniques to analyse and interpret the vast amount of available cricket data. By considering the aforementioned factors, the proposed model seeks to uncover hidden patterns using data manipulation to generate features that contribute to the game's outcome.

This framework provides a reliable foundation for predicting player matchups, enabling teams and selectors to make strategic decisions while selecting the 'Playing Eleven'. This will help by identifying key strengths and weaknesses of individual players and assessing their performance against specific opponents, teams can maximize their chances of success.

The Dataset

The dataset for each player is scraped from ESPN Cricinfo throughout their career (2000-2023*) which includes statistics against teams and year-wise statistics. This is done separately for each bowler and batter. The data for each pitch is also scraped and divided into first and second innings. The statistics for each player vs player and team vs team is scraped separately due to the scattered nature of the dataset.

The datasets are cleaned by removing unwanted columns, and the data types of numerical values are converted from strings after removing unwanted characters at play. The raw data is hence converted into the desired format for it to be utilised effectively.

Bowler Data

Overs	Mdns	Runs	Wkts	Econ	Pos	Inns	Opposition	Ground	Start Date	MatchNo	Balls	player
10.0	1	44.0	2.0	4.40	2	1	v India	Wankhede	14 Jan 2020	ODI # 4231	60.0	Pat Cummins
10.0	0	53.0	1.0	5.30	4	1	v India	Wankhede	14 Jan 2020	ODI # 4231	60.0	Adam Zampa
6.0	0	37.0	0.0	6.16	5	2	v India	Wankhede	17 Mar 2023	ODI # 4538	36.0	Adam Zampa
7.0	1	27.0	2.0	3.85	2	2	v India	Wankhede	17 Mar 2023	ODI # 4538	42.0	Marcus Stoinis
10.0	0	56.0	3.0	5.60	1	1	v India	Wankhede	14 Jan 2020	ODI # 4231	60.0	Mitchell Starc
9.5	0	49.0	3.0	4.98	1	2	v India	Wankhede	17 Mar 2023	ODI # 4538	59.0	Mitchell Starc
2.0	0	7.0	0.0	3.50	6	2	v India	Wankhede	17 Mar 2023	ODI # 4538	12.0	Glenn Maxwell
6.0	0	35.0	0.0	5.83	4	2	v India	Wankhede	17 Mar 2023	ODI # 4538	36.0	Cameron Green
10.0	1	56.0	1.0	5.60	5	1	v India	Wankhede	14 Jan 2020	ODI # 4231	60.0	Ashton Agar
9.0	0	31.0	0.0	3.44	3	2	v India	Wankhede	17 Mar 2023	ODI # 4538	54.0	Sean Abbott

Batter Data

	Runs	Mins	BF	4s	6s	SR	Pos	Dismissed	Inns	Opposition	Ground	Start Date	MatchNo	Batter
0	0.0	3	1.0	0	0	0.00	4	1	2	v Australia	Wankhede	17 Mar 2023	ODI # 4538	Suryakumar Yadav
1	13.0	-	10.0	2	0	130.00	8	1	1	v Australia	Wankhede	14 Jan 2020	ODI # 4231	Shardul Thakur
2	20.0	57	31.0	3	0	64.51	2	1	2	v Australia	Wankhede	17 Mar 2023	ODI # 4538	Shubman Gill
3	17.0	-	15.0	2	0	113.33	10	1	1	v Australia	Wankhede	14 Jan 2020	ODI # 4231	Kuldeep Yadav
4	16.0	-	14.0	0	1	114.28	4	1	1	v Australia	Wankhede	14 Jan 2020	ODI # 4231	Virat Kohli
5	4.0	15	9.0	1	0	44.44	3	1	2	v Australia	Wankhede	17 Mar 2023	ODI # 4538	Virat Kohli
6	25.0	-	32.0	2	1	78.12	7	1	1	v Australia	Wankhede	14 Jan 2020	ODI # 4231	Ravindra Jadeja
7	45.0	101	69.0	5	0	65.21	7	0	2	v Australia	Wankhede	17 Mar 2023	ODI # 4538	Ravindra Jadeja
8	3.0	11	8.0	0	0	37.50	1	1	2	v Australia	Wankhede	17 Mar 2023	ODI # 4538	Ishan Kishan
9	25.0	48	31.0	3	1	80.64	6	1	2	v Australia	Wankhede	17 Mar 2023	ODI # 4538	Hardik Pandya
10	10.0	-	15.0	2	0	66.66	1	1	1	v Australia	Wankhede	14 Jan 2020	ODI # 4231	RG Sharma
11	10.0	-	15.0	1	0	66.66	9	1	1	v Australia	Wankhede	14 Jan 2020	ODI # 4231	Mohammed Shami
12	0.0	-	0.0	0	0	-	11	0	1	v Australia	Wankhede	14 Jan 2020	ODI # 4231	Jasprit Bumrah

Pitches Data

	Name	Won batting first	Won batting second	Avg 1st Inngs Score	Std Dev - 1st Inngs Score	Avg 1st Inngs Wkts	Std Dev - 1st Inngs Wkts	Avg 2nd Inngs Score	Std Dev - 2nd Inngs Score	Avg 2nd Inngs Wkts	Std Dev - 2nd Inngs Wkts	Straight1	Straight2	Side1	Side2
1	Kennington Oval	18	30	257.02	60.53	4.52	3.31	226.65	55.40	3.21	3.11	68	66	67	70
2	Trent Bridge	12	14	269.96	96.53	4.33	3.69	234.59	68.15	3.30	3.30	72	64	64	64
3	County Ground Bristol	5	9	239.43	82.90	4.21	3.98	224.79	71.29	3.29	2.55	63	66	67	69
4	The Rose Bowl	14	15	255.28	66.64	4.72	3.29	216.34	70.43	2.83	2.67	82	82	76	76
5	The Cooper Associates County Ground	1	2	266.67	82.28	2.67	4.62	253.67	75.26	2.00	1.73	65	68	65	64
6	Emirates Old Trafford	15	13	237.29	74.87	4.82	3.78	203.71	66.62	2.89	3.50	70	73	65	77
7	Edgbaston	14	18	246.84	72.66	5.53	3.51	205.34	60.38	3.09	3.07	66	73	63	69
8	Headingley	8	12	266.25	55.49	6.35	2.52	248.40	44.00	3.05	2.95	70	65	72	66
9	Lord's	21	17	244.54	51.56	4.88	3.52	216.95	47.40	2.88	3.20	82	70	65	60
10	Riverside Ground	7	10	250.24	66.67	5.24	3.29	213.59	63.48	3.59	3.39	72	72	78	74
11	Sophia Gardens	8	16	239.92	72.27	3.64	3.66	213.96	55.55	3.12	3.02	60	64	74	72

The Model

The cleaned data is used to generate features using a machine learning model that affects the outcome of a match. This is done by dividing the initial data into parts such as overall stats and recent stats to get a general idea of the performance of a player, if data for the player does not exist, it is filled by utilising the data for similar players (based on role factor) and how they fair in different matchups.

The ground factors, such as ground dimensions, spin or pace friendly and batting or bowling friendly, are found using a logistic regression model. The obtained features are then used with the player factors to come up with probabilities of various outcomes that might occur in a delivery (the outcomes include: 'wicket', '0', '1', '2', '3', '4', '6', 'extras').
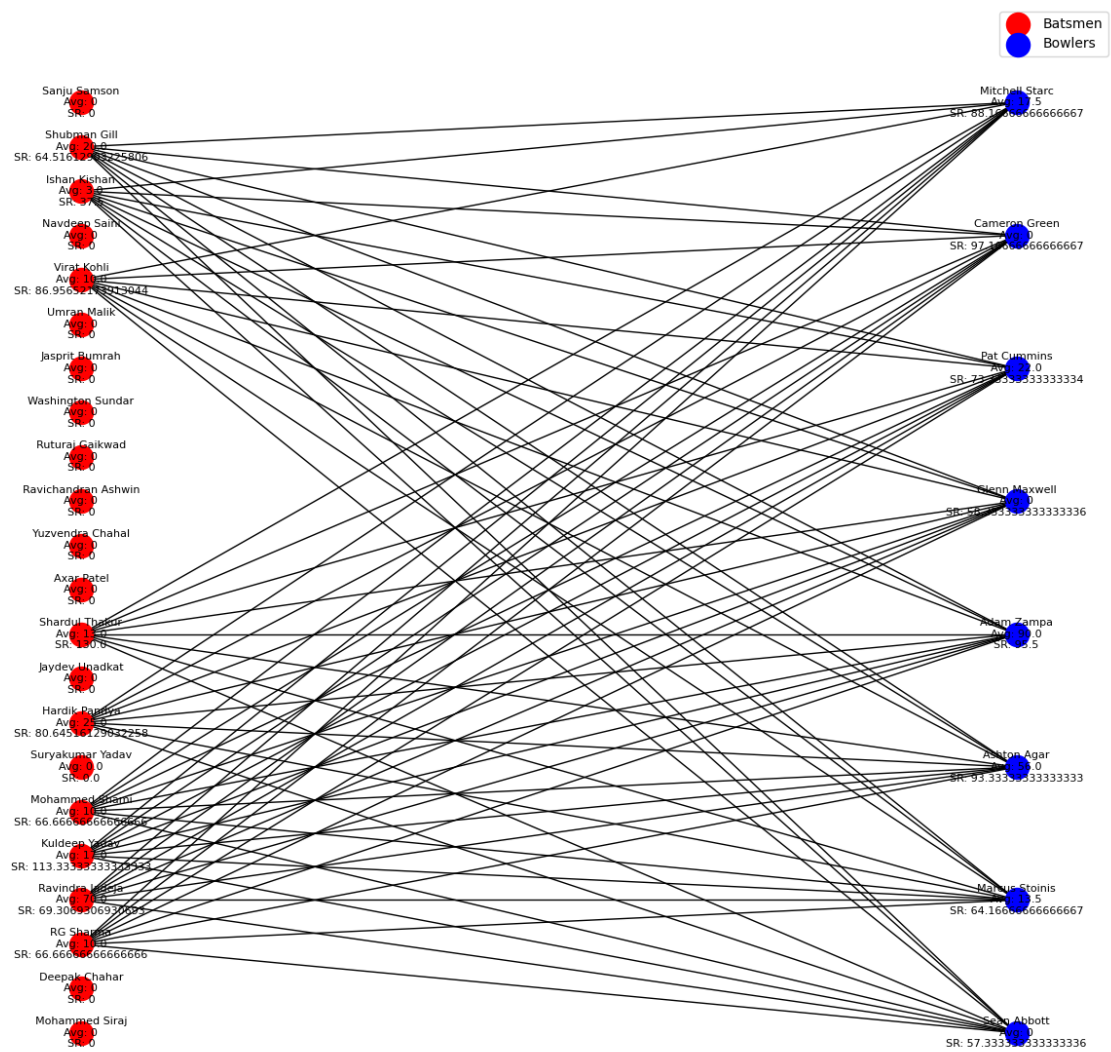
Pitch Features

	Name	Pace_Spin_Friendly \	Batting_Bowling_Friendly
0	Kennington Oval	Spin-Friendly	Batting-Friendly
1	Trent Bridge	Spin-Friendly	Batting-Friendly
2	County Ground Bristol	Spin-Friendly	Bowling-Friendly
3	The Rose Bowl	Pace-Friendly	Batting-Friendly
4	The Cooper Associates County Ground	Spin-Friendly	Batting-Friendly
5	Emirates Old Trafford	Pace-Friendly	Bowling-Friendly
6	Edgbaston	Pace-Friendly	Bowling-Friendly
7	Headingley	Pace-Friendly	Batting-Friendly
8	Lord's	Pace-Friendly	Bowling-Friendly
9	Riverside Ground	Pace-Friendly	Bowling-Friendly
10	Sophia Gardens	Spin-Friendly	Bowling-Friendly

	Name	Spin_Pace_Friendly \	Batting_Bowling_Friendly
0	Kennington Oval	0.494553	0.513462
1	Trent Bridge	0.486532	0.530783
2	County Ground Bristol	0.482488	0.474465
3	The Rose Bowl	0.504651	0.507876
4	The Cooper Associates County Ground	0.429359	0.529450
5	Emirates Old Trafford	0.508458	0.466963
6	Edgbaston	0.542821	0.487864
7	Headingley	0.613877	0.542407
8	Lord's	0.511923	0.475463
9	Riverside Ground	0.530994	0.495273
10	Sophia Gardens	0.455001	0.471840

A Bipartite graph is used to store the different probabilities and features between players. The highest probabilities of the suitable features will be used to recommend the top players, role-wise, that are most adaptable to the conditions and can tackle various members of the opponent's squad to be selected for the team allowing selectors to pick the best 'Playing Eleven' for the tournament or series.

The probabilities are further used to perform simulations of Team vs Team matchups. This is achieved using Python implementation to run ball by ball simulations of the match and give the average performance of each bowler and batter that played in the simulated matches.

Bipartite Graph

Sanju Samson
Avg: 0
SR: 0

Shubman Gill
Avg: 20.0
SR: 64.516129032258006

Ishan Kishan
Avg: 0
SR: 0

Navdeep Saini
Avg: 0
SR: 0

Virat Kohli
Avg: 0
SR: 86.95652173913044

Umran Malik
Avg: 0
SR: 0

Jasprit Bumrah
Avg: 0
SR: 0

Washington Sundar
Avg: 0
SR: 0

Ruturaj Gaikwad
Avg: 0
SR: 0

Ravichandran Ashwin
Avg: 0
SR: 0

Yuzvendra Chahal
Avg: 0
SR: 0

Axar Patel
Avg: 0
SR: 0

Shardul Thakur
Avg: 100.0
SR: 100.0

Jaydev Unadkat
Avg: 0
SR: 0

Hardik Pandya
Avg: 0
SR: 80.64516129032258

Suryakumar Yadav
Avg: 0
SR: 0.0

Mohammed Shami
Avg: 0
SR: 66.66666666666666

Kuldeep Yadav
Avg: 0
SR: 113.33333333333334

Ravindra Jadeja
Avg: 0
SR: 69.30693069306931

RG Sharma
Avg: 0
SR: 66.66666666666666

Deepak Chahar
Avg: 0
SR: 0

Mohammed Siraj
Avg: 0
SR: 0

Mitchell Starc
Avg: 5
SR: 88.16666666666667

Cameron Green
SR: 97.16666666666667

Pat Cummins
Avg: 0
SR: 77.33333333333334

Glenn Maxwell
Avg: 0
SR: 56.33333333333336

Adam Zampa
Avg: 0
SR: 63.5

Ashton Agar
Avg: 0
SR: 93.33333333333333

Marcus Stoinis
Avg: 5
SR: 64.16666666666667

Sean Abbott
SR: 57.33333333333336

Simulation of an Innings between India and Australia
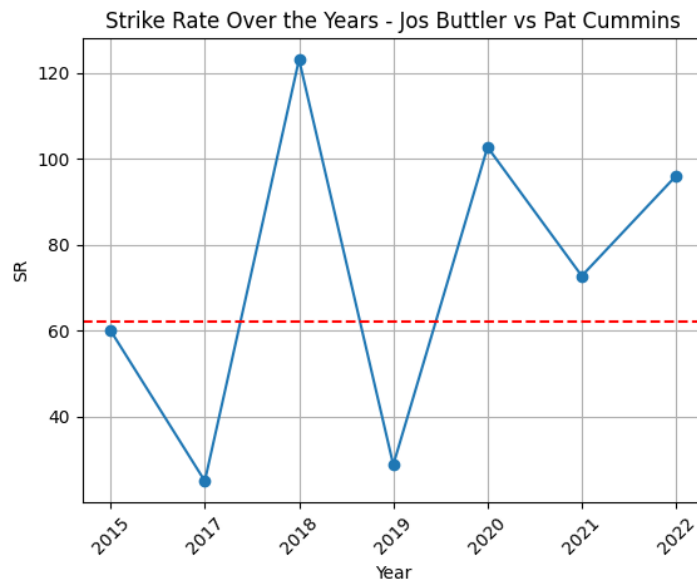
```
PS C:\Users\mdash\OneDrive\Desktop\starc> python simulate.py
Ashton Agar          8 overs        1 wickets      24 dots
Pat Cummins          10 overs       2 wickets      19 dots
Adam Zampa           9 overs        0 wickets      19 dots
Marcus Stoinis       6 overs        2 wickets      12 dots
Josh Hazlewood       9 overs        0 wickets      26 dots
Cameron Green        8 overs        1 wickets      19 dots


Rohit Sharma         99 runs        122 balls    5 4s    1 6s    Out : 37.4
Shubman Gill         8 runs         13 balls     0 4s    0 6s    Out : 7.6
Virat Kohli          30 runs        28 balls     0 4s    1 6s    Out : 16.3
Suryakumar Yadav        20 runs        23 balls      0 4s    1 6s    Out : 25.2
Hardik Pandya        50 runs        41 balls     1 4s    3 6s    Out : 38.1
Ishan Kishan         24 runs        32 balls     3 4s    0 6s    Not Out
Ravindra Jadeja         36 runs        37 balls      3 4s    1 6s    Out : 48.5
Kuldeep Yadav        7 runs         4 balls      1 4s    0 6s    Not Out
Mohammed Shami       0 runs         0 balls      0 4s    0 6s    DNB
Yuzvendra Chahal        0 runs         0 balls      0 4s    0 6s    DNB
Mohammed Siraj       0 runs         0 balls      0 4s    0 6s    DNB


 [[0, 0, 0, 0, 0, 2], [1, 0, 0, 1, 0, 0], [1, 1, 1, 1, 0, 1], [0, 4, 0, 2, 0, 1], [0, 4, 0, 1, 1, 0], [1, 0, 2
], [6, 0, 1, 1, 1, 0], [2, 1, 0, 1, 1, 1], [0, 1, 1, 1, 1, 1], [0, 1, 0, 1, 1, 1], [0, 0, 1, 2, 1, 1], [1, 4,
1], [0, 1, 0, 1, 1, 2], [1, 1, 0, 0, 0, 1], [0, 0, 1, 1, 0, 0], [0, 0, 0, 1, 0, 1], [0, 0, 0, 2, 0, 0], [0, 'w
0], [1, 0, 1, 0, 0, 1], [3, 1, 1, 0, 2, 1], [1, 2, 0, 6, 1, 2], [1, 1, 0, 0, 4, 0], [1, 6, 1, 1, 1, 1], [0, 0
1, 0], [0, 1, 1, 0, 0, 1], [2, 4, 6, 0, 1, 0], [1, 1, 1, 1, 0, 1], [0, 1, 1, 1, 0, 1], [0, 4, 0, 1, 2, 0], [0
, 1, 1]]


Total Runs: 274
Total Wickets: 6
Total overs bowled: 50/50
```

These features will be available on a dashboard for all teams and players. It will allow the user to input the squad of a team and will give the recommended team that can be formed based on the multiple parameters used in our model. The dashboard will also allow the user to perform the statistical analysis of a player or perform different matchups of their choice. It can also be used to predict upcoming matches and therefore help in building teams for fantasy leagues or games. Here is an example of the Strike Rate of Jos Buttler against Pat Cummins over the years.

Strike Rate Over the Years - Jos Buttler vs Pat Cummins

Sl. No.	Name	SRN	Individual Contribution
1)	P Sai Charan	PES2UG21CS364	• Scrape individual players data including team, from ESPN Cricinfo and organised the data into batters and bowlers. • Design the dashboard for team vs team matchups using NextJS and TailwindCSS • Authentication using Firebase and real-time access to the player statistics (axios and cheerio)
2)	Muhammad Ashar Reza	PES2UG21CS306	• Scrape grounds at world cup venues data

			• Clean all the scraped data so it can be utilised. • Implementation of matchup analysis using parameters for team vs team in python using the player vs player analysis. • Implementation of simulator for various probable player vs player and team vs team interactions using python.
3)	Pranit Prasant Pai	PES2UG21CS388	• Perform graphical analysis and assign various parameters including pitch, form, etc. to each player using Tableau and Power I • Display player vs player matchups based on user input • Prove the working of the model by analysing the ground factors in the 2019 World Cup and confirming using machine learning algorithm.
4)	Pratham R Shetty	PES2UG21CS392	• Implementation of the logistical regression model in python for ground factors to be calculated. • Implementation of the model in python for the probabilities required. • Implementation of bipartite graph to store probabilities and other factors. • Prove the working of the model by predicting the 2019 World Cup using machine learning model.

Software employed: Python libraries employing machine learning algorithms (Logistical regression), Flask, ReactJS, SQL, Tailwind CSS, Tableau

Features on Dashboard:

- Display detailed statistics of players– (overall, year-wise)
- Display ball to ball match details of all matches in ODI (can be extended to other formats)
- Player v player statistics
- Team v team statistics
- Team building to help with team management
- Predict upcoming matches
- Predict various player matchups
- Recommend players for Fantasy cricket games (such as helping in choosing the most valuable player)

Further Scope: Fill gaps in squads in Auctions (with Bid Price)

Faculty Mentors: Dr. Bharathi R and Dr. Sandesh B J