

A red cricket ball with white stitching is positioned on a green lawn. In the background, the edge of a wooden cricket bat is visible. The scene is softly lit, creating a calm and focused atmosphere.

Win Predictor for Cricket Games

Anish Deshpande

Faculty Advisor: Mr. Cruz

External Advisor: Mr. Prakash Giri

Background

- I've played cricket since the age of 9
- Gone on to play for the regional under 19 team and play cricket professionally in the Minor League for the Hollywood Master Blasters
- Scoring methods have changed over the last decade
 - Transfer from paper to online scoring

HOME CLUB	VISITOR	NUMBER OF	PLAYED AT	DATE			
ENTER IN							
IN	OUT	BALLS	RUNS SCORED	40	WICKETS	BOWLER	RUNS
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							
www.PrintablePaper.net							





CricClubs




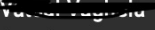













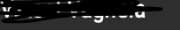
- Has an online scoring app that is used to score games in many leagues across the world, including the league that I play in
- Scores and stats can then be seen on the CricClubs website
- Keeps track of records and MVP points in order to give awards to players at the end of each tournament
- Has a live stream capability. Live score will be shown alongside YouTube live streams



My Motivation

- Lots of data is being collected, but we aren't making the best use of it
- Many ideas: commentary assistance, predictor, suggester, batsman strength and weakness finder, etc.
 - The Predictor and suggester attracted me because I have seen how much talk and debate goes into decisions about team compositions and toss decisions. I want to help solve this issue.

POINTS TABLE							GROUP : ALL	
#	Team	Mat	Won	Lost	N/R	Pts	Net RR	
1	 Davis Grizzlies	2	1	0	1	3	1.8200	
2	 BA Blazers	1	0	0	1	1	0.0000	
3	 SARE All Stars	1	0	0	1	1	0.0000	
4	 Stanford Strikers	2	0	1	1	1	-1.8200	
More Details								

BATTING		BOWLING		RANKING	
Player	Runs	Player	Wkts	Player	Points
  (DGR)	64	  (DGR)	4	  (DGR)	396
  (DGR)	34	  (SST)	4	  (SST)	199
  (SST)	27	  (SST)	2	  (DGR)	170

My Goals

- A **predictor** which would predict the winner of an upcoming game scored through the CricClubs app by using past game data
- A **suggester** that would:
 1. Suggest the optimal team compositions for both teams
 2. Suggest the team captain on whether they should bat or bowl first upon winning the toss

Significance

Widespread usage:

- CricClubs is used for scoring all across the world for thousands of different cricket leagues
- Developing a predictor / suggester for one league means that it can easily be implemented in other CricClub leagues

Predictor:

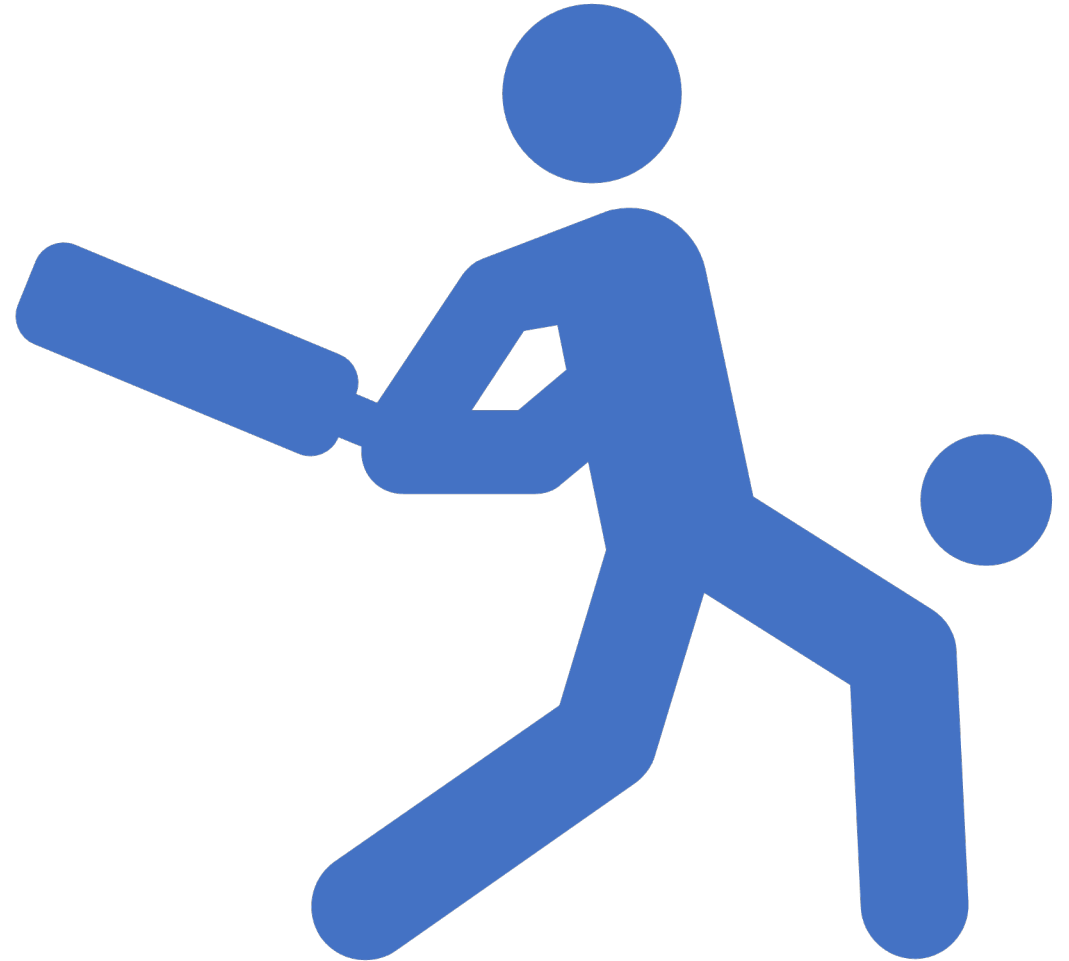
- Teams will get a sense of who is favored to win the game
- Can eventually result in a fun fantasy league system

Suggester:

- Teams will have help making important decisions before the game, such as deciding their playing 11 and deciding what to do upon winning the toss.

Methodology

- General idea that there would be certain factors (aka features) that would affect the outcome of a cricket game
- Started out by finding some of these factors that I thought would be significant
 - Then applied them to some basic formulas I drew up
 - Idea was that I would give a certain weightage to each of these factors. Then, I would play around with the weights until I ended up with my highest accuracy



PREDICTOR:

$$\text{Player Value} = \frac{\text{MVP Points this season}}{\text{\# of games played this season}}$$

$$\text{Team Value} = \frac{\Sigma \text{player values}}{\text{\# of players on team}}$$

$$\text{Relative Team Value A} = \frac{\text{Team Value A}}{\text{Team Value A or B (whichever is larger)}}$$

$$\text{Relative Team Value B} = \frac{\text{Team Value B}}{\text{Team Value A or B (whichever is larger)}}$$

(higher value of relative team value will be 1. Lower value will be between 0 and 1)

$$\text{Win Strength A} = \frac{\text{\# of wins for team A in past 10 games}}{10}$$

$$\text{Win Strength B} = \frac{\text{\# of wins for team B in past 10 games}}{10}$$

$$\text{Ground Strength A} = \frac{\text{\# of wins for team A in past 10 games at ground X}}{10}$$

$$\text{Ground Strength B} = \frac{\text{\# of wins for team B in past 10 games at ground X}}{10}$$

(ground X is the ground upon which teams A and B will be playing)

Total Strength A =

$$\beta_0 \text{Relative team value A} + \beta_1 \text{Win strength A} + \beta_2 \text{Ground strength A}$$

Total Strength B =

$$\beta_0 \text{Relative team value B} + \beta_1 \text{Win strength B} + \beta_2 \text{Ground strength B}$$

(where β represents a certain weightage given to each factor in the equation)

$$\text{Win \% A} = \frac{\text{Total Strength A}}{\text{Total Strength A} + \text{Total Strength B}} \times 100$$

$$\text{Win \% B} = 100\% - \text{Win \% A}$$

SUGGESTER:

(Values over 0.5 favor batting first. Values under 0.5 favor bowling first)

$$\text{Ground Tendency} = \frac{\text{games won batting first on ground X}}{\text{all games played on ground X}}$$

$$\text{Team tendency A} = \frac{\text{\# of games won by team A batting first}}{\text{all games won by team A}}$$

$$\text{A tendency vs B} = \frac{\text{\# of games won by team A batting first against team B}}{\text{all games played between team A and B}}$$

A tendency vs B on X =

$$\frac{\text{\# of games won by team A batting first against team B on ground X}}{\text{all games played between A and B on ground X}}$$

Total Tendency =

$$(\beta_0 \text{Ground tendency} + \beta_1 \text{Team tendency A} + \beta_2 \text{A tendency vs B} +$$

$$\beta_3 \text{A tendency vs B on X}) / 4$$

(Where β represents a certain weightage given to each factor in the equation)

If total tendency < 0.5 → bowl first

If total tendency >= 0.5 → bat first



Getting Started

- CEO of CricClubs Ganesh Nallapareddy gave me access to the CricClubs databases
 - Used SQL queries to extract appropriate data
- Imported data to R for the rest of my analysis

My External Advisor's Role

Part of the NCCA committee

Introduced me to the CEO of CricClubs

Reviewed my blogs and met weekly

Helped me code better and guided me to videos and links that could help me in my project

Helped me navigate through the CricClubs database

Progress

- Decided to focus on my predictor first
- Went away from the manual method of "playing with weights" and found some good machine learning algorithms for binary classification problems.
- The predictor involved binary classification (win = 1/ loss = 0)
- Possible algorithms to use: Xgboost, logistic regression, support vector machine
- My plan: use all these algorithms and see which model yields the highest accuracy (then tweak the most accurate model to make it even better)



Predictor

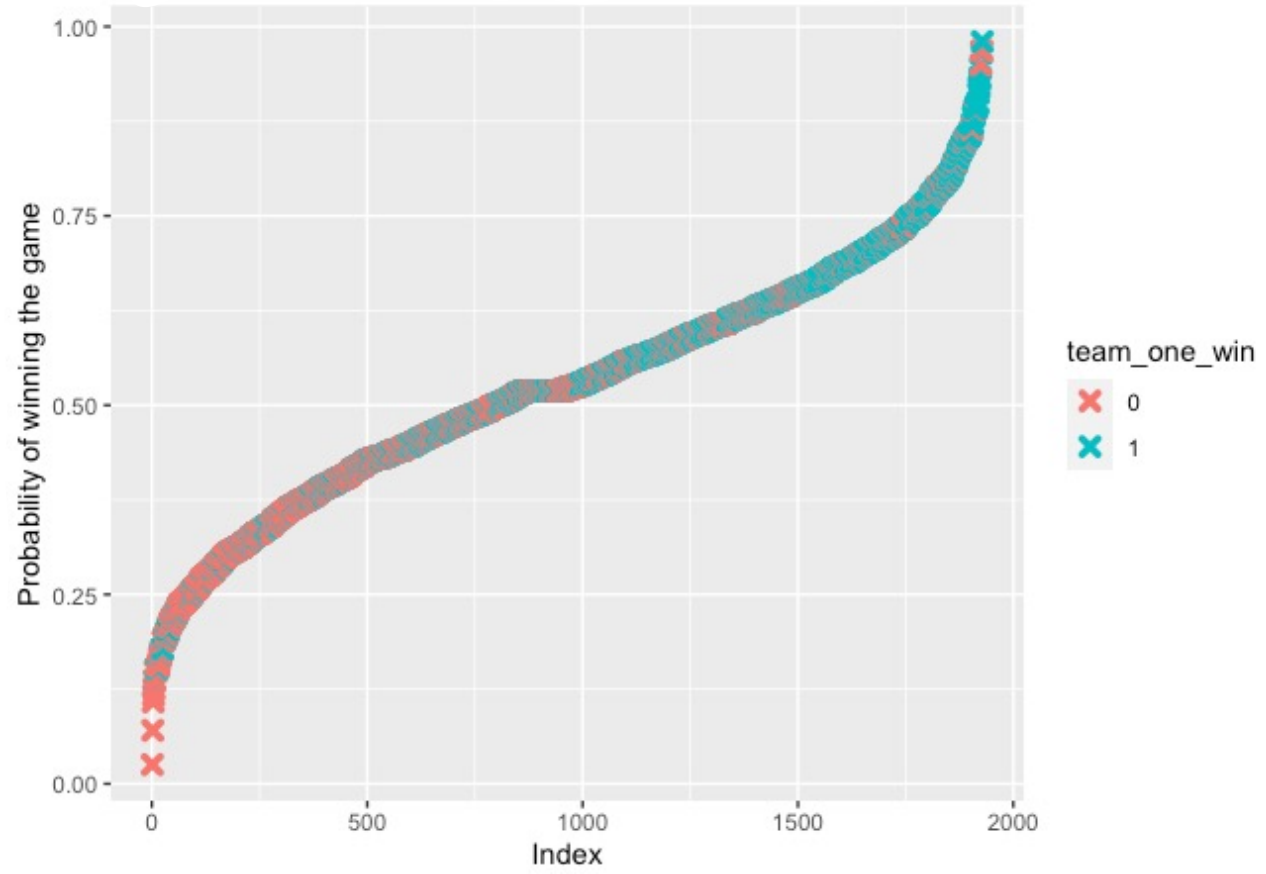
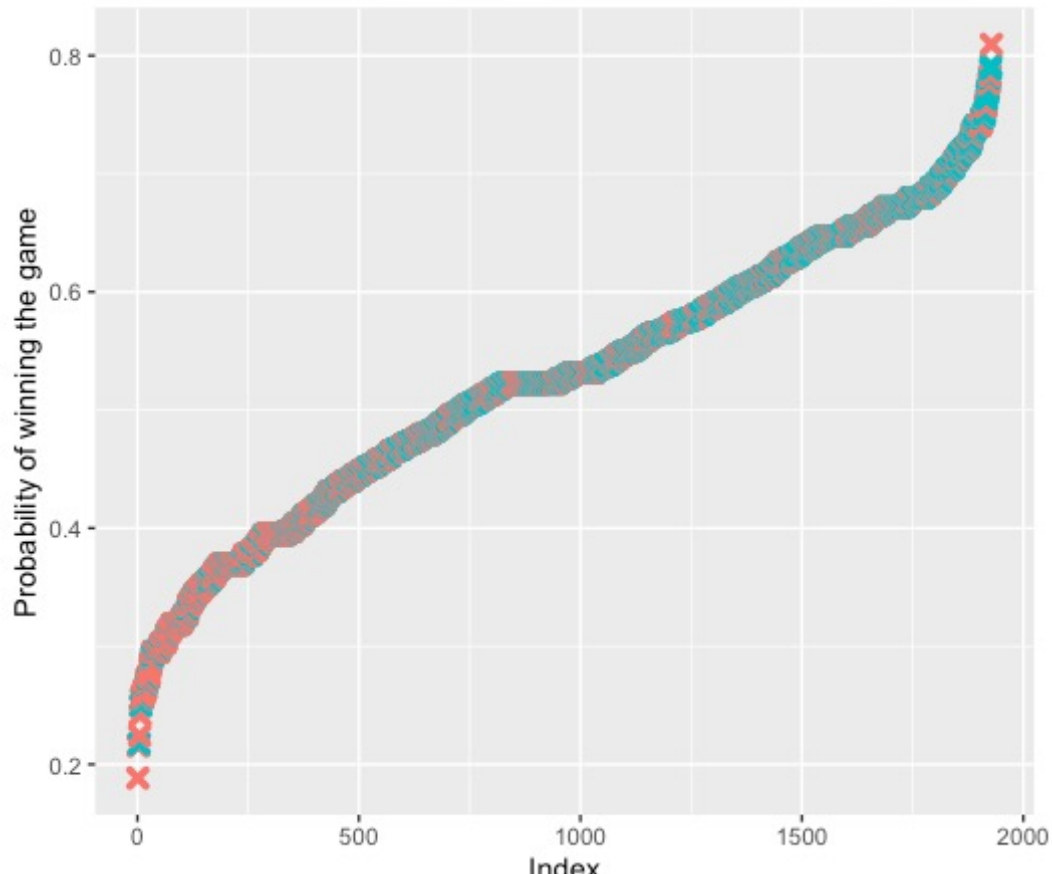
- Features I chose for my predictor were the following:
 1. Win Rate
 2. Win Rate against team B
 3. Relative Team Strength
 4. Relative Run Rate
 5. Relative Wicket Rate

I split my data of 2,042 games into a training and testing set

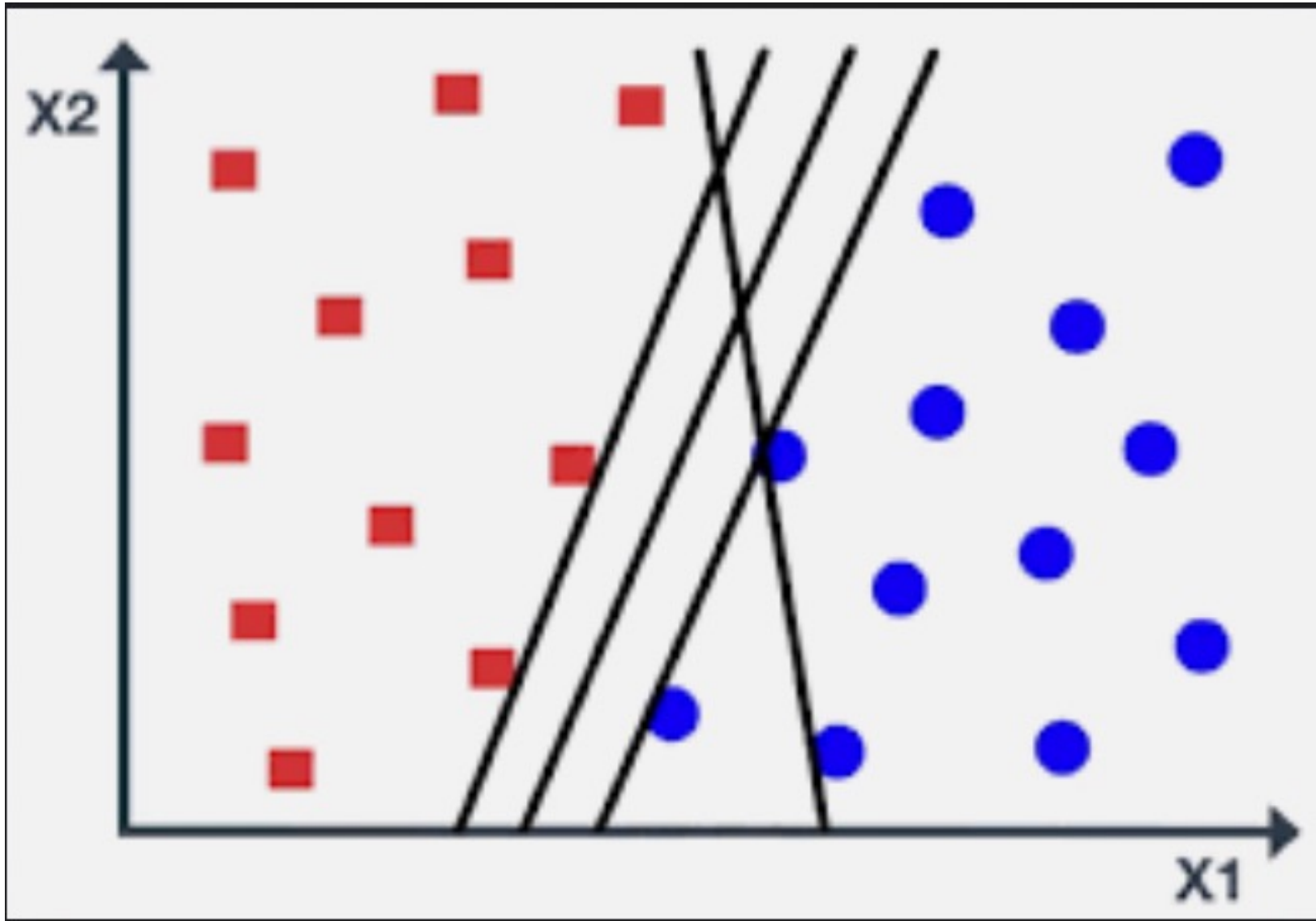
Found testing set accuracy of each model by comparing predicted results to actual results
- SVM model turned out to be the best with an accuracy of approx. 71%

	team_one_win	WinRate	WinRateAgstB	RelativeTeamStrength	RelativeRunRate	RelativeWicke
1	0	0.5000000	0.0	0.5000000	0.000000000	0.
2	0	0.3333333	0.5	0.5923338	0.322033898	2.
3	0	0.5000000	0.5	0.4157351	-0.113983051	-1.
4	1	0.3333333	0.5	0.5000000	0.000000000	0.
5	0	0.3333333	0.5	0.5000000	0.258333333	2.
6	0	0.0000000	0.5	0.5803926	-0.173333333	6.
7	1	0.5000000	0.5	0.5677985	0.067983950	-3.
8	0	0.3333333	0.5	0.5749926	-0.032913165	4.
9	1	1.0000000	0.5	0.5000000	0.087134503	-2.
10	1	0.7000000	0.5	0.5000000	0.000000000	0.
11	1	0.5555556	0.5	0.5000000	0.000000000	0.
12	0	0.0000000	0.5	0.5027684	0.000000000	0.
13	1	0.5000000	0.5	0.5362907	0.000000000	0.
14	0	0.0000000	0.5	0.5000000	0.000000000	0.
15	0	0.5555556	0.5	0.5000000	0.050000000	0.
16	1	0.8461538	0.5	0.5000000	0.444444444	2.

Logistic Regression



Support Vector Machine



Suggester

- Similar process as my predictor
- Found trouble trying to find the accuracy of my suggester
- Decided on a unique approach of using SVM with 4-way classification:
- 1 = Chose to bat first and won the game
- 2 = Chose to bowl first and won the game
- 3 = Chose to bat first and lost the game
- 4 = Chose to bowl first and lost the game

	GroundTendency	TeamTendency	AvsBTendency	AvsBonXTendency	TossGameWon
row names	0.5277778	0.5000000	0.5000000	0.5	2
2	0.5277778	0.0000000	0.5000000	0.5	3
3	0.5277778	0.0000000	0.5000000	0.5	3
4	0.4940239	1.0000000	1.0000000	1.0	4
5	0.4940239	1.0000000	0.5000000	0.5	3
6	0.4940239	0.5000000	0.5000000	0.5	2
7	0.5277778	0.5000000	1.0000000	1.0	4
8	0.5277778	0.0000000	0.5000000	0.5	3
9	0.5277778	0.2500000	1.0000000	1.0	4
10	0.4940239	0.4285714	0.5000000	1.0	1
11	0.4940239	0.2000000	0.5000000	1.0	4
12	0.5632184	0.5000000	0.5000000	0.5	3
13	0.5632184	1.0000000	1.0000000	1.0	1
14	0.4482759	0.5000000	0.5000000	0.5	3
15	0.4940239	0.2000000	0.5000000	0.5	3

Creating a Web App

- Decided to continue using R so that all coding was being done on one platform to avoid any confusion
- Used the R Shiny package
 - many new functions --> was challenging to learn
- Still a work in progress
- Eventually hope to implement this into the CricClubs website



Future Plans

1

Work on a suggester that will suggest optimal team compositions for an upcoming game

2

Be more exact with my predictions by giving a probability of win rather than simply "win" or "loss"

3

Do more research to see the relationship between suggesting and predicting in cricket games

Bibliography

- “Northern California Cricket Association NCCA.” *CricClubs*, <https://www.cricclubs.com/NCCA>.
- “Learn R, Python & Data Science Online.” *DataCamp*, <https://www.datacamp.com/>.
- *YouTube*, YouTube, 23 Aug. 2017, <https://www.youtube.com/watch?v=6tQhoUuQrOw>. Accessed 14 Mar. 2022.
- *Using Machine Learning to Predict the Outcome ...* – *Arxiv.org*. <https://arxiv.org/pdf/1511.05837.pdf>.
- Silver, Nate, Boice, Jay, Paine, Neil. “How Our NFL Predictions Work.” *FiveThirtyEight*, FiveThirtyEight, 5 Sept. 2018, <https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/>.
- Boice, Jay. “HOW OUR MLB Predictions Work.” *FiveThirtyEight*, FiveThirtyEight, 28 Mar. 2018, <https://fivethirtyeight.com/methodology/how-our-mlb-predictions-work/>.
- Coles, ~ Stuart. “Cricviz.” *Smartodds Loves Statistics*, 6 Aug. 2019, <https://smartodds.blog/2019/08/06/cricviz/#:~:text=WinViz%20is%20built%20upon%20CricViz's,the%20probability%20of%20each%20result>.
- Smith, Steve. “Cricket Growing in Popularity in San Diego with Minor League Team.” *KGTV*, KGTV, 9 Sept. 2021, <https://www.10news.com/sports/cricket-growing-in-popularity-in-san-diego-with-minor-league-team>.

Thank You!
Any Questions?

