FUNDAMENTALS OF ANALYTICS AND DISCOVERY INFORMATICS - PROJECT PRESENTATION

Topic : Prediction of the expected sports entertainment level of an international T20 cricket match

- By Vikhyat Dhamija

Introduction

Are you a cricket lover who is a great lover of a blitzkrieg form of a cricket i.e., a Twenty-20 match and have a desire to go to see the match in the stadium, but you are in a dilemma that whether you should spend your money and time going there ?

Are you a company who wants to be associated with highly entertaining T20 matches, in turn increasing their own popularity and revenue ?

*In this project we are trying to build a machine learning model to forecast or predict the **"Expected entertainment level of an International Twenty-20 cricket match".***

Motivation

"The television advertisement rates for the ICC Cricket World Cup 2019 have jumped to their highest, beating what Star India charged for Indian Premier League. The ad rates for India games range from INR 15-16 lakh (22K USD) for 10 seconds, two people in media buying agencies said. And for the much-awaited India-Pakistan match on June 16, a slot costs up to Rs 20 lakh (30K USD) for 10 seconds"[11]

Bloomberg

The Guardian

County	Price
Sussex	£304 (early bird £280)
Somerset	£279
Essex	£277
Northamptonshire	£265
Kent	£250
Warwickshire	£245 (early bird £220)
Middlesex	£245
Hampshire	£240
Yorkshire	£230
Worcestershire	£225

Source: The Guardian[12]

Literature Review

1. **Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning**

Key Points:

- Multivariate regression model to calculate the points of player based on ICC player points based on the factors like number of wickets taken, number of dot balls given, number of fours, number of sixes, number of catches, and number of stumpings
- Based on the individual's player points, the total weight of the team was calculated. Topmost 11 frequently played players of the team were used in the calculation of total teams' weight
- Based on RFE method , the 7 most useful features were detected.
- 10-fold cross validation with various algorithms like Naïve Bayes , Extreme Gradient Boosting , SVM and Logistic Regression, Random Forests and Multilayer Perceptron were used with MLP proved to be more accurate

This study made me understand how to approach the cricket predictions related problems , how the multivariate regression model can be used to calculate the features to be fed to the model, how the Recursive feature Elimination Techniques work to select the important features to be fed to the model.

Literature Review

2. Prediction of the outcome of a Twenty-20 Cricket Match

Key Points:

- For each instance corresponding to each match, the average of various parameters for players were averaged till the Date of the match like their Average Runs, Continuous Average Number of 4s, etc.
- In variation 1, 16 * 22 = 352 features were used per instance (11 per team and as two team played a match hence 22) i.e., the match in order to predict the outcome of the match.
- In variation 2, each player's features were aggregated to two values i.e., the Batting Aggregate and Bowling Aggregate where Time Scaling was used to scale the performances of the players
- In Variation 3, each player's feature i.e., the batting aggregate and bowling aggregate were converted into one value per player.

This study made me understand various approached to feature extraction and transformation for building the good Machine learning models.

Literature Review

3. Sport analytics for cricket game results using machine learning: An experimental study

Key Points

- Focused on different strategy to come out with the best model for predicting Cricket Game Results
- Two features set were used to predict the outcome of the match where first feature set was based on the features related to Home Team and another feature set was based on the Toss decision.
- Influential features had been identified using Correlation based feature selection, Information Gain etc. techniques.
- Evaluation of the models was performed using the 10-fold cross validation based on average predictive accuracy, recall and precision

This study made me understand that how we can apply various techniques for feature selection, how we can use different feature sets that are based on completely different approaches to prediction

Approach

Definition of a Concept

- Entertainment level will be generated by the mathematical formula:
  = (Total runs scored in 4’s and 6’s in both innings / Total runs scored in both innings) * 100

```
def label_function(x):
    if x > 0.60:
        return "H"
    elif x < 0.55:
        return "L"
    else:
        return "A"

merged_df["Elevel"] = merged_df["Elevelval"].apply(label_function)
```

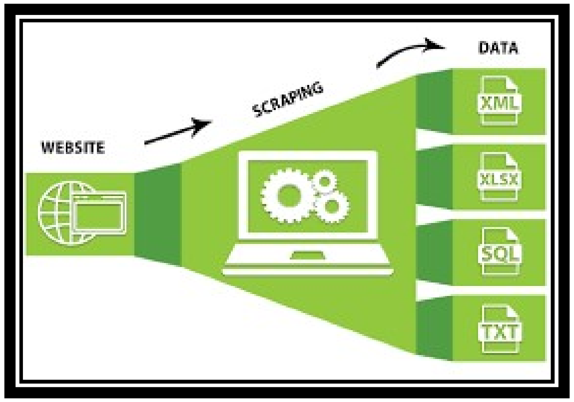
Approach

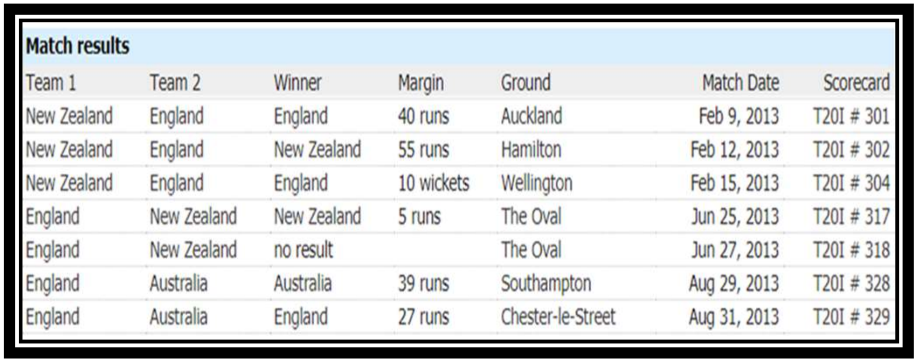
Attribute Set

S.No.	Attributes	Description
1	Team1	*Team 1* are the countries who plays the international T20 cricket. Here the current Top 10 Teams in ICC Twenty 20 Cricket ratings are considered, whose matches arise profound interest and viewership. These are : England, India, Australia, Pakistan, New Zealand, South Africa, Bangladesh, Afghanistan, West Indies, Sri Lanka
2	Team2	*Team 2* are the countries who plays the international T20 cricket. Here the current Top 10 Teams in ICC Twenty 20 Cricket ratings are considered, whose matches arise profound interest and viewership. These are : England, India, Australia, Pakistan, New Zealand, South Africa, Bangladesh, Afghanistan, West Indies, Sri Lanka
3	MatchTime	This will have three values : 1. day 2. Night 3. daynight
4	Ha1	This attribute will have value 1/0 depending on the fact that Team 1 has the home advantage or not.
5	Ha2	This attribute will have value 1/0 depending on the fact that Team 2 has the home advantage or not.
6	T1w	This attribute will have value 1/0 depending on the fact that Team 1 has won the toss or not.
7	T2w	This attribute will have value 1/0 depending on the fact that Team 2 has won the toss or not.
8	T1b	This attribute will have value 1/0 depending on the fact that Team 1 bat first or not.
9	Groundcapacity	This will have the values of the various cricket grounds seating capacities as it directly relates to the viewership of the match which may be a variable on which the performance of the players can be dependent as more viewers means more noise which may positively affect the results increasing the motivation of the players or they can increase the noise so as to create disturbance. So, it also forms an important feature to decide the performances of teams in the match.
10	AvgSR	This is the average batting strike rate of the top 5-6 players in the team line up. As the top 5-6 players in a team line up are the top batsmen of the team.
11	CRating	This will have the sum of the numbers assigned based on the ranks(ratings) of the Team1 and Team2 in the year when match was played as higher the rank of the two teams more entertaining the match will be.
12	Elevel	This is the class label for the Entertainment Level that has three values – High(H) , Average(A) and Low(L)

Data Collection

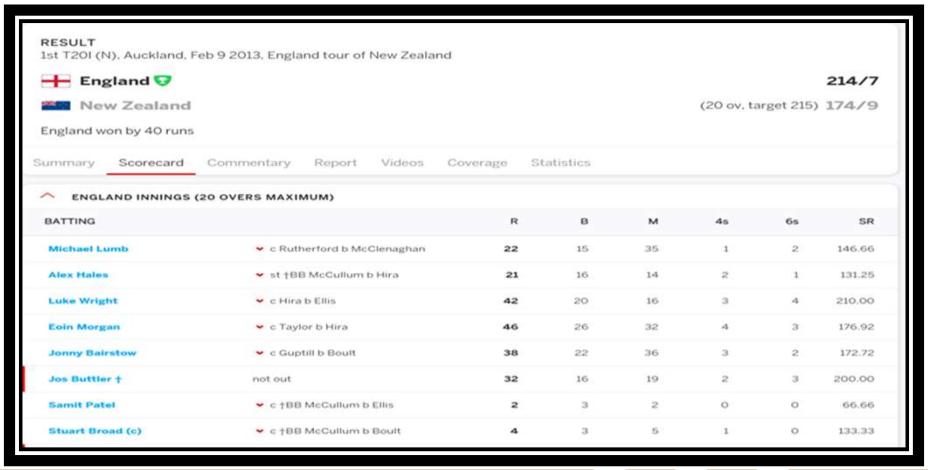
Web Scraping Script was built using BeautifulSoup Libraries and used to crawl ESPN cricinfo and Wikipedia





Match results

Team 1	Team 2	Winner	Margin	Ground	Match Date	Scorecard
New Zealand	England	England	40 runs	Auckland	Feb 9, 2013	T20I # 301
New Zealand	England	New Zealand	55 runs	Hamilton	Feb 12, 2013	T20I # 302
New Zealand	England	England	10 wickets	Wellington	Feb 15, 2013	T20I # 304
England	New Zealand	New Zealand	5 runs	The Oval	Jun 25, 2013	T20I # 317
England	New Zealand	no result		The Oval	Jun 27, 2013	T20I # 318
England	Australia	Australia	39 runs	Southampton	Aug 29, 2013	T20I # 328
England	Australia	England	27 runs	Chester-le-Street	Aug 31, 2013	T20I # 329



RESULT
1st T20I (N), Auckland, Feb 9 2013, England tour of New Zealand

England 214/7
New Zealand (20 ov, target 215) 174/9

England won by 40 runs

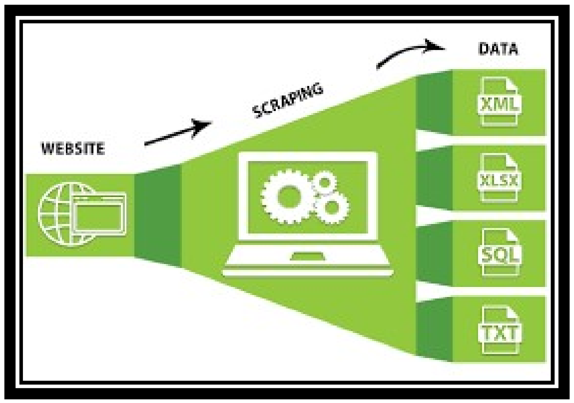
Summary Scorecard Commentary Report Videos Coverage Statistics

ENGLAND INNINGS (20 OVERS MAXIMUM)

BATTING		R	B	M	4s	6s	SR
Michael Lumb	c Rutherford b McClenaghan	22	15	35	1	2	146.66
Alex Hales	st †BB McCullum b Hira	21	16	14	2	1	131.25
Luke Wright	c Hira b Ellis	42	20	16	3	4	210.00
Eoin Morgan	c Taylor b Hira	46	26	32	4	3	176.92
Jonny Bairstow	c Guptill b Boult	38	22	36	3	2	172.72
Jos Buttler †	not out	32	16	19	2	3	200.00
Samit Patel	c †BB McCullum b Ellis	2	3	2	0	0	66.66
Stuart Broad (c)	c †BB McCullum b Boult	4	3	5	1	0	133.33

Data Collection

Web Scraping Script was built using BeautifulSoup Libraries and used to crawl ESPN cricinfo and Wikipedia



ICC Test Championship 28 April 2012[9]

Rank	Team	Matches	Points	Rating
1	England	44	5124	116
2	South Africa	32	3709	116
3	Australia	46	5153	112
4	India	46	5103	111
5	Pakistan	35	3781	108
6	Sri Lanka	38	3780	99
7	West Indies	34	2898	85
8	New Zealand	28	2366	85
9	Bangladesh	18	135	8

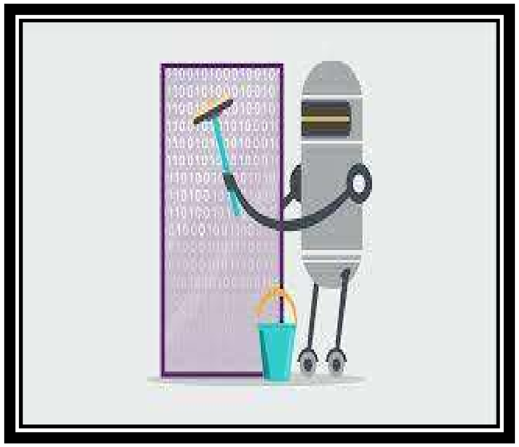
ICC ODI Championship 16 April 2012[9]

Rank	Team	Matches	Points	Rating
1	Australia	49	6030	123
2	South Africa	30	3549	118
3	India	55	6409	117
4	England	39	4333	111
5	Sri Lanka	52	5745	110
6	Pakistan	45	4710	105
7	New Zealand	31	2667	86
8	West Indies	32	2753	86
9	Bangladesh	36	2408	67
10	Zimbabwe	33	1511	46
11	Ireland	14	504	36
12	Netherlands	9	137	15
13	Kenya	9	74	8

ICC T20I Championship 31 March 2012[9]

Rank	Team	Matches	Points	Rating
1	England	14	1811	129
2	South Africa	12	1468	122
3	Sri Lanka	9	1056	117
4	New Zealand	14	1596	114
5	Pakistan	17	1817	107
6	Australia	15	1603	107
7	India	9	930	103
8	Ireland	10	946	95
9	West Indies	10	933	93
10	Afghanistan	6	500	83
11	Netherlands	5	321	64
12	Zimbabwe	9	463	51
13	Scotland	5	200	40
14	Canada	5	79	16
15	Kenya	6	75	13

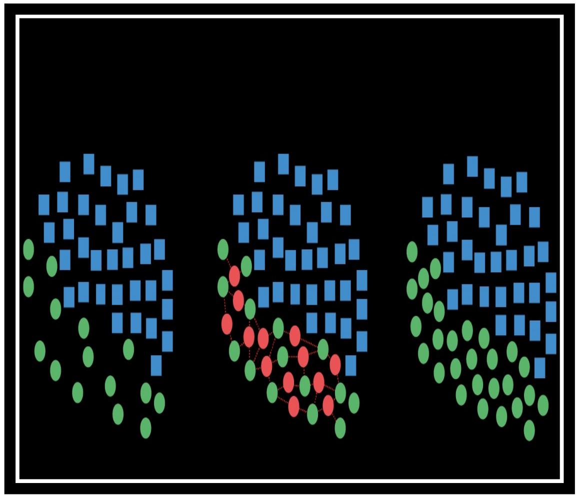
Data Cleaning & Transformation



Following steps were taken data :

- Merging of the Data Frames
- Duplicates were removed
- Missing Values were replaced by the mean values
- One hot encoding was performed on the categorical variables like the Team1 and Team2
- SMOTE technique was used to create a more balanced dataset.
- Correlation was checked between the various features for feature selection.

Synthetic Minority Oversampling Technique



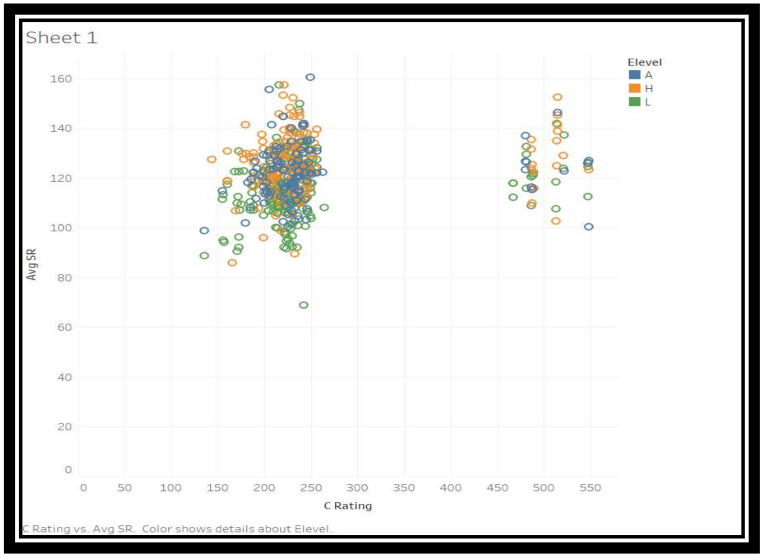
- This technique try to balance the minority classes.
- For one minority instance its K nearest neighbors in the feature space are selected where one neighbor is chosen at random.
- Synthetic instance is created by the convex combination of the chosen minority class instance and the selected neighbor.
- This technique leads to a balanced dataset as well as the increased number of instances.

Dataset

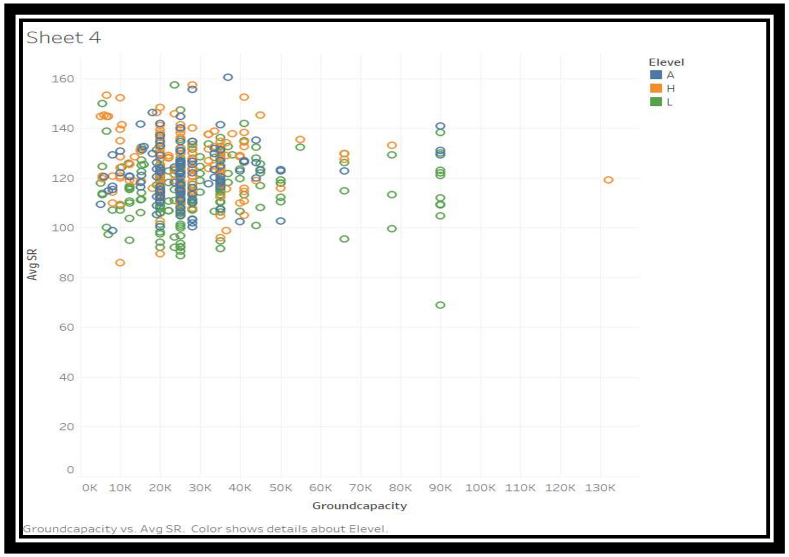
	Team1	Team2	MatchTime	Ha1	Ha2	T1w	T2w	T1b	Groundcapacity	AvgSR	CRating	Elevel
0	West Indies	England	daynight	1	0	0	1	1	12400	125.31	487.00	H
1	West Indies	England	daynight	1	0	1	0	0	8000	115.31	487.00	A
2	West Indies	England	daynight	1	0	1	0	1	8000	109.92	487.00	H
3	England	Pakistan	day	1	0	0	1	0	5500	124.65	547.00	L
4	New Zealand	England	day	1	0	0	1	1	18000	146.25	515.00	A
...	...	...	...	...	...	...	...	...	...	...	...	...
869	India	Pakistan	daynight	1	0	0	1	1	132000	119.18	210.00	H
873	Sri Lanka	Pakistan	night	1	0	1	0	1	35000	94.63	224.00	L
874	Sri Lanka	Pakistan	night	1	0	0	1	0	35000	91.52	224.00	L
879	Bangladesh	Pakistan	night	0	0	1	0	1	35000	122.34	235.00	H
883	Sri Lanka	Pakistan	night	0	0	1	0	1	35000	114.40	224.00	L

486 rows × 12 columns

Exploratory Analysis

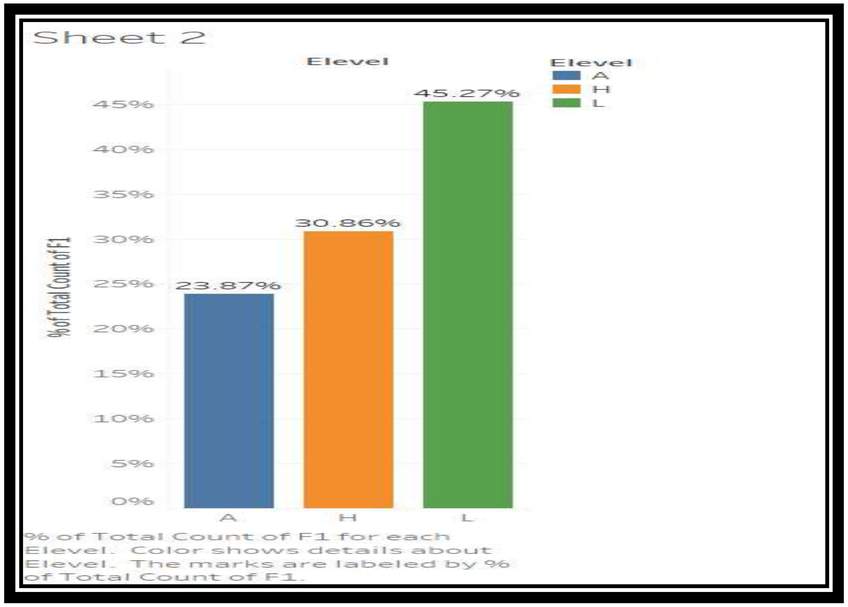


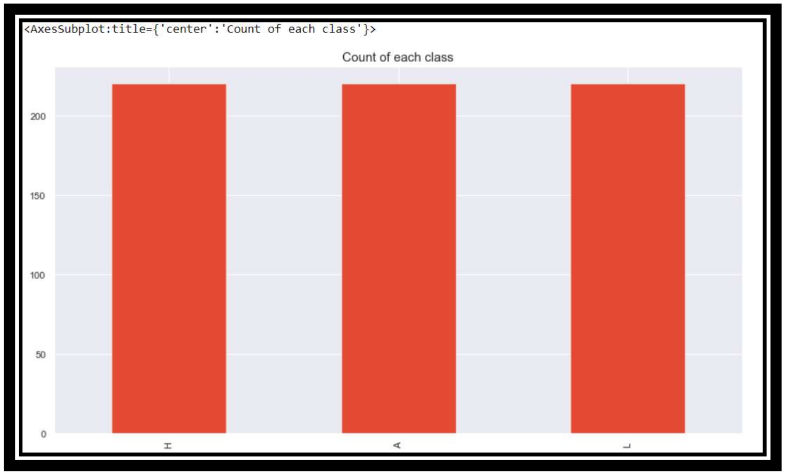
Combined Ratings Vs Average Batting Strike Rate



Ground Capacity Vs Average Batting Strike Rate

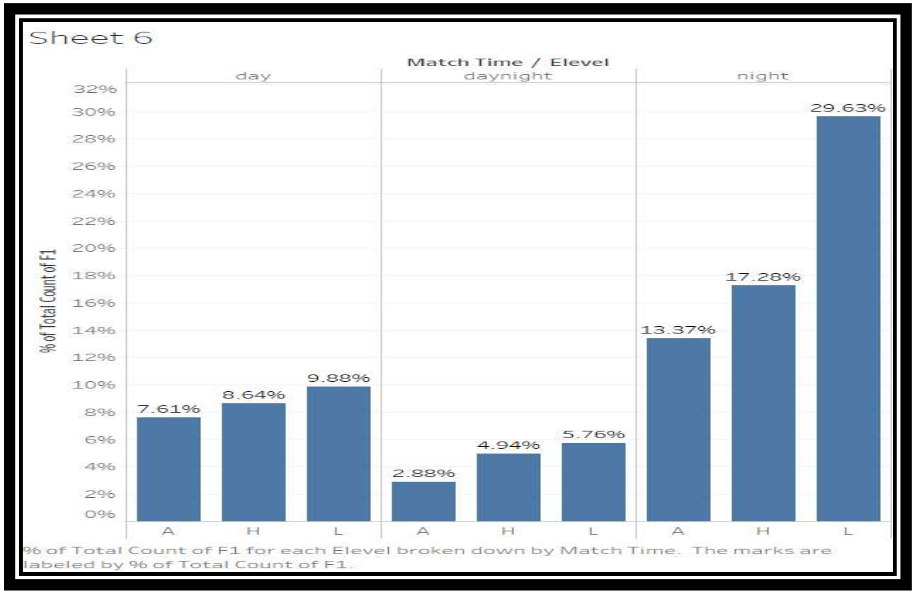
Exploratory Analysis





Balanced Dataset after SMOTE

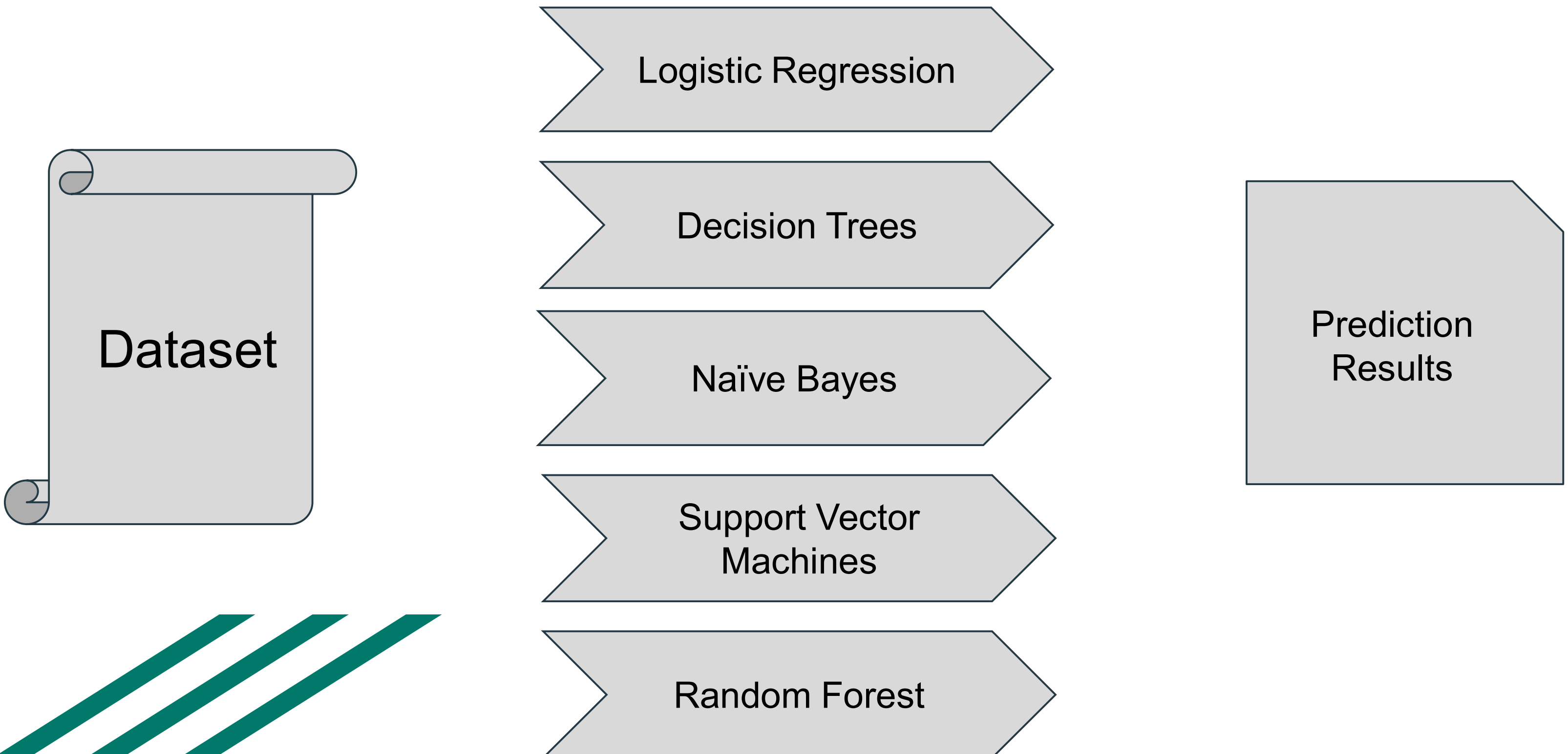
Exploratory Analysis



		Elevel		
Ha1	Team1	A	H	L
0	Afghanistan	3	3	3
	Australia	4	10	18
	Bangladesh	7	6	8
	England	5	9	14
	India	7	4	12
	New Zealand	3	2	17
	Pakistan	5	1	16
	South Africa	1	1	4
	Sri Lanka	3	3	7
	West Indies	2	1	1
1	Afghanistan			3
	Australia	16	8	23
	Bangladesh	2	7	7
	England	7	11	16
	India	9	14	17
	New Zealand	12	28	10
	Pakistan		4	4
	South Africa	16	21	11
	Sri Lanka	4	7	19
	West Indies	10	10	10

		Elevel		
Ha2	Team2	A	H	L
0	Afghanistan	2	1	
	Australia	9	15	10
	Bangladesh	7	6	10
	England	14	22	9
	India	13	17	20
	New Zealand	8	9	20
	Pakistan	22	17	49
	South Africa	7	10	28
	Sri Lanka	8	15	35
	West Indies	14	27	21
1	Bangladesh			1
	New Zealand		1	1
	Pakistan	4	1	3
	South Africa	3		2
	Sri Lanka	1	2	8
	West Indies	4	7	3

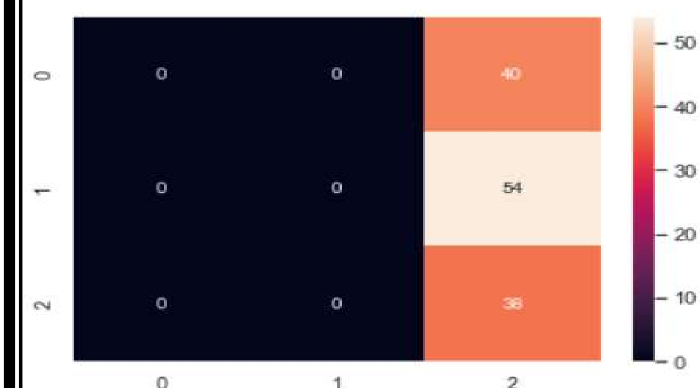
ML Algorithms



Evaluation Results

ZeroR Base Line Model

```
-------------------ZeroR on SMOTE Balanced Dataset----------
[[ 0  0 40]
 [ 0  0 54]
 [ 0  0 38]]
              precision    recall  f1-score   support

           A       0.00      0.00      0.00        40
           H       0.00      0.00      0.00        54
           L       0.29      1.00      0.45        38

    accuracy                           0.29       132
   macro avg       0.10      0.33      0.15       132
weighted avg       0.08      0.29      0.13       132

<AxesSubplot:>
```

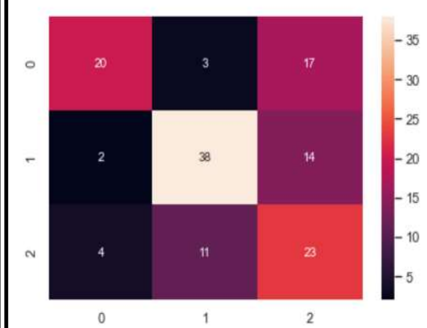


Evaluation Results

Random Forest

```
For the Random Forest based Ensemble classifier :
The average of the Macro averaged F1 score using the cross validation is: 0.6266006042556326
The average of the Macro averaged Precision score using the cross validation is: 0.6478044333386317
The average of the Macro averaged Recall score using the cross validation is: 0.6274001536098311
The average of the Accuracy using the cross validation is: 0.6367941129419293
```

```
------------------Random Forest Based Ensemble Algorithm on SMOTE Balanced Dataset--
Test Set Accuracy : 61.36363636363637 %


[[20  3 17]
 [ 2 38 14]
 [ 4 11 23]]
              precision    recall  f1-score   support

           A       0.77      0.50      0.61        40
           H       0.73      0.70      0.72        54
           L       0.43      0.61      0.50        38

    accuracy                           0.61       132
   macro avg       0.64      0.60      0.61       132
weighted avg       0.65      0.61      0.62       132

<AxesSubplot:>
```



Evaluation Results

Support Vector Machines

For the SVM based classifier :
The average of the Macro averaged F1 score using the cross validation is: 0.5836740468011709
The average of the Macro averaged Precision score using the cross validation is: 0.5897392862236653
The average of the Macro averaged Recall score using the cross validation is: 0.5900537634408601
The average of the Accuracy using the cross validation is: 0.5898736202207647

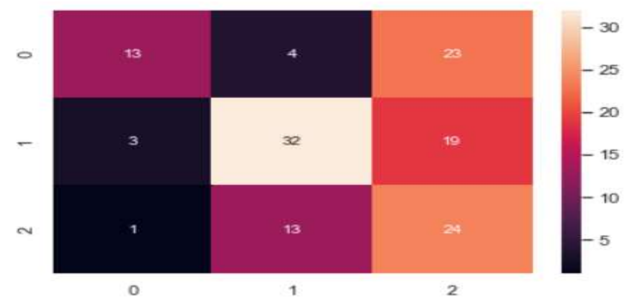
```
------------------SVM on SMOTE Balanced Dataset----------------------------
Test Set Accuracy : 62.121212121212125 %


[[23  2 15]
 [ 4 35 15]
 [ 3 11 24]]
              precision    recall  f1-score   support

           A       0.77      0.57      0.66        40
           H       0.73      0.65      0.69        54
           L       0.44      0.63      0.52        38

    accuracy                           0.62       132
   macro avg       0.65      0.62      0.62       132
weighted avg       0.66      0.62      0.63       132

<AxesSubplot:>
```

Evaluation Results

Decision Trees

```
For the Decision Tree classifier :
The average of the Macro averaged F1 score using the cross validation is: 0.514499132841474
The average of the Macro averaged Precision score using the cross validation is: 0.563819424269317
The average of the Macro averaged Recall score using the cross validation is: 0.527553763408602
The average of the Accuracy using the cross validation is: 0.5291313389857623
```

```
------------------Decision Tree-----------------------------
[[13  4 23]
 [ 3 32 19]
 [ 1 13 24]]
              precision    recall  f1-score   support

           A       0.76      0.33      0.46        40
           H       0.65      0.59      0.62        54
           L       0.36      0.63      0.46        38

    accuracy                           0.52       132
   macro avg       0.59      0.52      0.51       132
weighted avg       0.60      0.52      0.53       132

<AxesSubplot:>
```



Evaluation Results

Logistic Regression

```
For the Logistic Rgeression based classifier :
The average of the Macro averaged F1 score using the cross validation is: 0.47100740691750004
The average of the Macro averaged Precision score using the cross validation is: 0.48078801883530253
The average of the Macro averaged Recall score using the cross validation is: 0.4780145929339477
The average of the Accuracy using the cross validation is: 0.47774756039033756
```

```
------------------Logistic Regression on SMOTE Balanced Dataset--------------
Test Set Accuracy : 52.27272727272727 %


[[21 15  6]
 [ 8 17  9]
 [17  8 31]]
              precision    recall  f1-score   support

           A       0.46      0.50      0.48        42
           H       0.42      0.50      0.46        34
           L       0.67      0.55      0.61        56

    accuracy                           0.52       132
   macro avg       0.52      0.52      0.51       132
weighted avg       0.54      0.52      0.53       132

<AxesSubplot:>
```

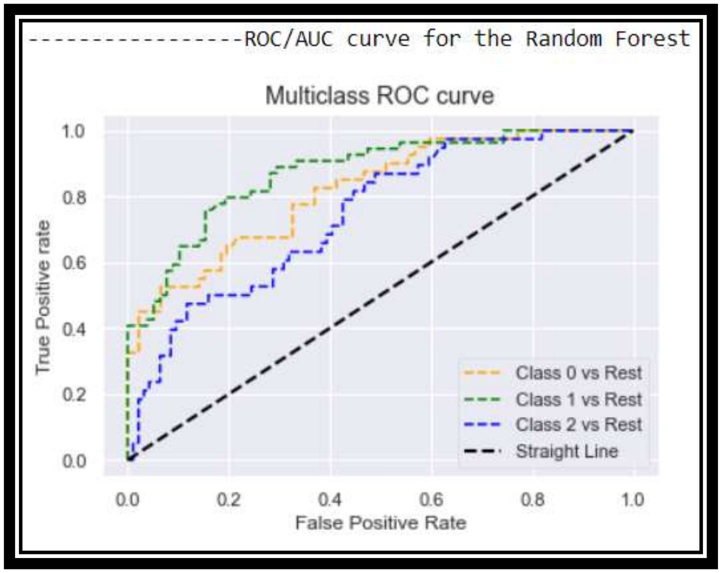
Evaluation Results

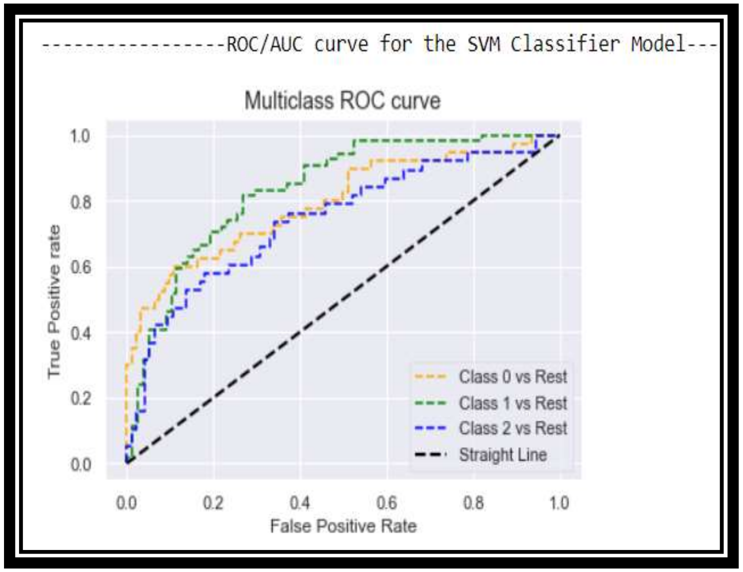
Naïve Bayes

```
For the Naive Bayes classifier :
The average of the Macro averaged F1 score using the cross validation is: 0.4005765731906107
The average of the Macro averaged Precision score using the cross validation is: 0.4033713216846847
The average of the Macro averaged Recall score using the cross validation is: 0.4096582181259601
The average of the Accuracy using the cross validation is: 0.40947048472244446
```

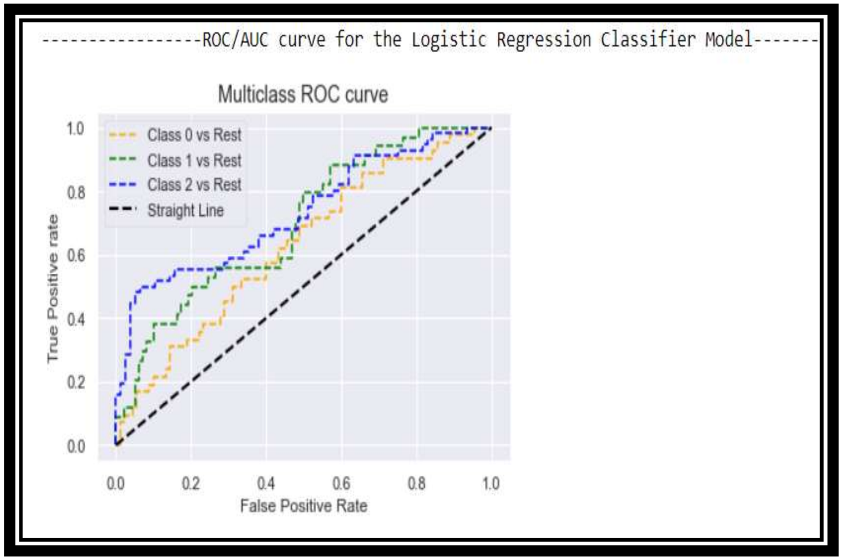
```
------------------Naive Bayes on SMOTE Balanced Dataset----------------------------
0.5
[[17 13 10]
 [14 29 11]
 [ 5 13 20]]
              precision    recall  f1-score   support

           A       0.47      0.42      0.45        40
           H       0.53      0.54      0.53        54
           L       0.49      0.53      0.51        38

    accuracy                           0.50       132
   macro avg       0.50      0.50      0.50       132
weighted avg       0.50      0.50      0.50       132

<AxesSubplot:>
```

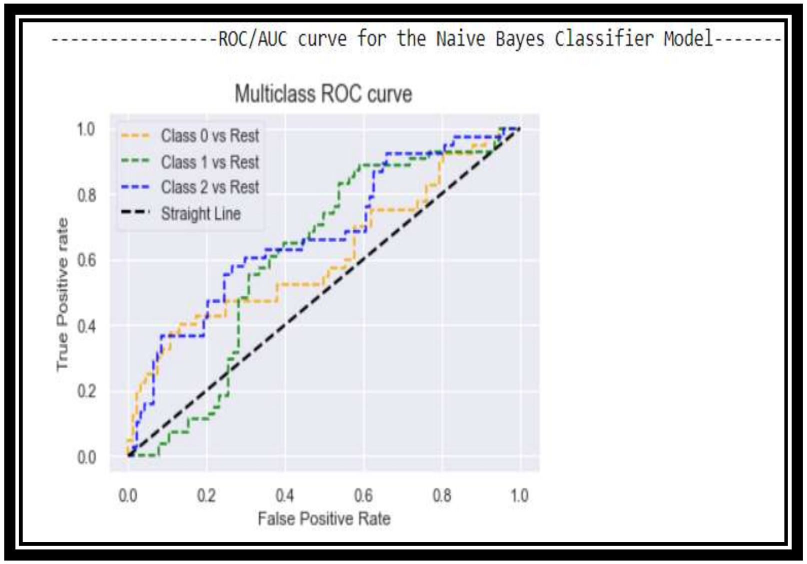
ROC-AUC Curves





ROC-AUC Curves





Discussion

- Apart from SVM and Random Forest, which are showing the accuracy, F1 Score, and other evaluation metrics above 60 %, others' metrics are hovering in the range of 50 to 60 % while Naïve Bayes performed badly.

- Reasons that SVM performed better can be the fact that :

✓ In SVM the best marginal distance is calculated between the line and the support vectors which highly reduce the errors.
✓ SVM use the geometrical properties or the spatial representation of data as the criteria for classification

- Reasons that Random Forest performed better than others especially Decision Trees :

✓ Decision tree algorithm at each node we make the best choice of splitting towards the purity of class but the problem with this approach is that we may not achieve the global optimum.
✓ Random Forest Classifier creates the various decision trees and choose the results through majority voting.

Discussion

- Naïve Bayes performance has been quite abysmal. The probable reason may be that the Naïve Bayes consider the features as independent . Naïve Bayes is also a probabilistic method and the SVM use the spatial representation.
- Higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes which shows that our SVM Model and Ensemble based Random Classifiers are performing much better than the other models developed through other algorithms, but the Naïve Bayes showed the poor performance.
- As the major target for our machine learning problem are the advertisers and agencies and people investing in the matches who want to know that how the match will be. So, precision of the Class "H" is very important so that they can put their best bet on the match and can retrieve the better return. And that seems to better be provided by the Random Forest Classifier based on the current Test dataset.
- Machine learning or Data Science problem all depends on the data as a food for producing better results, it seems that in order to bring better results more features and instances are required and hence our current dataset needs to be augmented.

Conclusion

- Evaluation metrics for the model generated through various algorithms used are higher than the baseline Zero R.
- Apart from SVM and Random Forest, which are showing the accuracy, F1 Score, and other evaluation metrics above 60 %, others' metrics are hovering in the range of 50 to 60 % while Naïve Bayes performed badly.
- The probable reasons for ensemble method like Random Forest performing well are :

➢ It generates various decision trees and then take majority voting, so the decision made tends to be more accurate.

➢ Random Forests produce unpruned and diverse trees which increase the resolution of the feature space.

➢ Random Forests get off the overfitting problem based on its randomness and voting mechanisms.

- SVM also has performed better than the others and the reason that seems to be causing this is the use of the geometrical representation by SVM to slice the data points into various classes

Future Work

1. Increase the feature space.

Currently relying on the features that directly affects the batting like the ratings, average batting strike rate , ground size and so on.

Capabilities of the bowlers also affect the performance of batsmen so the bowlers' data will be further scraped and that also will be used in the future enhancements for the project.

2. More algorithms will be used.

Other algorithms like gradient boosted decision trees, Multilayer Perceptron etc can also be used on this problem apart from the ones currently being used.

3. Scaling methodology

As learnt in the second reviewed paper, the parameters were time scaled to bring them on par with respect to time . Similarly, it is felt that the features like strike rates are increasing as the more player are tending to be more inclined to faster batting, so the older average strike rates of the team are not at par with the current average strike rates, hence they need to be scaled accordingly.

Time Log

Date	Time	Duration	Work
04/10/2021	18:00 – 00:00 EST	5	Studied the methodology and Libraries like Beautiful Soup Libraries of Python to be used for Web Scraping.
04/12/2021	11:00 – 16:00 EST	5	Worked on building script for scraping through the ESPN Cricinfo Website for getting the relevant attributes and storing them in the CSV file.
04/13/2021	08:00 – 13:30 EST	5.5	Worked on building script for scraping through the ESPN Cricinfo Website for getting the relevant attributes and storing them in the CSV file.
04/14/2021	09:00 – 14:00 EST	5	Worked on building script for scraping through the ESPN Cricinfo Website for getting the relevant attributes and storing them in the CSV file.
04/15/2021	19:30 – 00:00 EST	4.5	Worked on building script for scarping the Wikipedia Website for getting the relevant attributes and storing them in the CSV file.
04/16/2021	20:00 – 01:00 EST	5	Studied the methodology to perform various Data Cleaning tasks like Duplicate Removals, Filling the Null Values, Performing the joins using the Merge operations and various other necessary Data cleaning and Feature Engineering operations to be performed using the Python Pandas.

Time Log

04/17/2021	19:30 – 23:30 EST	4	Performed the Data Cleaning Tasks on the Dataset so as to come out with the combined and cleaned Data set to be used. Note that the calculations were performed to generate the Class Labels
04/18/2021	10:00 – 14:30 EST	4.5	Applied various Machine Learning Algorithms on the Dataset to check their performances
04/19/2021	09:00-13:00 EST 20:30- 23:30 EST	7	Based on the suggestions from Dr. Bill, studied about the SMOTE and applied SMOTE to generate more Balanced Dataset and increase the number of instances. Applied various Machine Learning Algorithms again on the Dataset to check their performances. Tried to drop some features and performed some hit and trial to check the performances.
04/20/2021	21:30 – 23:00 EST	2.5	Build the function for ZeroR algorithm and then check the performance of ZeroR model for our dataset in order to check that whether we are gaining from the machine Learning Algorithms.

Time Log

04/20/2021	08:00 – 12:00 EST 20:00 – 22:00 EST	6	Research ways to enhance the metrics and understood that one of the best ways is to introduce the best attributes or independent variables that can lead to improved prediction. Put the thought process to have such feature that can be extracted also through Web Scraping and came out with the Average Batting Strike rates of the Teams Playing. Modified the python script in order to extract the above-mentioned field from the website.
04/21/2021	11:00 – 14:00 EST	3	Reworked the Data Cleaning, merging tasks in order to produce the final Dataset again.
04/22/2021	10:00 – 14:00 EST	4	Based on the suggestion from the professor and for checking the relevancy of the attributes and to gain further insights about our data points, the exploratory analysis was performed using Data Visualization using Tableau. Feature correlation check was also performed through Pandas.
04/23/2021	21:00 – 00:00 EST	3	Applied various Machine Learning Algorithms again on the Dataset to check their performances. Here I got my performance improved for various models like SVM and Random Forest based classifier with parameters reaching 60-65 %.
04/23/2021	09:00 – 12:00 EST 13:00 – 17:00 EST	7	Studied the methodology to generate the ROC-AUC curves using Python Pandas. Created ROC-AUC curves for various Machine Learning Algorithms for comparing various Models.

Time Log

04/24/2021	12:00 – 16:00 EST 21:30 - 23:30 EST	6	Report Writing
04/25/2021	12:00 – 16:00 EST 21:30 - 23:30 EST	6	Report Writing
04/27/2021	19:00 – 00:00 EST	5	Presentation Preparation
05/02/2021	20:00 - 21:00 EST	1	Report Writing
05/03/2021	22:00 - 23:00 EST	1	Report Writing
05/04/2021	16:00 - 17:30 EST	1.5	Report Writing

References

1. Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning https://arxiv.org/pdf/1809.09813.pdf -
2. Prediction of the outcome of a Twenty-20 Cricket Match http://pages.cs.wisc.edu/~shruthir/Documents/MachineLearning_Final_Report.pdf -
3. Sport analytics for cricket game results using machine learning: An experimental study https://www.emerald.com/insight/content/doi/10.1016/j.aci.2019.11.006/full/pdf -
4. Quantitative Analysis of Forthcoming ICC Men's T20 World Cup 2020 Winner Prediction using Machine Learning. https://www.ijcaonline.org/archives/volume176/number32/singh-2020-ijca-920388.pdf -
5. Dynamic Winner Prediction in Twenty20 Cricket: Based on Relative Team Strengths https://dtai.cs.kuleuven.be/events/MLSA17/papers/MLSA17_paper_5.pdf -
6. Cricket Match Outcome Prediction Using Machine Learning http://www.ijasret.com/VolumeArticles/FullTextPDF/480_9.CRICKET_MATCH_OUTCOME_PREDICTION_USING.pdf -
7. Performance measure on multiclass classification https://www.youtube.com/watch?v=HBi-P5j0Kec -
8. Analytics Vidhya Website for various information https://www.analyticsvidhya.com/
9. Machine Learning Mastery website for various information https://machinelearningmastery.com/
10. Towards Data Science Website https://towardsdatascience.com/
11. https://www.bloombergquint.com/business/why-advertising-rates-have-hit-a-record-this-cricket-world-cup
12. https://www.theguardian.com/sport/2020/mar/11/cricket-ticket-prices-fans-good-value-money
13. https://adscholars.com/blog/advertisement-spend-ipl-2020/

Thank You