# CRICKET TEAM PREDICTION USING MACHINE LEARNING

**MICRO PROJECT REPORT**

**Submitted by**

## MAGULURI MANIKANTA-99210041073

**in partial fulfilment for the award of the degree of**

## BACHELOR OF TECHNOLOGY

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

****

**SCHOOL OF COMPUTING**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION KRISHNANKOIL 626 126**

APRIL 2025

**DECLARATION**

We affirm that the micro project work titled **“CRICKET TEAM PREDICTION USING MACHINE LEARNING ”** being submitted in partial fulfilment for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is the original work carried out by us. It has not formed part of any other project work submitted for the award of any degree or diploma, either in this or any other University.

**Student Name**

99210041073

MAGULURI MANIKANTA

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date:17-03-2025

**Signature of the Mentor**

Mrs.S.RESHNI

Assistant Professor Department of Computer Science and Engineering



**BONAFIDE CERTIFICATE**

Certified that this project report **“CRICKET TEAM PREDITION USING MACHINE LEARNING”** is the Bonafide work of “**MAGULURI MANIKANTA(99210041073)”** who carried out the Micro project work under my supervision.

**Mrs.S.RESHNI Dr. N. Suresh Kumar**

**SUPERVISOR HEAD OF THE DEPARTMENT**

**Assistant Professor Professor and Head**

Computer Science and Engineering Computer Science and Engineering

Kalasalingam Academy of Research Kalasalingam Academy of Research and Education and Education

Krishnankoil 626126 Krishnankoil 626126

Virudhunagar District. Virudhunagar District.

Submitted for the Micro Project Viva-voice examination held on

**Internal Examiner External Examiner**

**ACKNOWLEDGEMENT**

First and foremost, we thank the ‘Supreme Power’ for the immense grace showered on us which enabled us to do this project. We take this opportunity to express sincere thanks to the late,

**“Kalvivallal” Thiru T. KALASALINGAM, Chairman, Kalasalingam Group of Institutions, “Illayavallal” Dr. K. SRIDHARAN, Ph.D., Chancellor, Dr. S. SHASI ANAND, Ph.D., Vice President,** who is the guiding light for all the activities in our university.

We thank our Vice Chancellor **Dr. S. NARAYANAN, Ph.D.,** for guiding every one of us and infusing us with the strength and enthusiasm to work successfully.

We wish to express our sincere thanks to our respected Head of the Department **Dr. N. SURESH KUMAR,** whose moral support encouraged us to process through our project work successfully.

We offer our sincerest gratitude to our Project Supervisor, **Mrs.S.RESHNI,** for his patience, motivation, enthusiasm, and immense knowledge.

We are extremely grateful to our Micro Project Coordinator **Dr. P. Anitha**, Faculty In Charges **Dr. M. Rajasekaran, Mrs. B. Lavanya, Ms. P. J. Kiruthiga** for their constant encouragement in the completion of the Project.

Finally, we thank all, our Parents, Faculty, Non-Teaching Faculty, and our friends for their moral support.



**SCHOOL OF COMPUTING COMPUTER SCIENCE AND ENGINEERING MICRO PROJECT SUMMARY**

|  |  |  |
| --- | --- | --- |
| Micro Project Title | **CRICKET TEAM PREDICTION USING MACHINE LEARNING** | |
| Micro Project Team Members (Name with  Register No) | MAGULURI MANIKANTA-99210041073 | |
| Guide Name/Designation | Mrs.S.RESHNI  Assistant Professor, Department of Computer Science and Engineering | |
| Program Concentration Area | MACHINE LEARNING | |
| Technical Requirements | Multi Linear Regression, MinMax Scaler | |
| Engineering standards and realistic constraints in these areas | | |
| **Area** | **Codes & Standards / Realistic Constraints** | **Tick**  ✓ |
| Economic | This would involve forecasting how weather will affect crop  productivity and providing advise on how to minimize potential losses from unsafe scenarios.. |  |
| Environmental | The method supports sustainable agriculture by taking consideration of soil conditions, irrigation efficacy, and crop  selection to minimize environmental impact. |  |
| Social | Provides significant information for farmers and policymakers, resulting in improved food security, efficient land use, and  community wellbeing. |  |
| Ethical | Makes sure equal and neutral access to agricultural data, reducing misunderstanding while promoting equal chances for all producers. |  |
| Health and Safety | Allows farmers calculate the risks of soil deterioration, pest  infestations, and crop illnesses, decreasing potential losses and enhancing food safety. |  |
| Manufacturability | Software-based, scalable, and easy to integrate. |  |
| Sustainability | Supports eco-friendly farming and reduces waste. |  |

**ABSTRACT**

Player selection is a crucial part of every sport, and cricket is no exception. A variety of factors influence the players' performance. The team management and captain select eleven players from the full roster for each match. After a study of numerous factors and player results, the best eleven players are chosen. Each batter contributes by scoring runs, while each bowler contributes by taking wickets and allowing the least amount of runs possible.

This research aims to predict team success based on historical player performance. Individual player performances, as well as their contribution to the team, such as best batting performance among available batsmen, best bowling performance among available bowlers, and best allrounder performance, will be highly useful in determining the eleven players. To create the problem's prediction models, we employed the like Random Forest classifier, Naive Bayes, KNN, Logistic regression, Support Vector Machine (SVM) AND Decision tree classifier. The Random Forest classifier was shown to be the most reliable for the challenges presented.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **TITLE** | | **PAGE NO.** |
| **ABSTRACT** | | **6** |
| **LIST OF FIGURES** | | **8** |
| **LIST OF NPTEL/COURSE ERA/ UDEMY COURSES** | | **9** |
| **CHAPTER I** | **INTRODUCTION** | **10** |
| **1.1** | **Problem Statement** | |
| **1.2** | **Goal** | |
| **1.3** | **Objective** | |
| **CHAPTER II** | **LITERATURE REVIEW** **11** | |
| **CHAPTER III** | **EXPLORATORY DATA ANALYSIS** | **12** |
| **CHAPTER IV** | **MICRO PROJECT IMPLENTAION** | **15** |
| **CHAPTER V** | **RESULTS & DISCUSSION** | **17** |
| **CHAPTER VI** | **CONCLUSION & FUTURE SCOPE** | **17** |
| **COURSE CERTIFICATION** | | **18** |
| **REFERENCES** **19** | | |
| **PLAGIARISM REPORT** **20** | | |
| **APPENDIX** **21** | | |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURES** | **DETAILS** | **PAGE NO.** |
| 2.1 | Use case diagram | 11 |
| 3.1 | Df3.info() | 12 |
| 3.2 | Df3.describe() | 12 |
| 3.5 | Heat map | 14 |
| 4.1 | Total Matches played for season | 16 |
| 4.2 | Total matches Won By Team | 16 |

**LIST OF NPTEL/COURSE ERA/ UDEMY COURSES**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. NO.** | **COURSE NAME** | **COURSE DURATION** | **COURSE PLATFORM** |
| 1 | **Foundations of AI and Machine Learning** | >40 hours | COURSERA |

**CHAPTER – I INTRODUCTION**

## Problem Statement

In this project, we need to analyses the data to predict the team using machine learning algorithms. The Exploratory Data Analysis and visualization of the insight are very important for the selection and present the results. We will be required to clean the data as well as perform feature engineering to improve our results.

* 1. **Goal**

Maximum part of our current work is focused majorly in two directions:

* + - Cleaning data by making it worth using by dropping some columns and encoding some.
    - Predicting 11-member cricket squad by giving some features as an inputs and calculating their prediction

## Objective

The objective of this project is to predict a playing 11 squad of cricket considering data of more than 80+ nations and their players in national/international matches or leagues. With the help of machine learning techniques, we will predict 11 players according to the features.

.

### Description Of Data

This dataset contains complete information about All cricket events whether or national or international matches/leagues/ICC Events . Dataset prepared from the span of 2005-2019. There are many factors that can be analysed from this dataset.

We have two datasets(initially) namely:- 1. Match.csv (1950 rows\*25 columns)

2. Player\_performance .csv(1906 rows\*440 columns) Combined data:- (1950 rows\* 465 columns)

After cleaning of combined data, the size of datasets reduced to:-

1. Combined data:- (1870 rows\*35 Columns)

.

**CHAPTER II LITERATURE REVIEW**

The research paper that we have read, and the work done till date in field of predicting players is

basically regional and no such notable work has been done so far. They have prepared a dataset of some regional matches or teams; on the basis of that data they are predicting good players, but they are still not been able to figure out playing 11 using features. Also, some fantasy apps like dream11, fantasy cricket allows user to create their playing 11.

Muthuswamy and lam [1] predicted Indian bowler’s success against 7 international teams playing cricket and predicted nations against which they frequently played. They also predicted how many runs bowler is likely to give and wickets a bowler is going to take.

WikramaSinghe [2] predicted batsmen overall success in test cricket.

Barr and Kantor have [3] created a framework for determining performance of batsmen in limited overs.

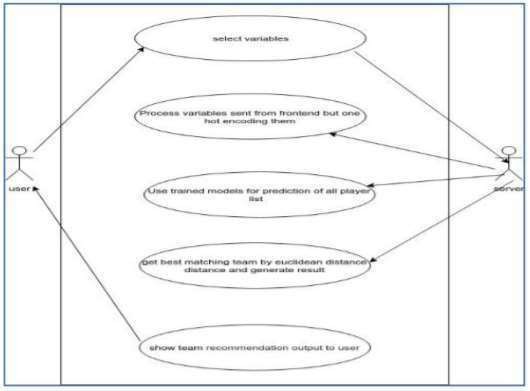


Fig 2.1 use case diagram

**CHAPTER III**

**EXPLORATORY DATA ANALYSIS**

With the use of summary statistics and graphical representations, exploratory data analysis refers to the crucial process of doing first investigations on data to uncover patterns, spot anomalies, test hypotheses, and check assumptions.

* To begin, I loaded the data collection and imported the relevant libraries (in this case, pandas, numpy, matplotlib, and seaborn).
* To get a closer look at the data, I used the pandas library's ".head()" function, which gives the first five observations in the data set. Similarly, “.tail()” retrieves the data set's last five observations.
* Using “.shape,” I calculated the total number of rows and columns in the data collection.
* Using df.info() o Only float and integer values are present in the data. o There are no null/missing values in any variable column.

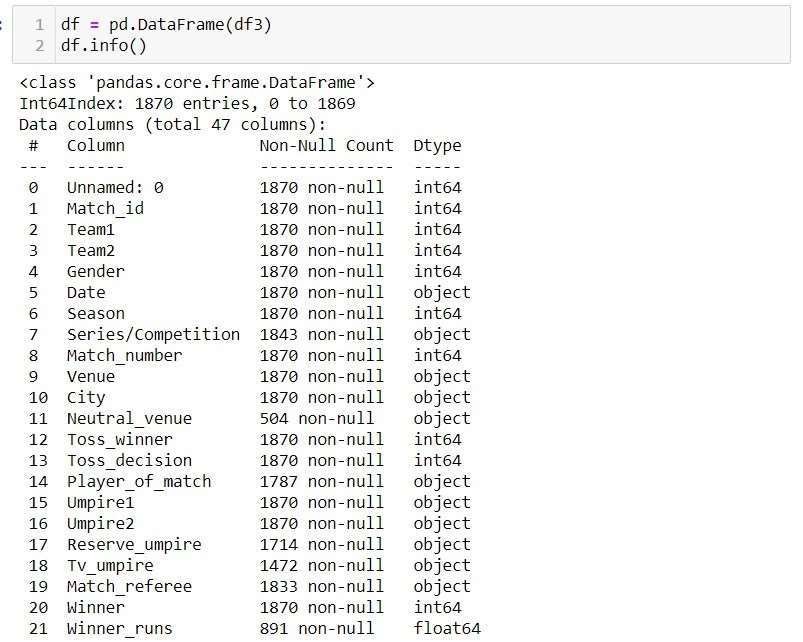


Fig 3.1 df3.info()

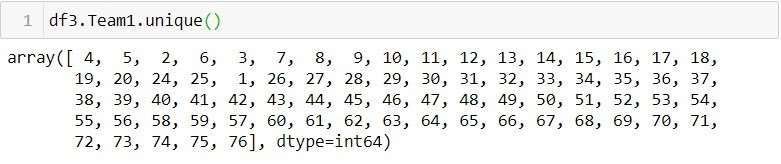
The pandas describe() method is extremely useful for obtaining various summary statistics. The count, mean, standard deviation, minimum and maximum values, and quantiles of the data are returned by this function

Fig3.2 df3.describe()

* The nature of the target variable/dependent variable is discrete and categorical.

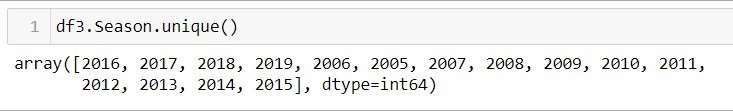
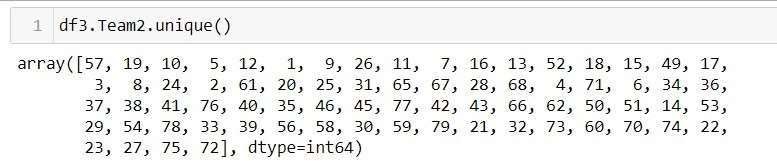


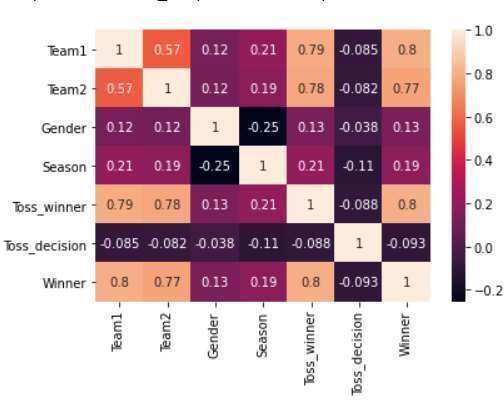
Fig3.3 df3.Season unique()

fig.3.4 team1.unique()



* Heat map visualisation o Positive correlation is represented by dark tones, whereas negative correlation is represented by lighter shades.

o If you set annot=True, you'll obtain values in grid-cells that show how characteristics are connected to one another.



### Fig3.5 heat map

**CHAPTER IV**

**MICRO PROJECT IMPLENTAION**

The implementation of this project is divided into the following steps –

* 1. Data collection

We were given both the datasets, so no data collection is done as such.

* 1. Data Preprocessing

We were given with two datasets namely match.csv and player\_performance.csv and combined them, after combining we drop some columns like Batting order, Inning batted, Bat Bowled, Bat Dismissed, Bat runs, 4s, 6s, ball faced, dots, batting team, inning bowled, over, runs, wicket, maiden, extra,etc. Also we used encoding for 6 columns:-

* + - Encoded values of team1/team2/toss\_winner/winner

{'Australia' : 1, 'Scotland' : 2, 'Papua New Guinea' : 3, 'Netherlands' : 4,'Kenya' : 5, 'Zimbabwe'

: 6, 'New Zealand' : 7, 'England' : 8, 'South Africa' : 9,'India' : 10,'Pakistan' : 11,'Bangladesh' : 12,'Adelaide Strikers' : 13,'Perth Scorchers' : 14,'Sydney Thunder' : 15, 'Brisbane Heat' : 16,'Sydney Sixers' : 17, 'Melbourne Stars' : 18, 'Hong Kong' : 19, 'West Indies' : 20,'Vanuatu' : 21, 'Maldives' : 22, 'Saudi Arabia' : 23,'Ireland' : 24, 'Afghanistan' : 25, 'Sri Lanka'

: 26, 'Malaysia' : 27, 'Denmark' : 28,'Bermuda' : 29, 'United States of America'

: 30, 'Italy' : 31, 'Japan' : 32,'United Arab Emirates' : 33,'Thailand' : 34,'Rising Pune Supergiant'

: 35, 'Mumbai Indians' : 36,'Kolkata Knight Riders' : 37,'Kings XI Punjab' : 38,'Delhi Daredevils' : 39,'Royal Challengers Bangalore' : 40, 'Sunrisers Hyderabad' : 41, 'Nigeria' : 42, 'Ghana' : 43, 'Botswana' : 44,'Sierra Leone' : 45,'Tanzania' : 46,'Qatar' : 47,'Bahrain' : 48, 'Melbourne Renegades' : 49, 'Rajasthan Royals' : 50, 'Chennai Super Kings' : 51,'Hobart Hurricanes' : 52, 'Delhi Capitals' : 53, 'Canada' : 54, 'Africa XI' : 55,'Deccan Chargers' : 56, 'Nepal' : 57, 'Argentina' : 58, 'Cayman Islands' : 59,'Pune Warriors' : 60, 'Singapore' : 61, 'Bhutan' : 62, 'Samoa' : 63, 'Belgium' : 64, 'Oman' : 65,'Kuwait' : 66, 'Jersey' : 67, 'China' : 68,

'South Korea' : 69, 'Uganda' : 70, 'Namibia' : 71,'Fiji' : 72, 'Norway' : 73, 'Guernsey' : 74, 'Suriname' : 75,'Gujarat Lions' : 76,'Zambia' : 77,'Asia XI' : 78, 'Kochi Tuskers Kerala' : 79}

* Encoded value of gender

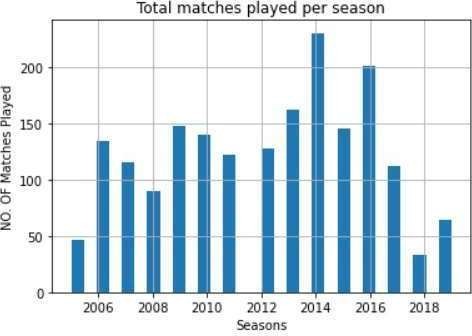
{'Female' : 0, 'Male' : 1}

* Encoded value of toss\_Decision

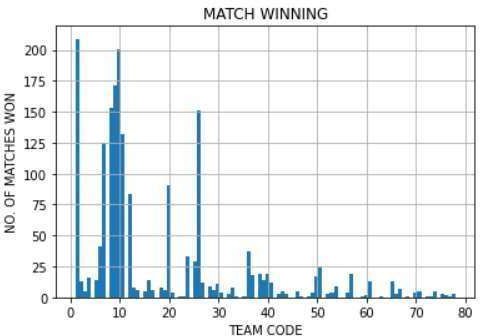
{'Field' : 0, 'Bat' : 1}

4.3 Feature selection

Features selection is done which can be used to build the model. The attributes used for feature selection are X coordinate and Y coordinate. On X we used Team1, Team2, Gender, Season , Toss winner, Toss decision, Winner as attributes and on Y coordinate, we use Player1, Player1, Player2, Player3, Player4, Player5, Player6, Player7, Player8, Player9, Player10, Player11, as an attribute.



**FIG4.1** Total Matches Played Per Season



**REFERENCES**

### Fig4.2 Total matches won By Team

**CHAPTER V**

## RESULTS & DISCUSSION

Reliability of data sources can affect predictions. Incomplete or inaccurate data can lead to poor model performance.

Cricket is influenced by numerous variables that may not be quantifiable, such as team morale, strategy, and unexpected events

Assess the importance of players’ performances in determining match outcomes. Star players’ contributions could be more predictive than overall team statistics.

**CHAPTER VI**

**CONCLUSION & FUTURE SCOPE**

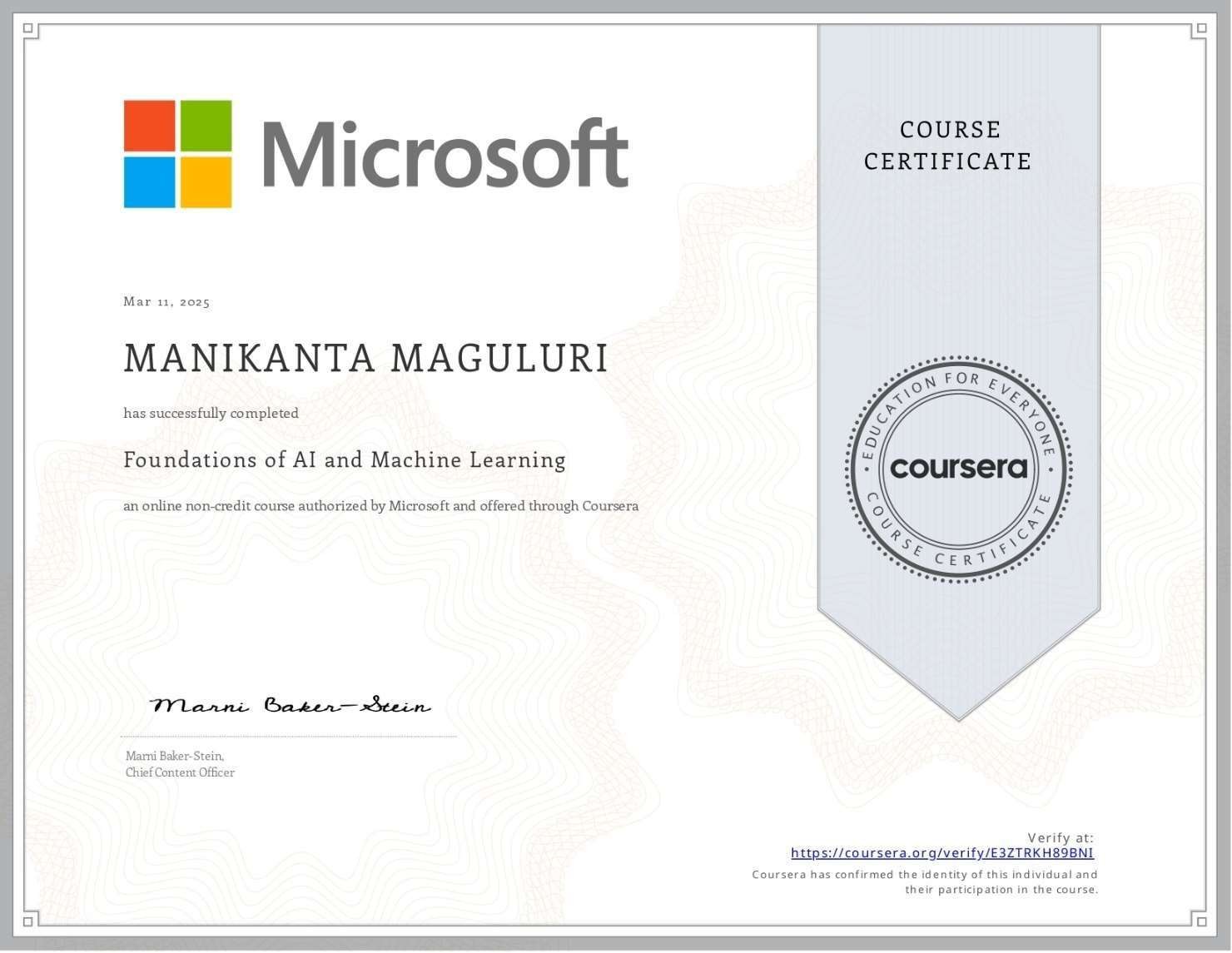
After implementing the algorithms, this section covers the result of accuracy comparison of Random Forest classification, Naive Bayes, KNN, Logistic regression, Support Vector Machine (SVM) AND Decision tree. All these algorithms come under the classification machine learning algorithm category. The algorithm was run to predict 11 cricket players.

The results obtained by following the various processes that come under machine learning. Data preprocessing includes dropping some rows in which cases come under more than 10 rows and also Converting string variables into numerical. After completing the data processing work, the dataset was divided into testing and training sets.

FUTURE SCOPE :

1. Data, in this case, is mixed for e.g. , test and limited over statistics are mixed up , so we can classify them and can choose playing 11 for each format.
2. Accuracy of team is quite low which can be improved further by more data cleaning and encoding.
3. We can also predict playing XI for each country.

# COURSE CERTIFICATION

****

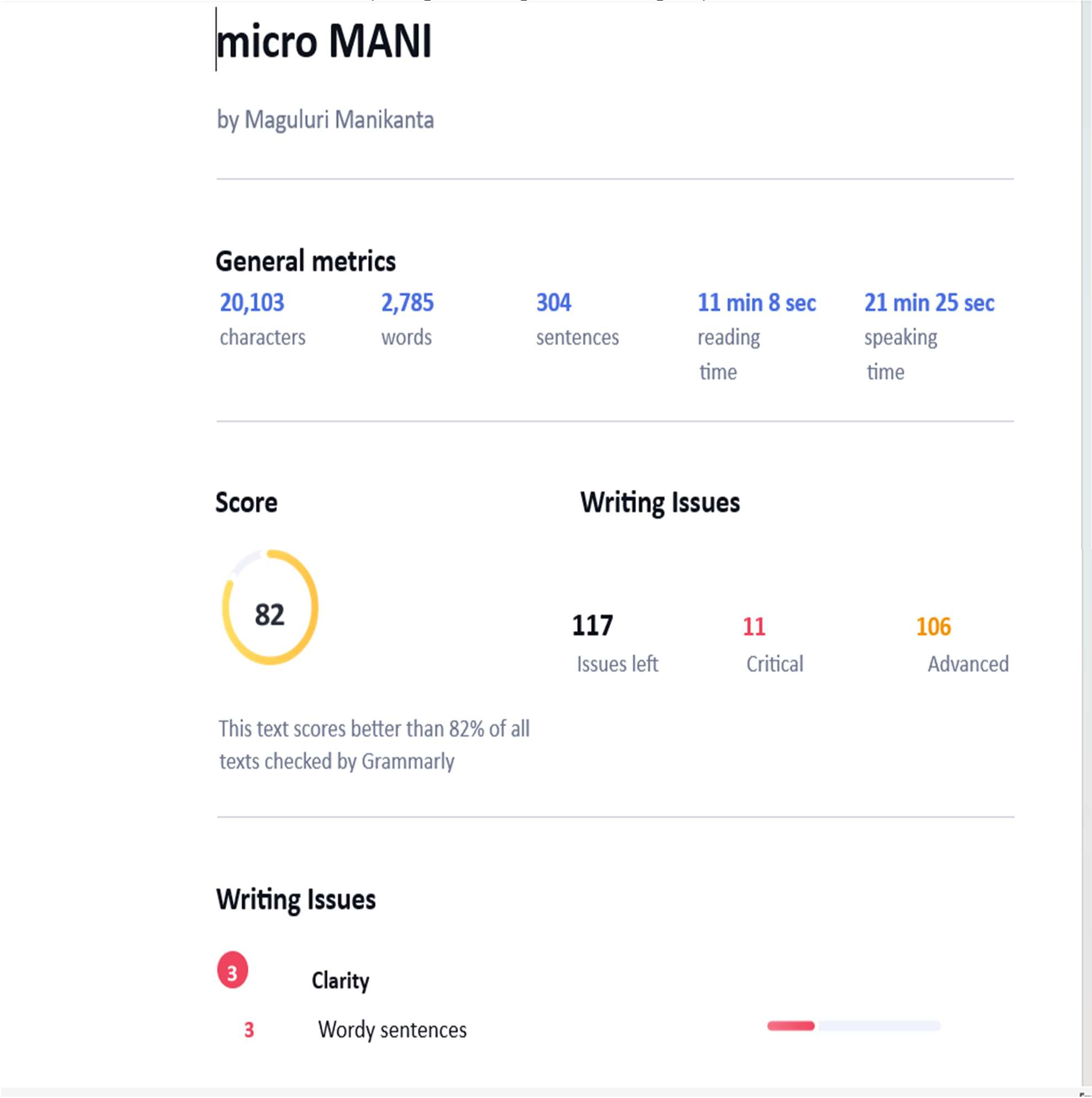
**REFERENCES**

1. S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks", Proceedings of IIE Annual Conference and Expo 2008, pp.1391-1395/
2. I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket", Journal of Human Sport & Exercise, vol. 9, no. 4, pp. 744-751, May 2014.
3. G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket", Operational Research Society, vol. 55, no. 12, pp. 1266-1274, December 2004.

1. Developing automated machine learning approach for fast and robust crop yield prediction using a fusion of remote sensing, soil, and weather dataset

**PLAGIARISM REPORT**

**(Project Report & Paper)**





**INTERNAL QUALITY ASSURANCE CELL MICRO PROJECT AUDIT REPORT**

This is to certify that the micro project work entitled “CRICKET TEAM PREDITION USING MACHINE LEARNING” categorized as an internal project done by MAGULURI MANIKANTA of the Department of Computer Science and Engineering, under the guidance of S. RESHNI during the Even semester of the academic year 2024 - 2025 are as per the quality guidelines specified by IQAC.

**Quality Grade**

**Deputy Dean (IQAC)**

**Administrative Quality Assurance Dean (IQAC) APPENEDIX**

**(Project Code)**

import numpy as np import pandas as pd

%matplotlib inline

import matplotlib.pyplot as plt import seaborn as sns

import pandas as pd

# Import RandomForestClassifier

from sklearn.ensemble import RandomForestClassifier

# Import train\_test\_split function

from sklearn.model\_selection import train\_test\_split, GridSearchCV

#Import scikit-learn metrics module for accuracy calculation from sklearn import metrics

#Import classification\_report

from sklearn.metrics import classification\_report from sklearn.preprocessing import LabelEncoder from sklearn.datasets import make\_classification df= pd.read\_csv("Match.csv")

df2= pd.read\_csv("Player\_performance.csv")

df3 = pd.merge(df,df2, how='inner', on = 'Match\_id') df3.drop(['Unnamed: 0','Date','Umpire1','Umpire2',

'Reserve\_umpire','Tv\_umpire','Match\_referee','Neutral\_venue', 'Method','Outcome','Winner\_runs','Winner\_wickets'], axis = 1)

df3[pd.isnull(df3['Team1'])]

df3[pd.isnull(df3['Team2'])]

df3[pd.isnull(df3['Venue'])]

df3[pd.isnull(df3['Winner'])] df = pd.DataFrame(df3) df.info()

tw=df3.groupby(["Toss\_winner"]).size() print("Total tosses won by team") tw.head() mw=df3.groupby(["Winner"]).size() print("Total Matches won by team") mw.head()

sns.heatmap(data=df3.drop(['Match\_id','Unnamed: 0','Date','Umpire1','Umpire2', 'Reserve\_umpire','Tv\_umpire','Match\_referee','Neutral\_venue', 'Method','Outcome','Winner\_runs','Winner\_wickets','Match\_number','Venue','City'],

axis=1).corr(), annot=True) df3.Season.unique() s=df3.Season a=print(s.value\_counts())

count= {'Season':['2005','2006','2007','2008','2009','2010','2011','2012',

'2013','2014','2015','2016','2017','2018','2019',],

'Total Matches played':['46','134','115','90','148','140','122','128'

,'162','230','145','201','112','33','64',]}

df4=pd.DataFrame(count) df4 df3['Season'].hist(bins=30) plt.xlabel("Seasons")

plt.ylabel("NO. OF Matches Played") plt.title('Total matches played per season') df3['Winner'].hist(bins=100) plt.xlabel("TEAM CODE") plt.ylabel("NO. OF TOSSES WON") plt.title('TOSS WINNING')

df3['Winner'].hist(bins=100) plt.xlabel("TEAM CODE") plt.ylabel("NO. OF MATCHES WON") plt.title('MATCH WINNING')

var\_mod = ['City','Toss\_winner','Toss\_decision','Venue'] le = LabelEncoder()

for i in var\_mod:

df[i] = le.fit\_transform(df[i]) df.head(5)

df3.Team1.unique() df3.Team2.unique() y = df3[['Players\_1']] y

#I will split data 20%

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2

, random\_state=10)

clf = RandomForestClassifier(max\_depth=1, random\_state=1) clf.fit(X, y)

clf = RandomForestClassifier(random\_state=0).fit(X\_train, y\_train) clf.predict\_proba(X\_test[:11])

p1=clf.predict(X\_test[:1]) print("Player 1 is: "+p1) a1=(clf.score(X\_test, y\_test))\*100 a1=round(a1,2)

print("Accuracy of Player 1 is: ",a1,"%") print("\033[1m" +"FINAL TEAM:-")

print("Player 1 is: "+p1) print("Player 2 is: "+p2) print("Player 3 is: "+p3) print("Player 4 is: "+p4) print("Player 5 is: "+p5) print("Player 6 is: "+p6) print("Player 7 is: "+p7) print("Player 8 is: "+p8) print("Player 9 is: "+p9) print("Player 10 is: "+p10) print("Player 11 is: "+p11) print("\n")

print("\033[1m" +"ACCURACY OF ALL PLAYERS:-")

print("-> Accuracy of Player 1 is: ",a1,"%") print("-> Accuracy of Player 2 is: ",a2,"%") print("-> Accuracy of Player 3 is: ",a3,"%") print("-> Accuracy of Player 4 is: ",a4,"%") print("-> Accuracy of Player 5 is: ",a5,"%") print("-> Accuracy of Player 6 is: ",a6,"%") print("-> Accuracy of Player 7 is: ",a7,"%") print("-> Accuracy of Player 8 is: ",a8,"%") print("-> Accuracy of Player 9 is: ",a9,"%") print("-> Accuracy of Player 10 is: ",a10,"%") print("-> Accuracy of Player 11 is: ",a11,"%")

print("\n")

at= (a1+a2+a3+a4+a5+a6+a7+a8+a9+a10+a11)/11 at=round(at,2)

print("Total Accuracy: ",at,"%") print("\033[1m"+"RANDOM FOREST CLASSIFICATION = ",at,"%")

print("\n")

print("LOGISTIC REGRESSION = 15.1%")

print("\n")

print("K-NEIGHBOUR CLASSIFICATION (KNN) = 18.2%")

print("\n")

print("SUPPORT VECTOR MACHINE CLASSIFICATION (SVM) = 13.4%")

print("\n")

print("DECISION TREE CLASSIFICATION = 14.4%")

print("\n")

print("NAVE BYES CLASSIFICATION = 16.1%")