Predictive Analysis in T-20 Cricket: Estimation and Prediction of fantasy points for IPL Players

MSc Research Project

Data Analytics

Pratheek Gogate
Student ID: X22159789

School of Computing

National College of Ireland

Supervisor: Mayank Jain

Student Name:	Pratheek Gogate		
Student ID:	X22159789		
Programme:	MSc Data analytics	**Year:**	2023-2024
Module:	MSc Research Project		
Supervisor:	Mayank Jain		
Submission Due Date:	14/12/2023		
Project Title:	Predictive Analysis in T-20 Cricket: Estimation and Prediction of fantasy points for IPL Players		

Word Count: 912   **Page Count:** 6

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Pratheek Gogate

Date: 14/12/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predictive Analysis in T-20 Cricket: Estimation and Prediction of fantasy points for IPL Players

Pratheek Gogate

Student ID: X22159789

1 Introduction

In data analytics there is always something called a backend, this includes hardware, software and basically everything which would help in smooth running of the project. So, this documentation includes all the details related to the system we used and software and different libraries for smooth running of the code.

2 System Specifications:

2.1 Hardware Requirements

This project was done on an Acer Aspire A515-57G laptop.
The laptop's hardware specifications are as follows:

- Processor: Intel Core i5 12th Generation
- Operating System: Windows 11
- RAM: 16 GB
- Storage: 512 GB

2.2 Software Requirements

During the time of this research, the following tools were used:

- Python version 3.11.1
- Microsoft Excel
- Jupyter Notebook
- Microsoft Word

2.2.1 Installation of python 3.11.1:

The required python version needs to be downloaded from official link[1].

2.2.2 Installation of Jupyter notebook:

Jupyter Notebook is the most preferred for executing the code as we can execute the code step by step in this and output of the each cell is displayed which would help us to easily debug the code. This will be installed from the terminal:

pip install jupyter

Then we can launch the jupyter notebook by using following command from terminal:

jupyter notebook

---

[1] https://www.python.org/ftp/python/3.11.1/python-3.11.1-amd64.exe

2.2.3 Required libraries Installation:

We need to install and import some of the important python libraries as they will help us in an analysing data.

First, they need to be installed using "pip install name" command in terminal or maybe we can do it in Jupiter notebook as well and then those needs to be imported using "import name".

For example, if we want to install NumPy then it would be:

pip install numpy

Then to import it, we need to write:

import numpy as np

Following are the list of libraries we used:

- Pandas
- NumPy
- Matplotlib
- scikit-learn
- TensorFlow

Software version summary table:

Software Name	Version	Download URL
Python	3.11.1	https://www.python.org/ftp/python/3.11.1/python-3.11.1-amd64.exe
Jupyter	1.0.0	https://jupyter.org/install

Libraries version summary table:

Library Name	Version	Download URL
numpy	1.26.2	https://files.pythonhosted.org/packages/dd/2b/205ddff2314d4eea852e31d53b8e55eb3f32b292efc3dd86bd827ab9019d/numpy-1.26.2.tar.gz
pandas	2.1.3	https://files.pythonhosted.org/packages/86/ff/662dde2193fc93b8547b073db20472b9676f944d907247a46c9c5bc45bfc/pandas-2.1.3.tar.gz
matplotlib	3.8.2	https://files.pythonhosted.org/packages/fb/ab/38a0e94cb01dacb50f06957c2bed1c83b8f9dac6618988a37b2487862944/matplotlib-3.8.2.tar.gz
scikit-learn	1.3.2	https://files.pythonhosted.org/packages/88/00/835e3d280fdd7784e76bdef91dd9487582d7951a7254f59fc8004fc8b213/scikit-learn-1.3.2.tar.gz
TensorFlow	2.15.0	https://files.pythonhosted.org/packages/93/21/9b035a4f823d6aee2917c75415be9a95861ff3d73a0a65e48edbf210cec1/tensorflow-2.15.0-cp311-cp311-win_amd64.whl

3   Implementation

3.1 **Data Collection:** Data is collected from kaggle and it is imported in jupyter notebook for further analysis.

```
In [3]: cricket_data.tail(5)
```

	match_id	season	start_date	venue	innings	ball	batting_team	bowling_team	striker	non_striker	...	extras	wides	noballs	byes	legbyes
200659	1254086	2021	2021-05-02	Narendra Modi Stadium, Ahmedabad	2	17.2	Delhi Capitals	Punjab Kings	SO Hetmyer	S Dhawan	...	0	NaN	NaN	NaN	NaN
200660	1254086	2021	2021-05-02	Narendra Modi Stadium, Ahmedabad	2	17.3	Delhi Capitals	Punjab Kings	SO Hetmyer	S Dhawan	...	0	NaN	NaN	NaN	NaN
200661	1254086	2021	2021-05-02	Narendra Modi Stadium, Ahmedabad	2	17.4	Delhi Capitals	Punjab Kings	SO Hetmyer	S Dhawan	...	0	NaN	NaN	NaN	NaN
200662	1254086	2021	2021-05-02	Narendra Modi Stadium, Ahmedabad	2	17.5	Delhi Capitals	Punjab Kings	SO Hetmyer	S Dhawan	...	1	1.0	NaN	NaN	NaN
200663	1254086	2021	2021-05-02	Narendra Modi Stadium, Ahmedabad	2	17.6	Delhi Capitals	Punjab Kings	SO Hetmyer	S Dhawan	...	1	1.0	NaN	NaN	NaN

5 rows × 22 columns

Figure 1: Sample data from dataset

3.2 Data Preprocessing:

After the dataset is downloaded, we import it in jupyter notebook and we filter the data which we want by using particular player conditions and we will explore more about the columns present in it and make necessary changes for model building.

```
In [10]: #Checking the Missing Value information
         dhoni_batting_stats.isnull().sum()

Out[10]: start_date           0
         match_id             0
         venue                0
         innings              0
         bowling_team         0
         ball                 0
         runs_off_bat         0
         boundary_count       0
         sixer_count          0
         half_century_count   0
         century_count        0
         strike_rate          0
         fantasy_points       0
         dtype: int64
```

Figure 2: Bull value count

3.3 Data Transformation:

Before actually building the model we need to take care about some of the columns which we cant directly use and in that case we need to transform them into different datatypes such that we do not have any issues while building the model.

3

```
Out[17]:  0       22
          1       15
          2       15
          3       12
          4       15
                  ..
          181      5
          182     36
          183     36
          184     36
          185     36
          Name: venue_encoded, Length: 186, dtype: int32
```

Figure 3: Transformed values of venue

4   Model Building:

After transformation and feature selection first, we need to split the data into train and test and then we need to build the model. In our case we used Machine learning techniques to estimate the fantasy points and used deep learning techniques to predict the fantasy points.

5   Evaluation:

Once the model is built, we need to evaluate the model with different evaluation metrices and evaluate whether that model fits the data or not. In our case we used r squared value, RMSE, MSE to evaluate.