

# Configuration Manual

MSc Research Project  
Programme Name

Sahil Chordia  
Student ID: x20203993

School of Computing  
National College of Ireland

Supervisor: Mr. Aaloka Anant

National College of Ireland  
Project Submission Sheet  
School of Computing



<b>Student Name:</b>	Sahil Chordia
<b>Student ID:</b>	x20203993
<b>Programme:</b>	Programme Name
<b>Year:</b>	2022
<b>Module:</b>	MSc Research Project
<b>Supervisor:</b>	Mr. Aaloka Anant
<b>Submission Due Date:</b>	15/12/2022
<b>Project Title:</b>	Configuration Manual
<b>Word Count:</b>	XXX
<b>Page Count:</b>	8

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.

**ALL** 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.

<b>Signature:</b>	
<b>Date:</b>	31st January 2023

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

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

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

<b>Office Use Only</b>	
Signature:	
Date:	
Penalty Applied (if applicable):	

# Configuration Manual

Sahil Chordia  
x20203993

## 1 Introduction

An explanation of the environment set up for the research project is given in detail in this configuration manual. This includes information about the programming language used, system configuration, and descriptions of the libraries and packages that were utilized. The Configuration Manual will be separated into the following six sections:

- Environmental Setup
- Libraries Required
- Data-set
- User Interface
- Implementation
- Code Repository

## 2 Environment Set-Up

### 2.1 System Specification

- Processor:- 11th Gen Intel(R) Core(TM) i3-1115G4 @ 3.00GHz 3.00 GHz
- Installed RAM:- 8.00 GB (7.76 GB usable)
- RAM:- 8.00 GB (7.76 GB usable)Installed RAM:- 8.00 GB (7.76 GB usable)
- SSD:- 256 GB

### 2.2 Technical Specifications

#### 2.2.1 Python

- Version 3.7.4
- Google Colab which is an online Python based online user-interface, was used throughout this research as it provides all the libraries pre-installed. Which avoid getting errors due to previous versions of libraries installed. The other reason to use google colab was that it provide a strong GPU for processing.

## 3 Data Sources

### 3.1 Libraries Required

- Numpy: - NumPy was used to carry out all the mathematical calculations and operations in this research.
- Pandas: - Pandas was used to read and work with data-set which was in the form of .csv file
- Seaborn: - Seaborn was used for plotting graphs and visuals.
- Matplotlib: - Matplotlib was used for plotting graphs and visuals.
- Plotly: - Plotly was used for plotting graphs and visuals.
- Scipy: - SciPy It provides more utility functions for stats, optimization, and signal processing
- Sk-Learn: - The most effective and reliable Python machine learning library is called Sklearn (Skit-Learn). Through a Python consistency interface, it offers a variety of effective tools for statistical modeling and machine learning, including classification, regression, clustering, and dimensionality reduction.

### 3.2 Importing Libraries

Figure 1 shows how the libraries used in this research were imported in the colab notebook.

```
[ ] import numpy as np
import pandas as pd
from pandas import DataFrame
from scipy.spatial import distance
import warnings
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
from sklearn.preprocessing import MinMaxScaler
from collections import Counter
from imblearn.over_sampling import SMOTE, ADASYN
from sklearn.preprocessing import RobustScaler, StandardScaler, LabelEncoder, LabelBinarizer
from sklearn.preprocessing import StandardScaler
from sklearn.naive_bayes import GaussianNB
```

Figure 1: Importing Libraries

## 4 Section 4

Data was downloaded from the above link, which is available in a comma-separated file (.csv).

Data Link: - <https://www.kaggle.com/datasets/umeshkumar017/fifa-21-player-and-formation-analysis>

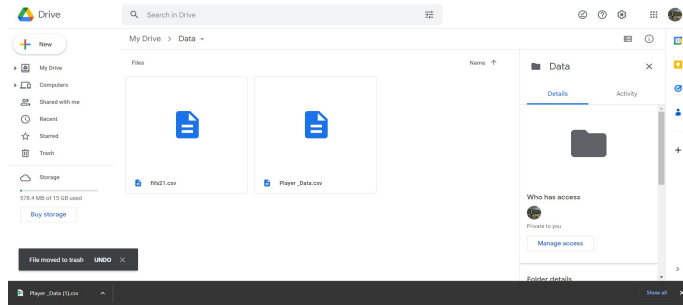


Figure 2: Mounting Drive

Data was stored in my drive showne link is given below and a image of data in google drive is 2:- Data-set link:- `’/content/drive/MyDrive/Data/fifa21.csv’` As I have used Google Colab for this research, I have uploaded my data to Google Drive and then mounted/connected my Google Drive to my Colab notebook, as shown in image 3.

```
[2] from google.colab import drive
    drive.mount('/content/drive')

Mounted at /content/drive
```

Figure 3: Mounting Drive

Once the data file was uploaded, I passed the link to the data as a parameter for `pd.read_csv()`, as shown in image 4.

```
[3] # Reading and checking the shape of the data
    data = pd.read_csv('/content/drive/MyDrive/Data/fifa21.csv')
    data.shape

(18541, 92)
```

Figure 4: Reading Data

After data has been linked to a Colab Notebook, I have printed the data to find the columns and attributes of the data.

```
data.head()
```

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	...	Penalties	Composure	Defensiv. Awareness
0	0	252093	Facundo Pellistri	18	<a href="https://cdn.sofifa.com/players/252093/60.png">https://cdn.sofifa.com/players/252093_60.png</a>	Uruguay	<a href="https://cdn.sofifa.com/flags/uy.png">https://cdn.sofifa.com/flags/uy.png</a>	71	87	Peñarol	...	66.0	61.0	35
1	1	179813	Edinson Cavani	32	<a href="https://cdn.sofifa.com/players/179813/60.png">https://cdn.sofifa.com/players/179813_60.png</a>	Uruguay	<a href="https://cdn.sofifa.com/flags/uy.png">https://cdn.sofifa.com/flags/uy.png</a>	86	85	Paris Saint-Germain	...	85.0	80.0	57
2	2	245541	Giovanni Reyna	17	<a href="https://cdn.sofifa.com/players/245541/60.png">https://cdn.sofifa.com/players/245541_60.png</a>	United States	<a href="https://cdn.sofifa.com/flags/us.png">https://cdn.sofifa.com/flags/us.png</a>	68	87	Borussia Dortmund	...	50.0	55.0	30
3	3	233419	Raphael Dias Ribeiro	23	<a href="https://cdn.sofifa.com/players/233419/60.png">https://cdn.sofifa.com/players/233419_60.png</a>	Brazil	<a href="https://cdn.sofifa.com/flags/br.png">https://cdn.sofifa.com/flags/br.png</a>	81	85	Flamengo	...	73.0	79.0	45
4	4	198710	James Rodríguez	28	<a href="https://cdn.sofifa.com/players/198710/60.png">https://cdn.sofifa.com/players/198710_60.png</a>	Colombia	<a href="https://cdn.sofifa.com/flags/co.png">https://cdn.sofifa.com/flags/co.png</a>	82	82	Everton	...	81.0	87.0	52

5 rows x 92 columns

Figure 5: Data

## 5 Exploratory Data Analysis

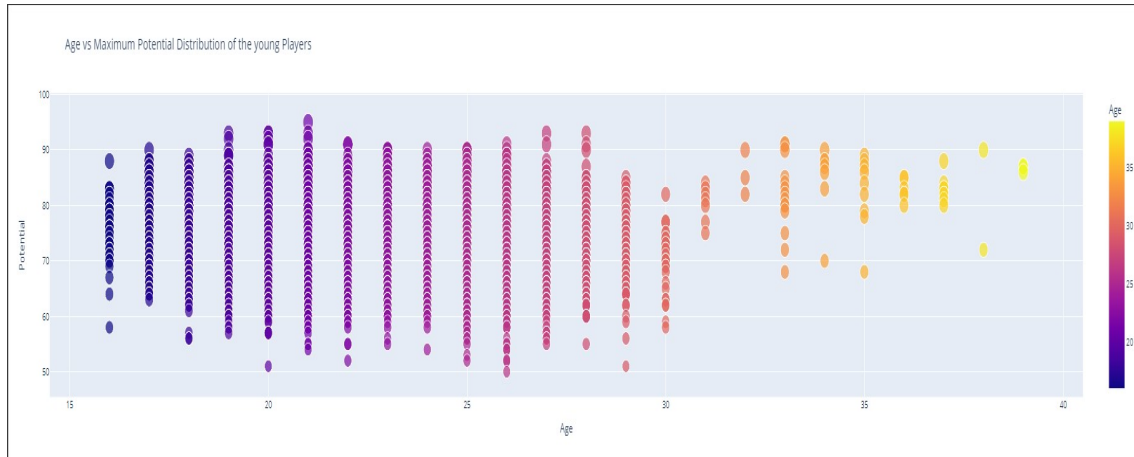


Figure 6: Age vs Potential Distribution of players

Figure 6 shows the Age wise potential distribution in the data set.

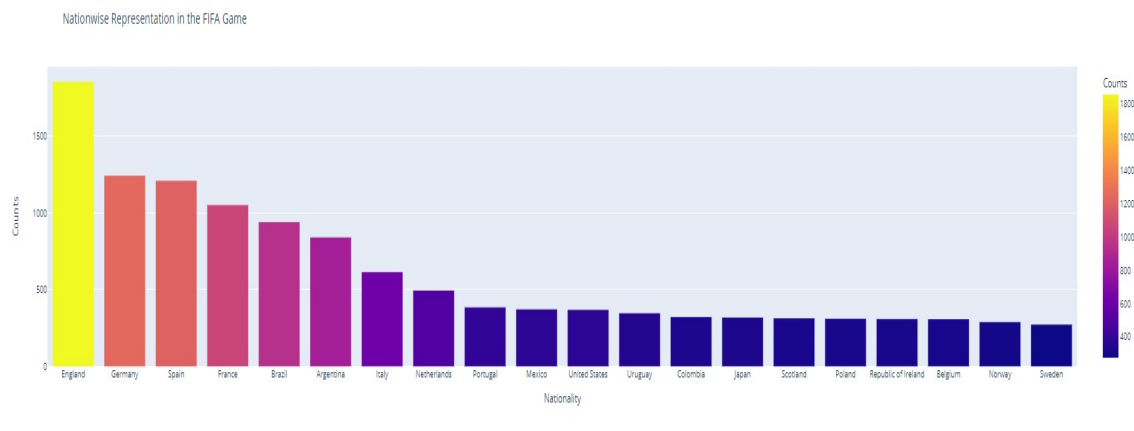


Figure 7: Nation-wise distribution of players

Figure 7 shows the Nation wise Player distribution in the data set.

Name	Age	Position	Overall	21	Oier Sanjurjo Maté	34	RDM	77
0 David Josué Jiménez Silva	34	CAM	86	22	Ferran Torres García	20	RM	81
1 Carles Puyol Saforcada	35	CB	83	23	Sergi Enrich Ametller	30	RS	77
2 Sergio Busquets Burgos	31	CDM	87	24	Roger Martí Salvador	29	RS	77
3 Iago Falqué Silva	30	CF	77	25	José María Callejón Bueno	33	RW	84
4 Xavier Hernández Creus	34	CM	86	26	Héctor Bellerín Moruno	25	RWB	80
5 David De Gea Quintana	29	GK	86	27	Iago Aspas Juncal	32	ST	84
6 Jordi Alba Ramos	31	LB	86					
7 Iñigo Martínez Berridi	29	LCB	80					
8 Marc Bartra Aregall	29	LCB	80					
9 Luis Alberto Romero Alconchel	27	LCM	85					
10 Luis Milla Manzanares	25	LDM	76					
11 David López Silva	30	LDM	76					
12 Marc Cucurella Saseta	21	LM	81					
13 Francisco Alcácer García	26	LS	81					
14 Mikel Oyarzabal Ugarte	23	LW	84					
15 Jonathan Castro Otto	26	LWB	81					
16 Daniel Carvajal Ramos	28	RB	86					
17 Sergio Ramos García	34	RCB	89					
18 Thiago Alcántara	29	RCM	85					
19 Daniel Parejo Muñoz	31	RCM	85					
20 Jorge Resurrección	28	RCM	85					

Figure 8: England Players

Figure 8 Shows list of players playing for England with their age, potential and position.

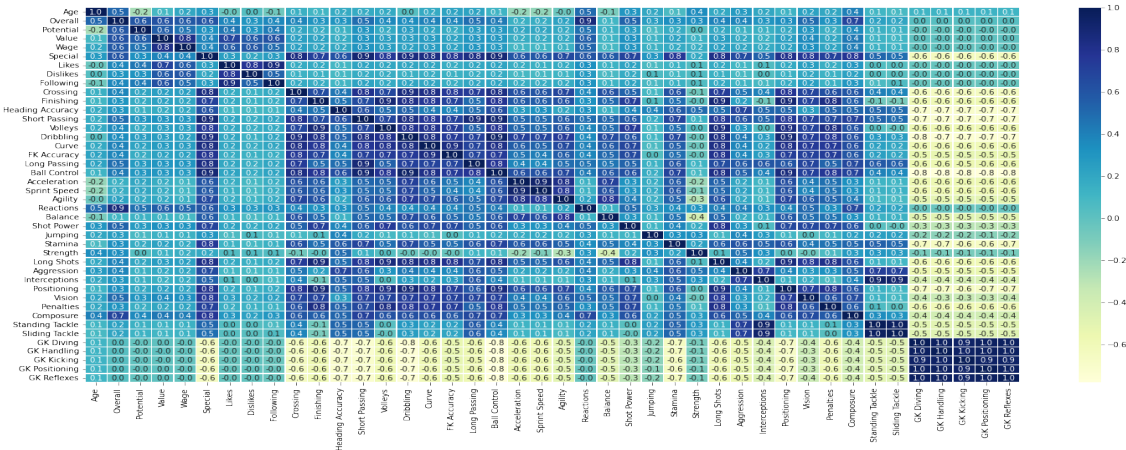


Figure 9: Correlation Matrix

Figure 9 displays the correlation matrix of this data.

## 6 Experiments

In this section, I have added snapshots of experiments conducted in my research.





## 6.2.2 Modeling for Experiment 2

```
[ ] from sklearn.model_selection import RandomizedSearchCV
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.naive_bayes import GaussianNB
mdl = GaussianNB()
mdl.fit(X_train, y_train)
y_pred = mdl.predict(X_test)
print('Training Accuracy: %f' % mdl.score(X_train, y_train))
print('Cross Validation score')
cv = cross_val_score(estimator = mdl, X = X_train, y = y_train, cv = 7)
SVC_cv = 7
print('k fold cross validation :', SVC_cv)
from sklearn.metrics import accuracy_score
mdl_a = accuracy_score(y_pred, y_test)
print('accuracy : %f' % mdl_a)
from sklearn.metrics import precision_score
mdl_p = precision_score(y_pred, y_test, average='weighted')
print('precision : %f' % mdl_p)
from sklearn.metrics import recall_score
mdl_r = recall_score(y_pred, y_test, average='weighted')
print('recall : %f' % mdl_r)
from sklearn.metrics import f1_score
mdl_f1 = f1_score(y_test, y_pred, average='weighted')
print('f1 score : %f' % mdl_f1)

Training Accuracy: 0.535
Cross Validation score: 0.5353333333333333
accuracy : 0.5353333333333333
precision : 0.5353333333333333
recall : 0.5353333333333333
f1 score : 0.5353333333333333
```

Figure 13: Modeling for experiment 2

## 6.3 Experiment 3

### 6.3.1 Feature selection for Experiment 3

```
df_new = df[df['Value'] > 0]
[ ] x = df.drop(['Potential', 'Standing Tackle', 'Or Kicking', 'Or Positioning', 'Or Handling'], axis = 1)
y = df_new['Potential']
print('Shape of x :', x.shape)
print('Shape of y :', y.shape)
Shape of x : (1803, 45)
Shape of y : (1803,)
x.columns
Index(['Name', 'Age', 'Nationality', 'Overall', 'Club', 'Value', 'Wage',
       'Special', 'Preferred Foot', 'Weak Foot', 'Skill Moves',
       'International Reputation', 'Work Rate', 'Body Type', 'Position',
       'Height', 'Weight', 'Likes', 'Dislikes', 'Training', 'Crossing',
       'Finishing', 'Heading Accuracy', 'Short Passing', 'Volleys',
       'Dribbling', 'Curve', 'Free Accuracy', 'Long Passing', 'Goal Control',
       'Acceleration', 'Sprint Speed', 'Agility', 'Reactions', 'Balance',
       'Shot Power', 'Shooting', 'Strength', 'Long Shots',
       'Aggression', 'Interceptions', 'Positioning', 'Vision', 'Penalties',
       'Composure', 'Sliding Tackle', 'Or Diving', 'Or Setpieces'],
      dtype='object')
```

Figure 14: Features for experiment 3

### 6.3.2 Modeling for Experiment 3

```
[ ] from sklearn.linear_model import LinearRegression
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.naive_bayes import GaussianNB
mdl = LinearRegression()
mdl.fit(X_train, y_train)
y_pred = mdl.predict(X_test)
print('Training Accuracy: %1.3f.' % mdl.score(X_train, y_train))
#10-fold cross validation score
cv = cross_val_score(estimator = mdl, X = X_train, y = y_train, cv = 8)
SVC_cv = 8
print("k fold cross validation :", SVC_cv)

Training Accuracy: 0.899.
k fold cross validation : 8
```

Figure 15: Modeling for experiment 3

## 7 Results

Figure 16 summarizes the results of all three experiments conducted through-out this research.

	Training Accuracy	Testing Accuracy	Precision	Recall	F-1 Score
Experiment 1	09%	08%	22%	08%	07%
Experiment 2	54%	53%	59%	53%	54%
Experiment 3	80%	78%	81%	78%	79%

Figure 16: Evaluation Summary

Figure 17 displays the results in the form of a bar-chart.

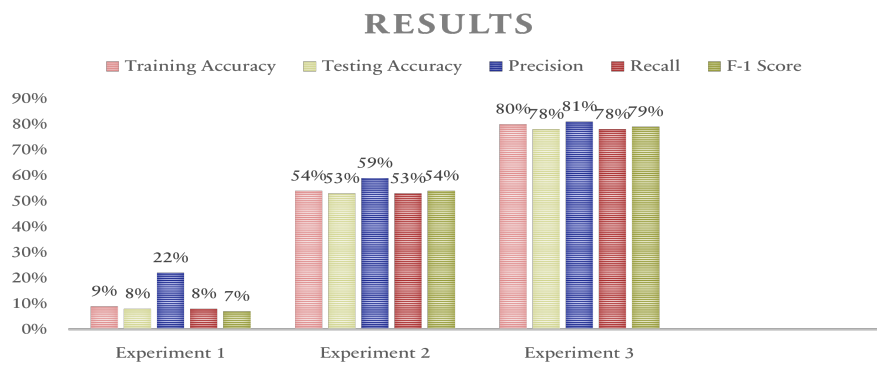


Figure 17: Results Chart