Data Intensive Computing CSE 587

Project Phase-3

TITLE: FIFA PLAYER ANALYSIS

NAME	UB ID
HarzeenAfshan	harzeena
Dhanush Roy Soma	dhanushr
Pavan Venkata Ramana Vadlamudi	pavanven

Problem Statement:

Our aim with the FIFA 2021 dataset is to examine and comprehend the various characteristics connected to every player who participated in FIFA. This contains comprehensive player data, including rating overall, potential, skill set, and physical characteristics. The collection additionally furnishes characteristics regarding the nationality, age, favorite foot, and club affiliation of the play.

AIM:

The project aims to develop a web-based application using Streamlit that predicts the position class (attacker, defender, or midfielder) of a football player based on their attributes. It utilizes machine learning techniques, specifically a Random Forest classifier, trained on FIFA-21 player data to make predictions. Users can input values for various player attributes, such as attacking, defending, skill, movement, mentality, and goalkeeping, through sliders in the web interface. Upon clicking the "Predict" button, the application generates a prediction for the player's position class.

1.(a) Working instructions to demo/use the finished product.

1. Before going to do any steps for implement install the required packages like streamlit, numpy, pandas, matplotlib, seaborn and sk-learn.

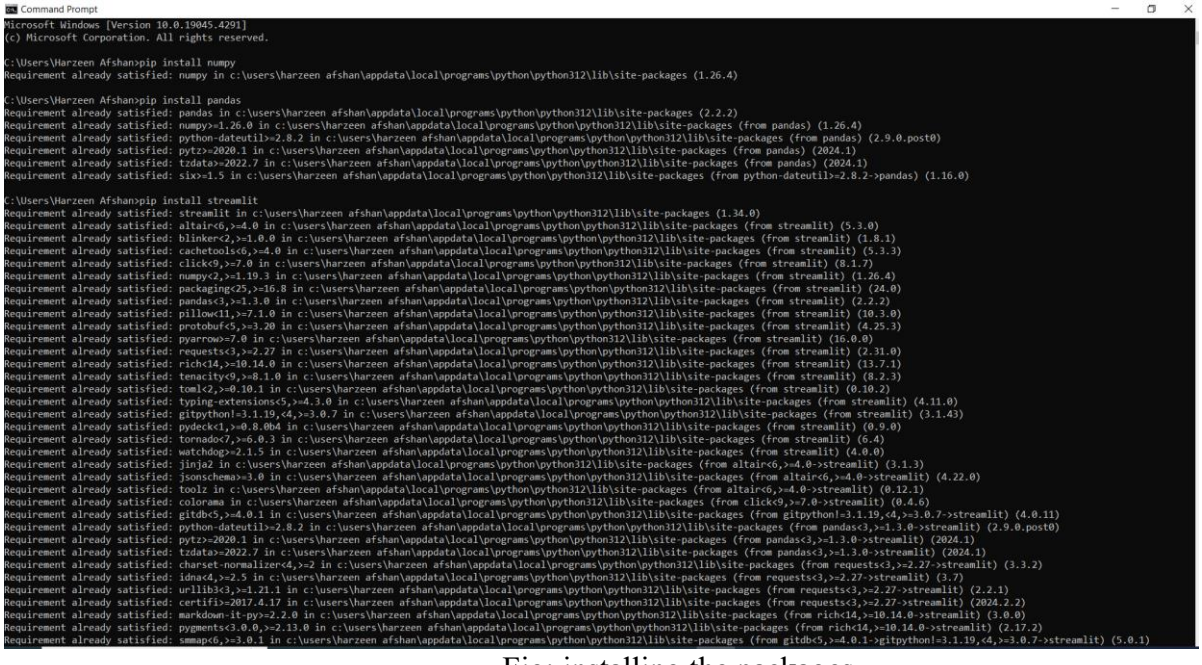


Fig: installing the packages

2. Extract the zip file that has been uploaded. The python file consisting of the code for phase – 3 can be found inside the folder named “phase3” .
3. Now, as we have developed using streamlit, we run the python file from the terminal.
4. Open the terminal, change the directory to the directory that consists of our python file (this folder also consists of train and test datasets).

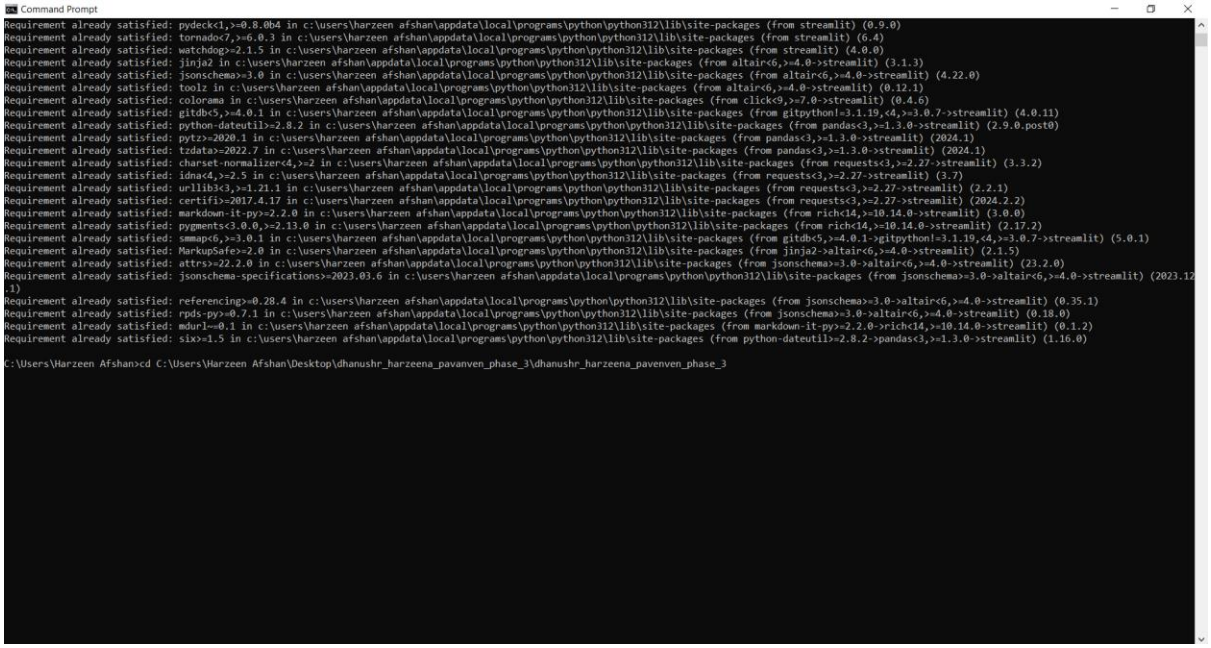


Fig: changing the directory

Here the folder dhanushr_harzeena_pavanven_phase_3 contains the phase3.py file which has the source file.

5. Once, we changed the directory to dhanushr_harzeena_pavanven_phase_3 in which we have the phase3.py source file run the following command: **“streamlit run phase3.py”**

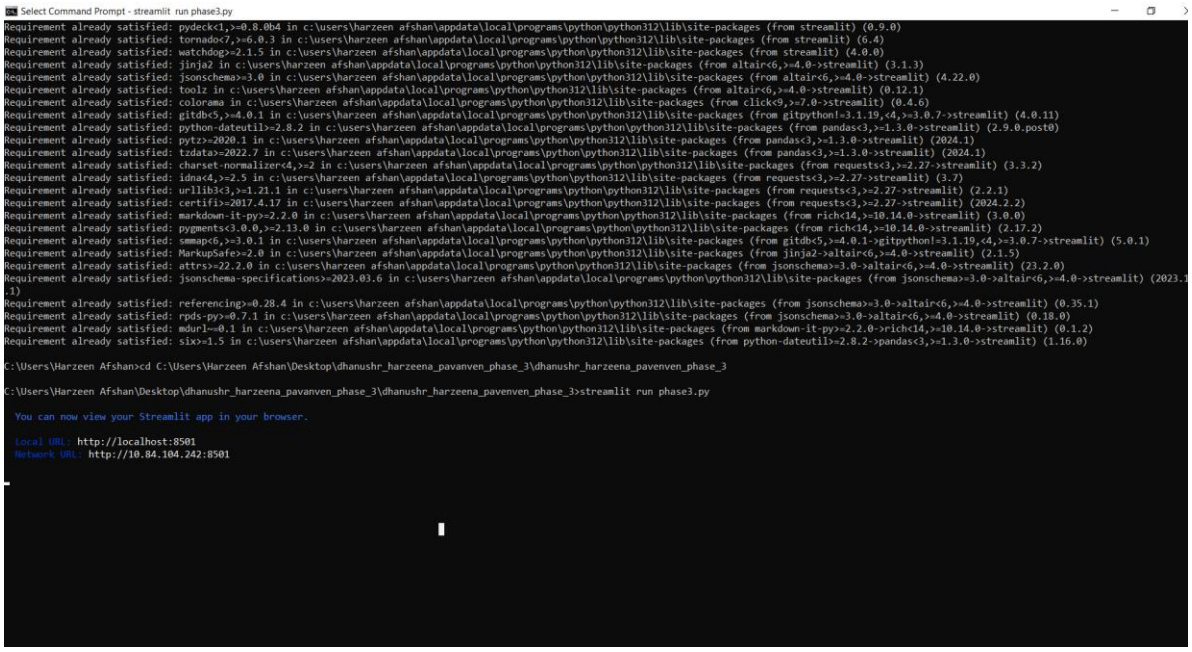


Fig: localhost address

On pressing enter the file will be executed,it redirect to the default web browser where the UI (output) is visualized.

6. This is how the UI looks like that the user to input for various fields and after entering the fields user presses the submit button it predicts the position of the player based on the input

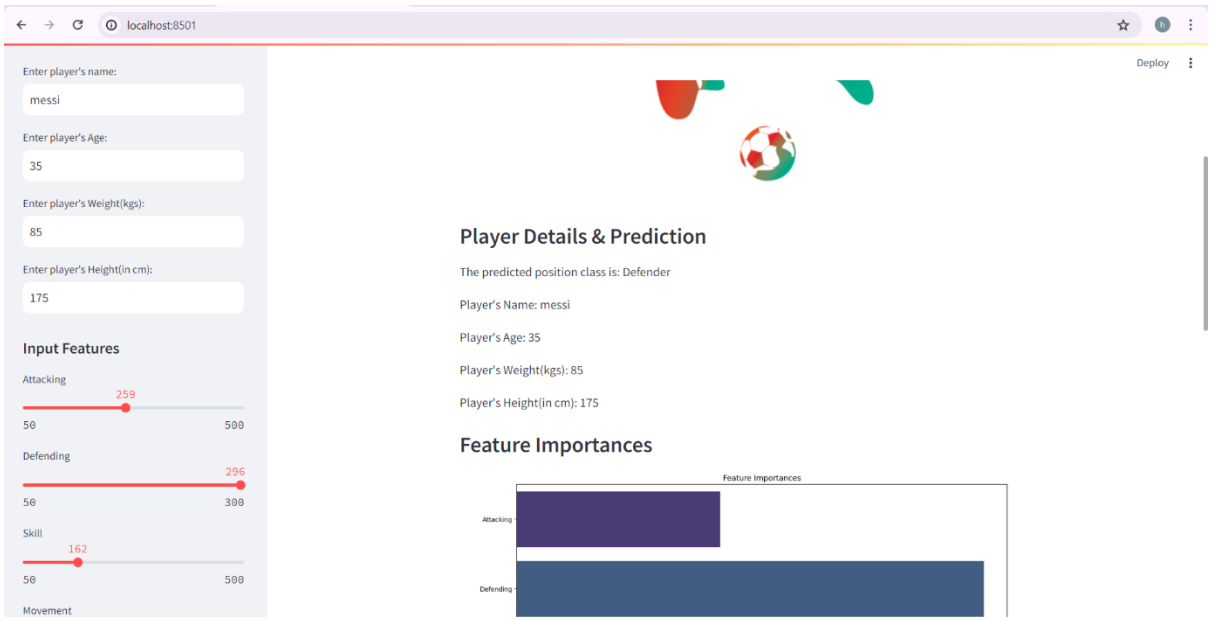


Fig: Screenshot of the UI

Also it shows the feature importances of the given data using bar graph and pie chart visualization.

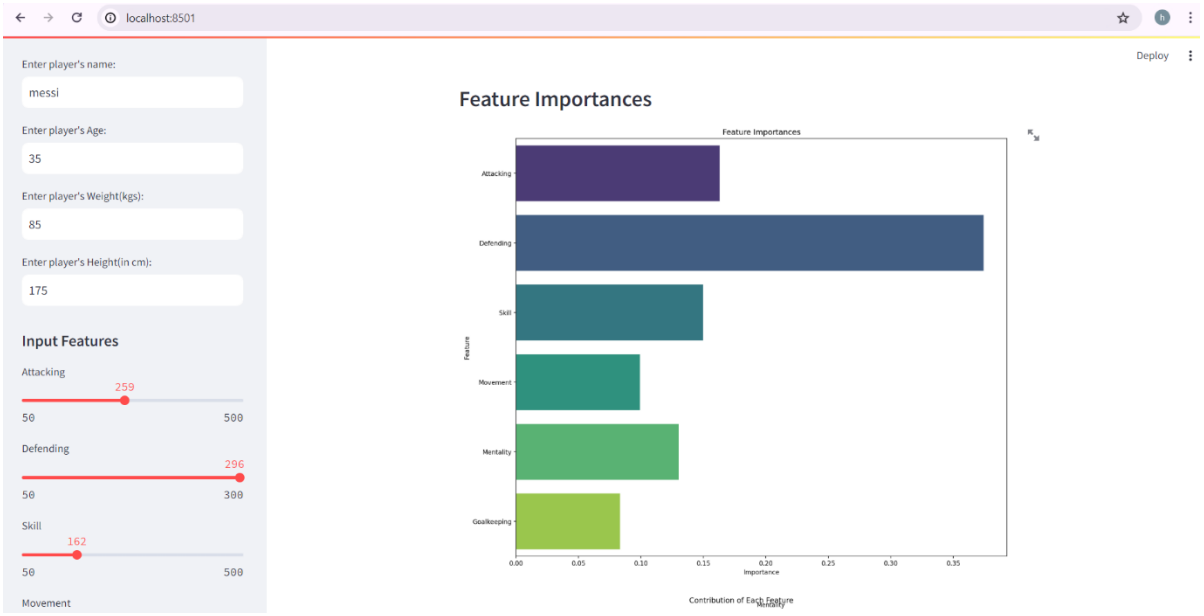


Fig: bar graph visualization of feature importances

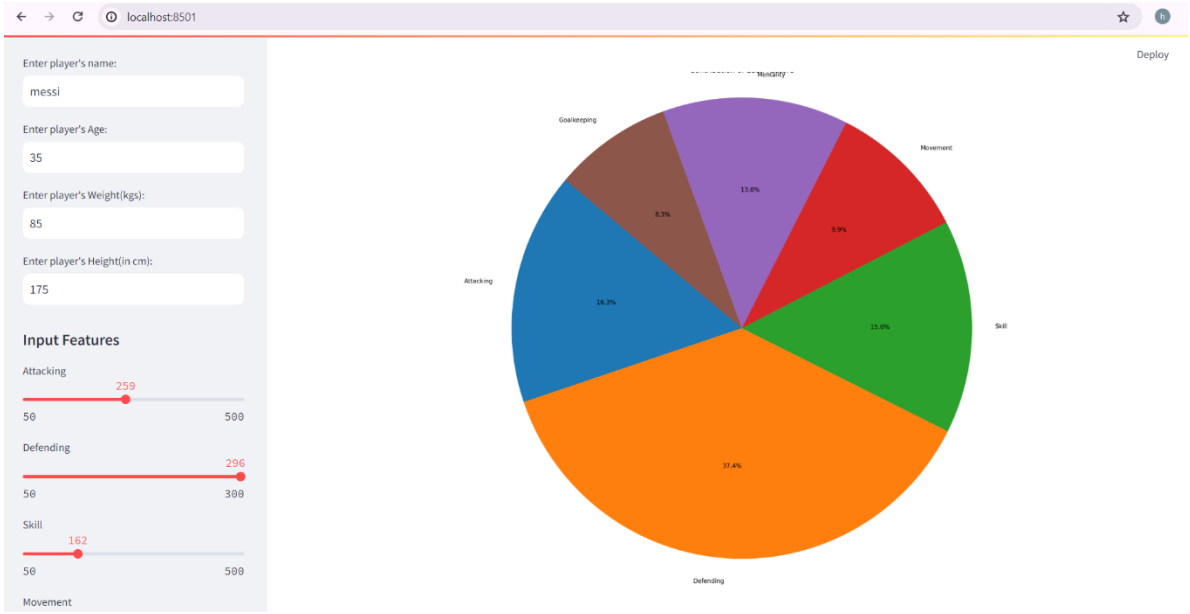


Fig: pic chart visualization of feature importance

1. (b) Model Implemented from Phase 2:

We have implemented 6 algorithms on our data set. The accuracies of each model are listed below:

Logistic Regession: 94.82

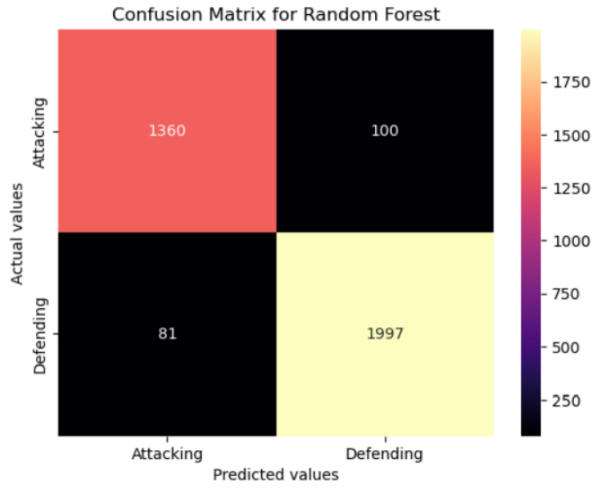
KNN: 95.05 (with K=5)

Naïve Bayes: 73.47

SVM:94.64

Random Forest: 95.30

Decision Tree: 94.64



We conclude from above accuracies and analysis on all the selected algorithms, the best suited algorithm for our model to classify if player was attacker or defender, Random Forest was best suited and least suited was Naïve bayes.

Feature Selection:

Tuning of Data: For this we made a new column in data frame as ‘position class’ which either stores Attacker[if best position is CAM (Center Attacking Midfielder), CF (Center Forward),LW (Left Winger),RW (Right Winger),ST (Striker)] or Defender[if best position is CB (Center Back),CDM (Center Defensive Midfielder),LB (Left Back),RB (Right Back),RWB (Right Wing Back), CM (Center Midfielder), LM (Left Midfielder), RM (Right Midfielder)] based on best position of player.

We selected the feature columns or feature variable as ‘Attacking’, ‘Defending’, ‘Skill’, ’Movement’, ‘Mentality’, ‘Goal Keeping’, and target variable is Position Class Variable.

1.(c) Recommendations related to the problem statement:

What can users infer from our product:

The application classifies players as attackers, defenders, or midfielders by analysing attributes including attacking, defending, skill, movement, mindset, and goalkeeping using random forest classifier techniques. Users can add player information to the prediction, such as name, age, height, and weight, also change the attributes from the side bar. Through visual representations such as feature importance, the tool offers insightful information about the significance of various player traits. The Football Player Position Predictor is a useful tool for coaches, scouts, and football lovers who want to maximize team composition and comprehend the dynamics of player placing. Its intuitive layout makes forecasts easy to use.

How does it help them to solve their problem?

The Football Player Position Predictor Application offers football franchises a valuable tool to streamline talent identification, optimize team composition, and enhance strategic planning. By inputting player attributes, franchises can quickly assess whether a player aligns with their team strategy and positional needs, facilitating efficient scouting and recruitment processes. The predictor's analysis of feature importance provides insights into critical player attributes for specific positions, empowering franchisees to make informed decisions regarding player development, resource allocation, and tactical approaches. Leveraging the predictive capabilities of the tool can provide franchises with a competitive advantage, enabling them to build cohesive teams, allocate resources effectively, and enhance their chances of success on the field.

Ideas on how to extend the product:

We want to keep improving the Football Player Position Predictor so that football teams can use it more and more easily. Our goal is to offer a flexible platform that gives customers the knowledge and skills they need to assess talent and make strategic decisions. Our goal is to make complicated processes easier to understand and enable wise decision-making by providing features that are easy to use, like player comparison tools and real-time player performance updates. Our objective is to develop a solution that can be customized to meet the specific requirements of every franchise, allowing them to optimize productivity and streamline processes. We aim to make the predictor a useful tool for teams in various leagues and play levels, with an emphasis on scalability and usability, ultimately assisting in their success both on and off the field.

Conclusion:

To conclude, we have taken fifa_2021 as our dataset performed data preprocessing, performed EDA, and implemented various 6 algorithms and found that Random forest classifier was a best-fit model to analyze if the player was an attacker, defender, or midfielder for the first two phases and for the third phase we developed an application where user can utilize the application and using random forest classifier the predictions are shown and the detailed explanation how the application was created and how predictions are shown by the model in application are demonstrated in the video and also in the report. The project-related files are attached in a zip file uploaded including phase 1, phase 2 and phase 3.

References:

[1]. Gathering data set from Kaggel: link is given below:
https://www.kaggle.com/datasets/yagunnersya/fifa-21-messy-raw-dataset-for-cleaning-exploring/code
[2]. https://en.wikipedia.org/wiki/Exploratory_data_analysis
[3]. NIST on EDA, https://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm, last viewed February, 2021.
[4]. John Tukey Biography, https://mathshistory.st-andrews.ac.uk/Biographies/Tukey/, last viewed 2021.
[5]. https://scikit-learn.org/stable/
[6]. https://streamlit.io/