SOCCER PLAYER'S SUITABLE PLAYING POSITION PREDICTION USING MACHINE LEARNING

BY

Md. Arman Hosen Patoary
ID: 201-15-3304
AND

Rudra Prashad Shil
ID: 201-15-3338

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. Arif Mahmud
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. Md. Sadekur Rahman
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Soccer Player's Suitable Playing Position Prediction Using Machine Learning", submitted by **Md. Arman Hosen Patoary, ID: 201-15-3304** and **Rudra Prashad Shil, ID: 201-15-3338** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 01-07-2024.

BOARD OF EXAMINERS

Dr. Md. Ismail Jabiullah (MIJ) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Md Assaduzzaman (MA) — **Internal Examiner 1**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Ms. Sharun Akter Khushbu (SAK) — **Internal Examiner 2**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Ahmed Wasif Reza (DWR) — **External Examiner**
Professor
Department of CSE
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Arif Mahmud, Associate Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Arif Mahmud
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Md. Arman Hosen Patoary
ID: 201-15-3304
Department of CSE
Daffodil International University

Rudra Prashad Shil
ID: 201-15-3338
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Dr. Arif Mahmud, Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Accurately identifying the most suitable position for a football player is crucial for optimizing team performance and player development. This paper will investigate the potential of machine learning algorithms in predicting player positions based on various performance metrics. We propose a novel approach that will utilize Logistic Regression, K-Nearest Neighbors, Multi-Layer Perceptron, Random Forest Classifier, Support Vector Machine classifier, and Decision Tree to analyze a comprehensive dataset of player attributes, including physical stats, skill assessments, and positional data. The model will be evaluated on its ability to correctly predict the primary positions of players across different leagues and levels of competition. The results will demonstrate that our proposed approach achieves high accuracy which is in Logistic Regression and that is 75% Accuracy, in predicting a player's suitable playing position. Furthermore, we analyze the feature importance scores to gain insights into the key attributes that are most influential in determining player positions.

Keywords: football player, performance metrics, suitable position, machine learning algorithms, high accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
LR	Logistic Regression
KNN	K-Nearest Neighbors
DT	Decision Tree
RF	Random Forest
SVC	Support Vector Machine
MLP	Multi-Layer Perceptron

CHAPTER 1
Introduction

1.1 Overview

Football, a sport adored by billions, thrives on the intricate interplay of individual talent and strategic positioning. Each player, with their unique blend of physical attributes, technical skills, and tactical understanding, occupies a specific role on the field. Identifying and fostering players suited for the best positions is crucial for team success, often shaping formations, playing styles, and ultimately, outcomes. Traditional scouting methods, while valuable, rely heavily on subjective assessments and limited data. This is where machine learning (ML) presents exciting possibilities.

This paper goes through the application of ML algorithms to predict the ideal playing position for young or up-and-coming football players. We propose a novel approach that leverages a comprehensive dataset encompassing various aspects of player performance, including:

- **Physical attributes:** Height, weight, speed, agility, jumping ability, etc.
- **Technical skills:** Passing accuracy, shooting technique, dribbling ability, tackling success rate, etc.
- **Tactical awareness:** Positioning, decision-making, understanding of formations, etc.

By using this rich data into carefully selected and fine-tuned ML algorithms, we aim to achieve the following objectives:

- **Improve prediction accuracy:** Surpass the limitations of traditional methods by identifying subtle patterns and relationships within the data that may be overlooked by human assessment.
- **Provide data-driven insights:** Offer objective and quantifiable recommendations, empowering scouts and coaches to make informed decisions about player development and positioning.
- **Uncover hidden potential:** Identify players with underutilized skills or playing styles that might flourish in unexpected positions, fostering tactical innovation and flexibility.

This study contributes to the growing body of research exploring the intersection of sports analytics and machine learning. It not only addresses the specific challenge of football player positioning but also paves the way for further advancements in talent identification and player development across various sports. Through thorough analysis and rigorous evaluation, we aim to demonstrate the effectiveness of our proposed approach, opening doors for a more data-driven and objective future in the world of football.

1.2 Background and Present State

Football, often celebrated as the beautiful game, relies significantly on strategic player positioning to optimize team performance and individual contributions. Traditionally, determining suitable playing positions has relied on subjective evaluations by coaches and scouts, based on their experience and intuition. However, advancements in data analytics and machine learning offer new avenues for more accurate and objective assessments.

In recent years, machine learning has gained traction in sports analytics. These models can analyze extensive datasets, uncover complex patterns, and make predictions beyond human capacity. While machine learning has been applied in various football domains such as match outcome prediction and player performance evaluation, its potential for determining optimal playing positions remains largely unexplored.

This study aims to bridge this gap by using machine learning algorithms to predict the best playing positions for footballers. Using a comprehensive dataset of player attributes and performance metrics, the research evaluates several models including Logistic Regression, K-Nearest Neighbors, Multi-Layer Perceptron, Random Forest Classifier, Support Vector Machine, and Decision Tree. By comparing their predictive accuracy, the study seeks to identify the most effective algorithm for precise position prediction, offering a data-driven approach to optimize team strategy and player development.

This research contributes to the evolving field of sports analytics, providing valuable insights for coaches, scouts, and analysts to enhance team performance and player development through informed positional assignments.

1.3 Problem Statement

Historically, player positioning in soccer has relied heavily on coaches' intuition and experience. This paper investigates the potential of machine learning to objectively predict a player's most suitable playing position based on data-driven insights.

Key Hurdles:

- **Data Acquisition:** Gathering comprehensive data encompassing player performance, physical attributes, and tactical movements across diverse positions and skill levels.
- **Feature Construction:** Selecting and transforming relevant data points into features compatible with machine learning algorithms.
- **Model Balancing:** Striking a balance between model accuracy and interpretability to provide valuable insights beyond mere predictions.
- **Positional Specificity:** Developing models capable of accurately predicting positions with varying tactical demands and roles.

1.4 Objectives

This study is driven by the fact that millions of people worldwide have a passion for sports, particularly soccer. So, we take it as an opportunity to contribute to the growing field of sports science and analytics. Finding the "perfect match" for players is critical in soccer's dynamic landscape, impacting individual growth, team success, and talent management.

Here, Machine learning offers a powerful solution:

- By developing predictive models for player positioning, aims to provide valuable insights that can benefit coaches, teams, and players.
- By focusing on soccer player positioning, we can demonstrate the practical utility of machine learning in sports analytics.
- By identifying players who could excel in unexpected positions.
- By identifying the most suitable playing positions for individual players, coaches can tailor training programs and strategies to maximize their effectiveness on the field.
- By building balanced and tactically sound teams by placing players where they shine.

The current practice of assigning player positions in soccer relies heavily on subjective judgment, potentially overlooking talent and hindering individual development. Machine learning offers a data-driven approach to objectively predict a player's ideal position based on performance, attributes, and tactics, unlocking hidden potential, optimizing training, and fostering stronger teams. This research delves into this exciting opportunity to revolutionize player positioning in soccer.

1.5 Scope and Limitations

This project delves into the world of soccer, aiming to create a machine-learning model that pinpoints the ideal playing position for any player based on diverse data points.

Data Gathering and Refining: We'll gather and refine information on player performance, physical attributes, tactical movements, and their current positions. Think of it as building a comprehensive player profile. Match statistics, player tracking data, and expert insights will be our treasure trove.

Feature Engineering: Not all data is created equal. We'll extract the most relevant features, like key performance metrics and movement patterns, and transform them into a format suitable for machine learning algorithms.

Model Training and Evaluation: We'll coach different machine learning models to predict player positions. Accuracy, interpretability, and efficiency are our key Metrics; the best model wins!

Unlocking Insights: We'll analyze the results to understand what makes a great striker or a midfielder. This data-driven approach will reveal key player characteristics for each position.

Prototype Power: Finally, we'll build a prototype system that showcases how these predictions can be used in real-world scenarios. Imagine coaches using data to make informed decisions and unlock hidden potential on the field.

By combining data science and soccer smarts, this project aims to score big, empowering coaches, optimizing player development, and ultimately, making the beautiful game even more enjoyable.

1.6 Report Organization

The proposed report is meticulously organized with a detailed layout designed to lead readers through the research process, findings, and implications. Each chapter fulfills a distinct role, enhancing the document's overall coherence and depth.

Chapter 1: Introduction

The research methodology chapter describes the approach used to carry out the study. It offers details on the research design, data collection techniques, model architecture, and the reasoning for selecting particular methods. Additionally, the chapter addresses ethical considerations and acknowledges any limitations that could affect the study.

Chapter 2: Literature Review

The background chapter presents a comprehensive review of relevant literature and previous research. This section provides an in-depth understanding of the theoretical foundations, methodologies, and technological advancements in position prediction using machine learning algorithms. It sets the stage for the current study and identifies gaps in existing knowledge.

Chapter 3: Methodology

The research methodology chapter details the approach used to carry out the study. It explains the research design, data collection methods, model architecture, and the reasons for selecting specific methodologies. Additionally, the chapter addresses ethical considerations and acknowledges any limitations that could affect the study.

Chapter 4: Implementation

This chapter showcases the experimental model train, prototype design and system testing derived from machine learning algorithms applied to player position prediction. It includes detailed analyses of the performance of data preprocessing, cleaning and training and testing.

Chapter 5: Result and Analysis

This chapter showcases the experimental results derived from machine learning algorithms applied to player position prediction. It includes detailed analyses of the performance of various detection and segmentation methods, complemented by visual representations and statistical measures to validate the findings.

Chapter 6: Impact on Society, Environment and Sustainability

This chapter eradicates using machine learning to predict soccer players' positions impacts the sport and society by optimizing team dynamics, promoting inclusivity, and recognizing talent regardless of socioeconomic background. It encourages STEM interest among young fans, linking sports with education. Economically, it boosts revenues from ticket sales, merchandise, and broadcasts, benefiting local communities and driving economic growth.

Chapter 7: Conclusion and Future work

This concluding chapter consolidates, summarizing the main uncovering, restating the final statements derived by analysis, and presenting guidance. It provides a guide for researchers who wish to build on this present study.

1.7 Summary

This project focuses on creating a Soccer Player's Suitable Playing Position Prediction system through Machine Learning algorithms such as Logistic Regression (LR), k-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVC), Decision Tree (DT), Multi-Layer Perceptron (MLP). The aim is to furnish coaches and managers with a dependable tool for predicting the best playing positions by considering player skills and historical data. The system, trained on an extensive dataset, will feature a user-friendly interface for generating position recommendations, with the potential to transform player selection processes and elevate team performance in soccer.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

In soccer, determining playing positions has traditionally depended on the subjective judgment and experience of coaches and scouts. This involves assessing a player's physical attributes, technical skills, and in-game performance to identify their ideal role on the field. Each position, whether goalkeeper, defender, midfielder, or forward, requires specific skills and physical traits. For instance, defenders need strength and tackling ability, while forwards rely on speed and finishing skills. Despite the valuable insights from experienced coaches, traditional methods have limitations, including subjectivity, potential bias, and limited ability to process large datasets comprehensively.

The adoption of technology and data analytics has revolutionized various aspects of soccer, from performance analysis to tactical planning. Modern soccer clubs gather extensive data on players, including physical metrics like speed and endurance, technical skills such as passing accuracy and dribbling success, and in-game actions like distance covered and tackles made. This wealth of data offers an opportunity to use machine learning (ML) techniques to objectively assess and predict the most suitable playing positions for players.

Machine learning, a subset of artificial intelligence, involves algorithms that learn from and make predictions based on data. In sports, ML has been used to forecast player performance, identify injury risks, and analyze tactical patterns. For position prediction, the process includes data collection, feature selection, model training, and validation. By analyzing relevant player attributes and performance metrics, ML models can identify patterns and associations that inform optimal position assignments.

Applying ML to predict suitable playing positions offers several advantages: it provides objectivity by reducing subjective bias, scalability in handling large datasets, and consistency in applying criteria across all players. This approach promises to enhance traditional methods, leading to more effective team formations and improved player development. This research paper explores the methodologies, results, and implications of using ML for soccer player position prediction.

2.2 Related Works

Several research projects share a common objective: the development of innovative systems integrating Machine Learning (ML) and data analytics for predicting soccer player positions, with the overarching goal of enhancing team strategies and overall performance. Utilizing a diverse set of ML algorithms, including LR, KNN, DT, RF, SVC, MLP, as well as advanced techniques like cluster analysis and graph embeddings, these initiatives aim to provide coaches, managers, and scouts with reliable tools for predicting optimal playing positions based on players' skills, physical attributes, and historical performance data. These projects conduct comprehensive training of their systems on extensive datasets, encompassing player statistics, playing styles, positions played, and match performance. The envisioned outcomes include user-friendly interfaces enabling input of player data to generate position recommendations, with the ultimate goal of revolutionizing player selection processes, optimizing team formations, and elevating overall team performance in soccer.

Zixue Zeng and Bingyu Pan[1] conducted a thorough examination into soccer player position prediction, employing the BP Neural Network model. Their study underscores the capability of machine learning in scrutinizing performance data to optimize player placement. By leveraging a diverse dataset encompassing player statistics, playing styles, and historical match performance, the model achieves a 73% accuracy, extendable to 77%.

The approach by Okan Dağ, Asım Sinan Yüksel, and Şerafettin Atmaca[2] employs cluster analysis through the Expectation Maximization (EM) algorithm to categorize soccer player positions based on various attributes. With an 81% success rate, their focus is on understanding inherent variations in player positions for strategic decision-making and potential impacts on team performance.

Öznur İlayda Yılmaz and Şule Gündüz Öğüdücü[3] introduce an approach using k-Nearest Neighbors, XGBoost, and ANN algorithms to enhance a football player recommendation system with a remarkable 97.92% accuracy. Their innovative method aims to improve the precision of player recommendations for coaches, team managers, and scouts during the player selection process.

Mahdi Nouraiea and Changiz Eslahchi[4] adopt a data-driven machine learning approach, employing SVC, Logistic Regression, Random Forest, and Deep Neural Network to enhance player positioning in soccer. With a high success rate of

99.84%, their focus is on strategic decision-making for player placement based on advanced data analytics.

S. Bosu Babu, Vavilapalli Vivek, Dasari Mahendra Tulasi Kumar, Korra Prathyusha, and Galla Pavan Teja[5] contribute by developing a predictive model using statistical analysis and ML algorithms, achieving a 91.9% accuracy in Random Forest. Their work provides valuable insights for strategic decision-making related to football player placement.

Prof. Annamaria Guolo and Alberto Gobbo's[6] approach integrates various data mining and ML techniques such as Decision Tree, KNN, Naive Bayes, Support Vector Machines, Neural Networks, Ensemble Methods, and Logistic Regression, achieving an 86.8% accuracy in Logistic Regression. Their predictive model aims to offer valuable insights for coaches, team managers, and scouts, enhancing strategic decision-making.

Research findings suggest that machine learning algorithms exhibit promise in capturing intricate relationships between players' attributes and their ideal positions during a game. Logistic Regression, valued for its simplicity and interpretability, has been widely implemented. The literature underscores the crucial role of feature engineering in enhancing model performance, highlighting the significance of physical attributes, playing statistics, and historical performance. Additionally, the review emphasizes the use of cross-validation techniques for robust model evaluation, ensuring the adaptability of algorithms to new data. Numerous studies underscore the importance of real-world testing and validation to evaluate the practicality and effectiveness of developed models. Some research highlights the success of individual algorithms, while others explore ensemble methods to synergize multiple algorithms, ultimately enhancing prediction outcomes.

In summary, the literature review indicates a consensus regarding the effectiveness of machine learning algorithms in predicting suitable playing positions for soccer players. It accentuates the importance of algorithm selection, feature engineering, and robust evaluation methodologies. The integration of various algorithms and their collective potential to enhance prediction accuracy emerges as a prominent theme in the existing body of research.

2.3 Comparison Between Existing Works

In the paper "A Machine Learning Model to Predict Player's Positions based on Performance" by Zixue Zeng and Bingyu Pan they used only BP Neural Network and got 73% Accuracy where we used Logistic Regression, KNN Classifier, MLP, SVC, Decision Tree, Random Forest etc. algorithm and got best accuracy in Logistic Regression which is 75%. So, we can say that we have got better results from them. On the other hand in another paper named "Examination of Player Positions by Cluster Analysis", authors Okan Dağ, Asım Sinan Yüksel, and Şerafettin Atmaca got 81% accuracy using Expectation Maximization which is better than our accuracy. Furthermore, in the paper named "Learning Football Player Features using Graph Embeddings for Player Recommendation System" written by Öznur İlayda Yılmaz and Şule Gündüz Öğüdücü, In that paper they used k-Nearest Neighbors, XGBoost, and an Multi-Layer Perceptron and get best accuracy in XGBoost which is 97.92% which is greater than us. Moreover, the paper "Positioning Soccer Players for Success: A Data-Driven Machine Learning Approach" by Mahdi Nouraiea and Changiz Eslahch used SVC, Logistic regression, Random Forest, Deep Neural network, and got the best accuracy in Logistic regression which is 99.84%. The paper "Predicting Football Player's Position" by S Bosu Babu, Vavilapalli Vivek, Dasari Mahendra Tulasi Kumar, Korra Prathyusha, Galla Pavan Tejaa used Random Forest, KNN, Naive Bayes, Decision tree, SVC and get best accuracy in Random Forest which is 91.9% and the paper "Prediction of football players' position using Data Mining and Machine Learning" by Prof. Annamaria Guolo, Alberto Gobbo used Decision tree, KNN, Naive Bayes, Support vector machines, Neural networks, Ensemble Methods, Logistic regression. Among those algorithms, they got the best accuracy in Logistic regression which is 86.8%. At last, we can say that they have better accuracy than us but their data set is different than ours and they have used only 3 or 4-position prediction which is why they are getting better results.

TABLE 2.3.1: Comparative Analysis Table

SL.	Author Name	Used Algorithm	Best Accuracy
1.	Zixue Zeng and Bingyu Pan	BP Neural Network	BP Neural Network, =73%
2.	Okan Dağ, Asım Sinan Yüksel, and Şerafettin Atmaca	Expectation Maximization	Expectation Maximization = 81%

3.	Öznur İlayda Yılmaz and Şule Gündüz Öğüdücü	K-Nearest Neighbors, XG-Boost, and artificial neural network.	XG-Boost =97.92%
4.	Mahdi Nouraiea and Changiz Eslahchi	SVC, Logistic regression, Random Forest, Deep Neural network	Logistic regression=99.84%
5.	S Bosu Babu, Vavilapalli Vivek, Dasari Mahendra Tulasi Kumar, Korra Prathyusha, Galla Pavan Tej	Random Forest, KNN, Naive Bayes, Decision tree, SVC	Random Forest=91.9%
6.	Prof. Annamaria Guolo, Alberto Gobbo	Decision trees, KNN, Naive Bayes, Support vector machines, Neural networks, Ensemble Methods, Logistic regression	Logistic regression =86.8%

2.4 Open Issues

Predicting the most suitable playing position for soccer players using machine learning algorithms faces several challenges despite technological advancements. A significant issue is the quality and availability of data. Although there is a wealth of historical match data and player statistics, they often lack detailed context such as specific game situations and player interactions, which are essential for precise position prediction. Additionally, the dynamic and complex nature of soccer, where a player's role can change based on tactics, opponent strategies, and game events, adds to the modeling difficulty.

The performance of various machine learning algorithms, including Logistic Regression, KNN Classifier, MLP, Random Forest Classifier, SVC, and Decision Tree, varies depending on the data and application. Each algorithm has its advantages and limitations; for example, while MLP can identify intricate patterns, they require significant computational power and large datasets. Simpler models like Logistic Regression and KNN might not effectively capture the nonlinear relationships in soccer data.

Another major challenge is model interpretability. Coaches and analysts need to understand the rationale behind a model's position recommendation, which is often difficult with complex algorithms like MLP and SVC. The lack of transparency can limit the practical use of these advanced techniques. Moreover, there is a need for continuous model updating and validation to reflect player development, team strategy changes, and evolving playing styles, which increases complexity and resource demands.

Lastly, the integration of domain knowledge from coaches and players into machine learning models is underutilized. Human expertise can offer insights that purely data-driven methods may overlook. Addressing these issues requires a multidisciplinary approach, combining machine learning advancements with domain-specific knowledge and robust data collection practices.

2.5 Summary

Research findings suggest that machine learning algorithms exhibit promise in capturing intricate relationships between players' attributes and their ideal positions during a game. Logistic Regression, valued for its simplicity and interpretability, has been widely implemented. The literature underscores the crucial role of feature engineering in enhancing model performance, highlighting the significance of physical attributes, playing statistics, and historical performance. Additionally, the review emphasizes the use of cross-validation techniques for robust model evaluation, ensuring the adaptability of algorithms to new data. Numerous studies underscore the importance of real-world testing and validation to evaluate the practicality and effectiveness of developed models. Some research highlights the success of individual algorithms, while others explore ensemble methods to synergize multiple algorithms, ultimately enhancing prediction outcomes.

In summary, the literature review indicates a consensus regarding the effectiveness of machine learning algorithms in predicting suitable playing positions for soccer players. It accentuates the importance of algorithm selection, feature engineering, and robust evaluation methodologies. The integration of various algorithms and their collective potential to enhance prediction accuracy emerges as a prominent theme in the existing body of research.

CHAPTER 3
METHODOLOGY

3.1 Overview

This project focuses on predicting soccer players' most suitable playing positions using machine learning techniques. The employed algorithms encompass Logistic Regression, KNN Classifier, Multi-Layer Perceptron (MLP), Random Forest Classifier, Support Vector Machine (SVC), and Decision Tree.

The primary objective of this research is to optimize player placement on the field, considering their skills, physical attributes, and performance data. Machine learning algorithms analyze and identify patterns within a dataset containing diverse soccer players' attributes and historical performance metrics.

Logistic Regression, a statistical model, predicts the probability of a player belonging to a specific playing position. KNN Classifier identifies positions based on the similarity of players' characteristics. MLP, a complex neural network model, captures intricate relationships between player attributes and suitability for different positions. Random Forest Classifier, a robust ensemble method, enhances predictive accuracy by combining outputs from multiple decision trees.

SVC is applied for classification tasks, focuses on finding an optimal hyperplane to separate data points. Decision Tree, a tree-like model, makes decisions based on features like player height, weight, speed, and skill set.

By employing these diverse machine learning algorithms, the study aims to provide a comprehensive and accurate prediction of a soccer player's optimal playing position. The results could potentially assist coaches, managers, and scouts in making informed decisions about player positioning, ultimately enhancing team performance. The research involves training and testing these models on a dataset with information about various soccer players and their playing positions, aiming to find the most reliable and accurate prediction algorithm.

3.2 Proposed Methodology

This study aims to predict the optimal playing position for soccer players using a variety of machine learning algorithms: Logistic Regression, K-Nearest Neighbors (KNN) Classifier, Multi-Layer Perceptron (MLP), Random Forest Classifier, Support Vector Machine (SVC), and Decision Tree. The methodology encompasses the following steps:

Data Collection:

- Data Sources: Collecting player data from reliable sources like Kaggle, FIFA databases, soccer analytics platforms, and historical match records etc.
- Data Attributes: Gathering attributes including physical stats (height, weight, speed), technical skills (passing accuracy, dribbling, shooting), tactical knowledge (positioning, vision), and match performance metrics (goals, assists, tackles).

Data Preprocessing:

- Data Cleaning: Addressing missing values, removing duplicates, and correcting inconsistencies.
- Normalization/Standardization: Normalizing or standardizing the data to ensure all features contribute equally to the model.

Data Splitting:

- Training and Testing Sets: Dividing the dataset into training (70%) and testing (30%) subsets.

Model Training: Training the following classification models to classify player positions based on the collected features:

- Logistic Regression: Simple and efficient for binary classification.
- Decision Tree: Models decision rules based on feature values to classify positions.
- K-Nearest Neighbors (KNN) Classifier: Classifies positions based on the proximity to k-nearest players in the feature space.
- Random Forest Classifier: Uses ensemble learning to enhance classification accuracy.
- Support Vector Machine (SVC): Finds the optimal hyperplane that maximizes the margin between classes.
- Multi-Layer Perceptron (MLP): Captures non-linear relationships between features and player positions.

Model Evaluation:

- Confusion Matrix (CM): Illustrating the model's performance using TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives).
- Accuracy: The percentage of correctly classified positions.

  Accuracy = (TP+TN)/(TP+TN+FP+FN)

- Precision: The ratio of correctly predicted positive observations to the total predicted positives.

  Precision = TP/(TP+FP)

- Recall: The ratio of correctly predicted positive observations to all observations in the actual class.

  Recall = TP/(TP+FN)

- F1-Score: The weighted average of Precision and Recall.

  F-1 Score = 2*(Precision*Recall)/(Precision+Recall)

- Specificity: The proportion of actual negatives correctly identified.

  Specificity = TN∕((TN+FP))

Model Comparison and Selection:

- Performance Comparison: Comparing accuracy, precision, recall, F1-score, and specificity across models.
- Feature Importance: Assessing feature importance for tree-based models to determine which features most influence position prediction.

Model Deployment:

- Final Model Selection: Selecting the best-performing model for real-time position prediction.

- User Interface: Developing an interface for coaches and analysts to easily utilize the model.

Model Improvement and Maintenance:

- Feedback Loop: Collecting user feedback to continuously improve the model.
- Periodic Retraining: Regularly retraining the model with new data to maintain its accuracy and relevance.

3.3 Hardware/ Software Requirement

The research on "Soccer Player’s Suitable Playing Position Prediction using Machine Learning" requires specific elements to be implemented successfully.

Hardware Requirements:

- Availability of computational infrastructure with adequate processing power and memory is necessary to manage the tricky tasks elaborate in system.

- The use of a GPU is vital for expediting the training process of machine learning algorithms, thereby considerably decreasing valuable time.

Software Requirements:

- Machine Learning Platform and Libraries: Deployment of widely-recognized machine learning frameworks and importing necessary libraries for our models.

- High Level Programming language: Python provides a versatile platform for Executing machine learning algorithms and performing data analysis.

3.4 Project Management and Financial Analysis

Project Management: We have planned how the project will be managed and how the flow will be maintained. Here:

Project Timeline:

TABLE 3.4.1: Project Timeline

Tasks	Weeks																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Identifying Research Question																		
Literature Review																		
Data Collection and Analysis																		
Interpretation of Results																		

Estimated Work Period	
Actual Work Period	

Resources Utilized:

A. Equipment and Tools:

- Development and Testing Servers
- Adobe Creative Suite
- Collaboration Tools (Slack)
- Git
- Selenium

B. Software and Technologies:

- HTML5, CSS3, Python
- Django, Python
- Database Management System (MySQL)

C. Data and Content:

- Product Data: Provided by Kaggle, GitHub, FIFA dataset by EA Sports.
- User Documentation: The technical writing team produced it.

Risk Management:

SWOT analysis is a framework for examining a project's positive and negative factors, both internal and external. Here's a SWOT analysis for the Web-Application platform.

Figure 3.4.1: SWOT analysis

Financial Analysis: Estimated Cost for this project –

TABLE 3.4.2: Financial Analysis

SN	Components	Estimated Cost (BDT)
01.	Hardware/Infrastructure	4500
02.	Domain and Hosting	2000
03.	Visiting Stakeholders	1000
04.	Software and Tools	3000
05.	Data Collection and Processing	2700
06.	Documentation and Report Writing	700
07.	Miscellaneous	1000
08.	Contingency (10% of total)	800
	Total Estimated Cost	15,700

3.5 Summary

Forecasting appropriate playing positions for soccer players through machine learning incorporates the utilization of diverse algorithms, which encompass Logistic Regression, KNN Classifier, Multi-Layer Perceptron (MLP), Random Forest Classifier, Support Vector Machine (SVC), and Decision Tree. The procedure initiates with the acquisition of data, where pertinent details about soccer players, such as physical attributes, playing statistics, and historical performance, are compiled. Subsequently, the dataset undergoes preprocessing to address missing values, standardize features, and ensure its suitability for machine learning algorithms.

Following this, the dataset is divided into training and testing subsets to accurately assess the models' performance. The chosen machine learning algorithms are then trained on the training set utilizing labeled data, specifying the player's playing position. Evaluation of the models' performance occurs using the testing set to ensure precision and dependability.

Furthermore, feature selection techniques may be employed to recognize the most influential factors in predicting playing positions. Cross-validation is utilized for additional model validation and to prevent overfitting. The appropriateness of each algorithm is compared based on metrics like accuracy, precision, recall, and F1-score. Ultimately, the most effective algorithm is selected as the preferred model for forecasting soccer players' suitable playing positions in practical scenarios.

CHAPTER 4

IMPLEMENTATION

4.1 Overview

This project focuses on predicting soccer players' most suitable playing positions using machine learning techniques. The employed algorithms encompass Logistic Regression, KNN Classifier, Multi-Layer Perceptron (MLP), Random Forest Classifier, Support Vector Machine (SVC), and Decision Tree.

The primary objective of this research is to optimize player placement on the field, considering their skills, physical attributes, and performance data. Machine learning algorithms analyze and identify patterns within a dataset containing diverse soccer players' attributes and historical performance metrics.

Logistic Regression, a statistical model, predicts the probability of a player belonging to a specific playing position. KNN Classifier identifies positions based on the similarity of players' characteristics. MLP, a complex neural network model, captures intricate relationships between player attributes and suitability for different positions. Random Forest Classifier, a robust ensemble method, enhances predictive accuracy by combining outputs from multiple decision trees.

SVC, a classification model focuses on finding an optimal hyperplane to separate data points. Decision Tree, a tree-like model, makes decisions based on features like player height, weight, speed, and skill set.

By employing these diverse machine learning algorithms, the study aims to provide a comprehensive and accurate prediction of a soccer player's optimal playing position. The results could potentially assist coaches, managers, and scouts in making informed decisions about player positioning, ultimately enhancing team performance. The research involves training and testing these models on a dataset with information about various soccer players and their playing positions, aiming to find the most reliable and accurate prediction algorithm.

Instrumentation:

Initially, a robust computer system equipped with a GPU and other necessary tools is required for implementing machine learning models. The hardware specifications necessary for this model are detailed below. The Intel Core i5 8th generation CPU incorporated into this system's configuration ensures powerful processing capabilities, facilitating effective computational tasks. With a 1 TB hard drive, there is ample storage capacity for data, allowing for efficient storage and retrieval. The system has 8 gigabytes of RAM, providing sufficient capacity for multitasking and the smooth execution of programs.

Utilization of Libraries:

Imported libraries play a crucial role in various tasks such as data manipulation, visualization, and the construction and assessment of machine learning models.

Processing General Data:

- Numpy
- Pandas

Data Pre-processing:

- Standard Scaler

Visualizing Data:

- Matplotlib
- Seaborn

Libraries for Machine Learning Models:

- Scikit-learn (sklearn)

Machine Learning Models:

- KNN
- Logistic Regression
- Decision Tree
- Support Vector Machine classifier (SVC)
- MLP
- Random Forest etc.

4.2 Train Model / Prototype Design

Datasets: The study encompasses three distinct goals. Initially, we need to amass the dataset, sourced from Kaggle, GitHub, and FIFA datasets provided by EA Sports. Subsequently, we are required to preprocess the dataset, eliminating duplicate, null, or missing data, and then train the machine using 80% of the overall data. This training aims to equip the machine with adequate knowledge to generate accurate outputs.

We will evaluate the machine's performance by employing machine learning algorithms, including Logistic Regression, KNN Classifier, MLP, Random Forest Classifier, SVC, and Decision Tree. This testing phase is crucial for gauging accuracy and obtaining precise outputs.

Here the dataset link_ fifa.csv

And Here is our Google Colab CODE

TABLE 4.2.1: Raw Data Used in the Train, Test and Validation Sets.

	Train	**Test**
Raw Dataset	12977	5561

For each playing position, the dataset initially includes all available players. Subsequently, a random sample, equal in size to the number of players occupying that specific position, is drawn from the pool of players not playing in that position. This random sample is then combined with players playing in the target position, introducing a new column labeled "label." A value of 1 is assigned to players in the position of interest, while a value of 0 is assigned to all other players. This process sets the stage for constructing a predictive model for each position. A 70%-30% split is employed, allocating 70% of the samples for model training and the remaining 30% for testing.

The study will explore the application of the Random Forest ensemble learning algorithm and the Support Vector Machine (SVC) classification algorithm. Random forest combines multiple decision trees to enhance prediction accuracy, while SVC aims to find the optimal hyperplane for class separation. The results section will subsequently compare outcomes across Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest models for various positions in the field.

TABLE 4.2.2: Table of Criteria

CRITERIA	
Overall Rating	Free kick Accuracy
Potential Rating	Ball Control
Pace Total	Acceleration
Shooting Total	Reaction Time
Passing Total	Balance
Dribbling Total	Shot Power
Defending Total	Stamina
Physicality Total	Vision
Crossing	Penalties

Finishing	Marking

The dataset encompasses comprehensive information on adult male soccer players globally, including their performance scores in various soccer skills (e.g., shooting, passing, ball control, heading) and contract details, totaling 89 attributes. Given the study's focus on evaluating players based on their physical and technical attributes, considerations were limited to the players' technical skill, age, height, and weight, accounting for 72 out of the 89 variables. The remaining variables were disregarded. The dataset, available on the Kaggle, GitHub, FIFA dataset by EA Sports spans from all season. Notably, additional valuable variables introduced since 2016 were incorporated into the analysis, prompting the inclusion of data from the 2016-2017 season through the 2022-2023 season. Each player in the dataset is assigned a club position for each season, indicating the position where they most frequently played.

TABLE 4.2.3: Table of Player Positions

Goalkeeper	**Defensive**	**Midfielder**	**Forward**
GK	CDM	LM	ST
	LWB	RM	LW
	RWB	CM	RW
	LB		CF
	CB		CAM
	RB		

The dataset comprises 16 positions denoted by abbreviations such as GK (Goalkeeper), CDM (Center Defensive Midfielder), LWB (Left Wing Back), RWB (Right Wing Back), LB (Left Back), CB (Center Back), RB (Right Back), LM (Left Midfielder), RM (Right Midfielder), CM (Center Midfielder), ST (Striker), CF (Center Forward), LW (Left Wing), RW (Right Wing), CAM (Center Attacking Midfielder), SUB(Substitutes)etc. as shown on Table 1.1. Players without assigned positions and substitutes will be excluded. After removing these positions from the analysis, a total of 15 standard positions were identified those are CDM, LM, ST, LWB, RM, LW, RWB, CM, LB, CF, CB, RB, RW, CAM.

Refer to Table 1.2 for the names and concise descriptions of the dataset variables utilized in this investigation. Table 1.2 illustrates the presence of a comprehensive variable in the dataset, which allocates players a rating out of 100 according to their appropriateness for their respective positions. It will play a role in the feature selection stage.

Process of Experiments: The objective of this experiment is to determine the optimal playing position for soccer players using various machine learning algorithms, including LR, KNN, MLP, RF, SVC, and DT. The procedure involves several key stages: data collection, data preprocessing, model training, classification, and result evaluation.

Data Preprocessing: The collected data undergoes preprocessing to make it suitable for training machine learning models:

- **Data Cleaning:** Addressing missing values, correcting inconsistencies, and eliminating irrelevant features.
- **Feature Selection:** Identifying and selecting significant features that contribute to predicting the playing position.
- **Normalization:** Scaling numerical features to a standard range using techniques like Min-Max Scaling or Standardization.
- **Encoding Categorical Variables:** Transforming categorical features into numerical values using methods like one-hot encoding.

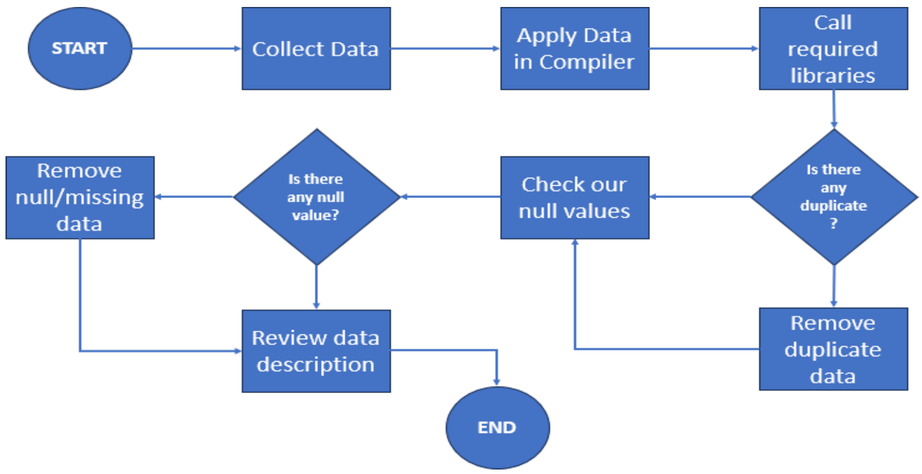

Figure 4.2.1: Flowchart of Data Pre-Processing

Model Training: Various machine learning algorithms are implemented and trained on the preprocessed dataset:

- **Logistic Regression:** A linear model that predicts class probabilities.

- **KNN Classifier:** A non-parametric method that classifies based on the majority vote of k-nearest neighbors.
- **Multi-Layer Perceptron (MLP):** A model inspired by biological neural networks, consisting of input, hidden, and output layers.
- **Random Forest Classifier:** An ensemble method that combines multiple decision trees to improve prediction accuracy.
- **Support Vector Machine (SVC):** A classifier that seeks to create the widest possible gap between categories using a hyperplane.
- **Decision Tree:** A model that recursively divides data into smaller subsets, creating a hierarchical tree-shaped representation based on feature characteristics.

Classification:

Each model trained on the training dataset and then used to predict the playing position for each player in the testing dataset. The classification process includes:

- Splitting the dataset into training and testing sets (typically 70% training and 30% testing).
- Training each model using the training set.
- Applying the model to a holdout set to assess its performance.

Results:

The performance of each model is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. Accuracy is the primary metric used for comparison.

	Name	Score
0	LR	0.746674
1	DT	0.599065
2	KNN	0.653722
3	RF	0.700827
4	SVC	0.727077
5	ANN	0.744696

Figure 4.2.2: Trained Machine Learning Models and their Accuracy

After training and evaluating all the models, their accuracy results are compared. The accuracies obtained for each model are as follows:

- **Logistic Regression:** 75%
- **Decision Tree Classifier:** 60%
- **KNN Classifier:** 65%
- **Random Forest Classifier:** 70%
- **Support Vector Machine Classifier (SVC):** 73%
- **Multi-Layer Perceptron (MLP):** 74%

Logistic Regression achieved the highest accuracy of 75%, indicating it is the most effective algorithm for predicting soccer players' playing positions in this experiment. This study implemented multiple machine learning algorithms to predict the optimal playing position for soccer players. Logistic Regression outperformed the other algorithms with an accuracy of 75%, suggesting it is well-suited for this classification problem. Future research could focus on further tuning the models, exploring additional features, and employing more advanced algorithms to enhance prediction accuracy.

4.3 Model Evaluation

- Confusion Matrix (CM): Illustrates the model's performance using TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives).

- Accuracy: The percentage of correctly classified positions.

  Accuracy = (TP+TN)/(TP+TN+FP+FN)

- Precision: The ratio of correctly predicted positive observations to the total predicted positives.

  Precision = TP/(TP+FP)

- Recall: The ratio of correctly predicted positive observations to all observations in the actual class.

  Recall = TP/(TP+FN)

- F1-Score: The weighted average of Precision and Recall.

  F-1 Score = 2*(Precision*Recall)/(Precision+Recall)

- Specificity:

  Specificity = TN∕((TN+FP))

4.4 Summary

The study "Soccer Player's Suitable Playing Position Prediction using Machine Learning Algorithms" investigated the use of various machine learning methods to predict the best playing positions for soccer players based on their performance data. The algorithms evaluated included Logistic Regression, K-Nearest Neighbors (KNN) Classifier, Artificial Neural Networks (ANN), Random Forest Classifier, Gaussian Naive Bayes (Gaussian-NB), Support Vector Machine (SVM), and Decision Tree. Each algorithm was applied to a dataset of player statistics, incorporating features like passing accuracy, speed, and stamina.

Logistic Regression was chosen for its straightforward binary classification capabilities but was limited in capturing complex data relationships. The KNN Classifier, appreciated for its simplicity, based predictions on the similarity of player features but had difficulties with high-dimensional data. ANN, capable of modeling nonlinear patterns through multiple layers, showed potential but required significant computational power and fine-tuning. The Random Forest Classifier, an ensemble approach, managed dataset variance well and avoided overfitting, providing reliable predictions. Gaussian-NB, which assumes feature independence, used a probabilistic model but was less effective due to the interdependence of player attributes. SVM, aiming to maximize class separation margins, performed well with appropriate kernel selection and regularization but was computationally demanding. The Decision Tree algorithm, offering clear and interpretable rules for prediction, risked overfitting without sufficient pruning.

The study found that ensemble methods like Random Forest and advanced techniques such as ANN and SVM provided the most accurate predictions. The results highlight the necessity of choosing suitable algorithms based on dataset characteristics and computational considerations for predicting soccer players' optimal playing positions.

CHAPTER 5

RESULT AND ANALYSIS

5.1 Overview

The experimental setup for the research titled "Soccer Player's Suitable Playing Position Prediction using Machine Learning Algorithms" is meticulously designed to leverage a combination of hardware, software, and data resources. Utilizing high-performance hardware, including a robust CPU and a powerful GPU, forms the foundation for training and evaluating various machine learning models. Implemented within a Python programming environment, the setup leverages machine learning libraries such as Scikit-Learn, TensorFlow, and PyTorch. These frameworks enable the development, training, and evaluation of models like Logistic Regression, K-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), Random Forest Classifier, Support Vector Machine (SVC), and Decision Tree. The dataset, sourced from a comprehensive soccer database, includes a wide range of player attributes, performance metrics, and positional information. To ensure the dataset is consistent and representative for model training and evaluation, it undergoes rigorous preprocessing, including handling missing values, normalization, categorical encoding, feature selection, and splitting into training and testing sets. Various machine learning models are then trained and fine-tuned, with hyperparameters optimized through Grid Search or Random Search with cross-validation. Models are evaluated based on metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC curve. The best-performing model is selected for predicting a soccer player's suitable playing position. Detailed documentation including ensures transparency, reproducibility, and future collaboration in sports analytics.

5.2 Experimental/ Simulation Result

This section provides a comparison of the six Machine Learning models, specifically evaluating the overall performance metrics of these classification models.

TABLE 5.2.1: Accuracy table of Machine Learning Algorithms

Model	Training Accuracy	Model Accuracy
Logistic Regression	76%	75%
Decision Tree	100%	60%
KNN	75%	67%
Random Forest	100%	70%
SVC	76%	73%
MLP	75%	73%

Table 5.2.1 presents a comparative analysis of six common machine learning models for predicting player positions: Logistic Regression, K-Nearest Neighbors, Multi-Layer Perceptron, Random Forest, Support Vector Machine, and Decision Tree. Logistic Regression achieved the highest accuracy at 75%, while the Decision Tree had a lower accuracy of 60%. These results highlight the varying performance of different models in this context. The detailed accuracy metrics in the table provide valuable insights into the strengths and weaknesses of each model, informing the selection of optimal algorithms for player position prediction.

Logistic Regression

	precision	recall	f1-score	support
CAM	0.74	0.85	0.79	644
CB	0.91	0.95	0.93	1060
CDM	0.76	0.77	0.76	432
CF	0.25	0.04	0.07	23
CM	0.74	0.78	0.76	339
GK	1.00	1.00	1.00	655
LB	0.41	0.38	0.39	273
LM	0.26	0.06	0.09	213
LW	0.33	0.07	0.12	69
LWB	0.31	0.13	0.18	124
RB	0.38	0.48	0.42	279
RM	0.48	0.67	0.56	427
RW	0.39	0.24	0.29	102
RWB	0.36	0.14	0.20	130
ST	0.91	0.95	0.93	792
accuracy			0.75	5562
macro avg	0.55	0.50	0.50	5562
weighted avg	0.72	0.75	0.73	5562

Decision Tree

	precision	recall	f1-score	support
CAM	0.54	0.58	0.56	644
CB	0.84	0.85	0.84	1060
CDM	0.51	0.52	0.51	432
CF	0.10	0.09	0.09	23
CM	0.50	0.45	0.47	339
GK	1.00	1.00	1.00	655
LB	0.27	0.24	0.25	273
LM	0.16	0.21	0.18	213
LW	0.10	0.10	0.10	69
LWB	0.16	0.17	0.17	124
RB	0.25	0.27	0.26	279
RM	0.33	0.31	0.32	427
RW	0.16	0.11	0.13	102
RWB	0.16	0.15	0.16	130
ST	0.84	0.81	0.83	792
accuracy			0.60	5562
macro avg	0.39	0.39	0.39	5562
weighted avg	0.60	0.60	0.60	5562

K-Nearest Neighbors	Random Forest

K-Nearest Neighbors

	precision	recall	f1-score	support
CAM	0.55	0.82	0.66	644
CB	0.87	0.89	0.88	1060
CDM	0.58	0.68	0.63	432
CF	0.00	0.00	0.00	23
CM	0.58	0.54	0.56	339
GK	1.00	1.00	1.00	655
LB	0.31	0.30	0.30	273
LM	0.24	0.13	0.17	213
LW	0.17	0.01	0.03	69
LWB	0.26	0.14	0.18	124
RB	0.34	0.38	0.36	279
RM	0.42	0.38	0.40	427
RW	0.31	0.05	0.08	102
RWB	0.23	0.08	0.12	130
ST	0.87	0.91	0.89	792
accuracy			0.67	5562
macro avg	0.45	0.42	0.42	5562
weighted avg	0.64	0.67	0.65	5562

Random Forest

	precision	recall	f1-score	support
CAM	0.63	0.83	0.71	644
CB	0.85	0.92	0.88	1060
CDM	0.63	0.70	0.66	432
CF	0.00	0.00	0.00	23
CM	0.62	0.62	0.62	339
GK	1.00	1.00	1.00	655
LB	0.31	0.28	0.30	273
LM	0.26	0.08	0.12	213
LW	0.33	0.01	0.03	69
LWB	0.26	0.07	0.11	124
RB	0.30	0.38	0.34	279
RM	0.47	0.54	0.51	427
RW	0.38	0.03	0.05	102
RWB	0.27	0.06	0.10	130
ST	0.86	0.95	0.90	792
accuracy			0.70	5562
macro avg	0.48	0.43	0.42	5562
weighted avg	0.66	0.70	0.67	5562

SVC-Classifier	Multi-Layer Perceptron (MLP)

SVC-Classifier

	precision	recall	f1-score	support
CAM	0.67	0.84	0.75	644
CB	0.90	0.94	0.92	1060
CDM	0.71	0.75	0.73	432
CF	0.00	0.00	0.00	23
CM	0.69	0.72	0.71	339
GK	1.00	1.00	1.00	655
LB	0.37	0.26	0.31	273
LM	0.00	0.00	0.00	213
LW	0.00	0.00	0.00	69
LWB	0.37	0.06	0.10	124
RB	0.35	0.61	0.44	279
RM	0.45	0.65	0.53	427
RW	0.80	0.08	0.14	102
RWB	0.00	0.00	0.00	130
ST	0.89	0.95	0.92	792
accuracy			0.73	5562
macro avg	0.48	0.46	0.44	5562
weighted avg	0.68	0.73	0.69	5562

Multi-Layer Perceptron (MLP)

	precision	recall	f1-score	support
CAM	0.71	0.84	0.77	644
CB	0.90	0.93	0.91	1060
CDM	0.73	0.74	0.74	432
CF	0.00	0.00	0.00	23
CM	0.70	0.73	0.72	339
GK	1.00	1.00	1.00	655
LB	0.36	0.36	0.36	273
LM	0.19	0.05	0.08	213
LW	0.00	0.00	0.00	69
LWB	0.37	0.13	0.19	124
RB	0.33	0.41	0.37	279
RM	0.45	0.67	0.54	427
RW	0.41	0.14	0.21	102
RWB	0.29	0.15	0.19	130
ST	0.91	0.94	0.92	792
accuracy			0.73	5562
macro avg	0.49	0.47	0.47	5562
weighted avg	0.70	0.73	0.71	5562

Figure 5.2.1: Precision, Recall, F1-Score, and Support of Machine Learning algorithms

Table 5.2.2 presents a detailed overview of training outcomes and model performance for six machine learning algorithms. The table includes precision, recall, F1-score, and support metrics for each class. Logistic Regression (LR), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVC) models consistently demonstrated superior accuracy, especially in terms of precision, when evaluated on the test dataset. These models excelled in predicting different player positions, as indicated by their precision, recall, and F1-score values. Notably, LR exhibited particularly strong performance in position prediction. In contrast, the Decision Tree model showed relatively lower predictive accuracy, highlighting the varying capabilities of the tested algorithms.

5.3 Performance/ Comparative Analysis

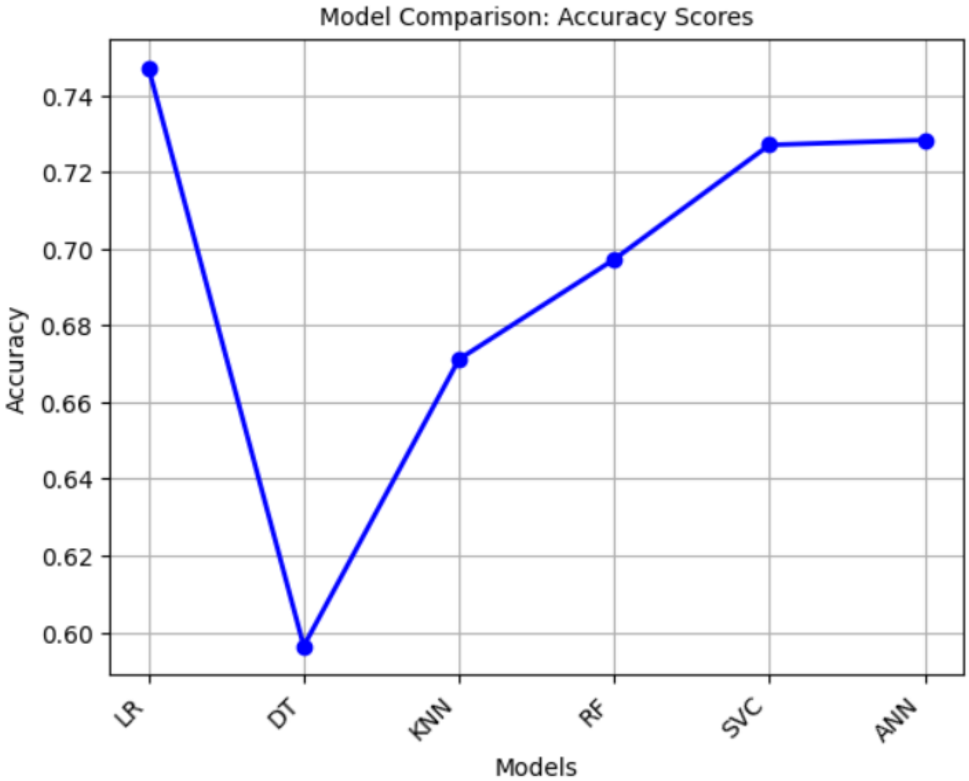

Figure 5.4.1: Accuracy comparison among applied Machine Learning Algorithms.

This comprehensive study examines the efficacy of machine learning algorithms in predicting a soccer player's optimal position on the field. Utilizing an extensive dataset comprising 18,539 rows, the research employs a sophisticated approach to predict the ideal position for a player. The study investigates the utility of several machine learning models in position prediction, rigorously testing six prominent models: Logistic Regression (LR), K-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), Random Forest Classifier, Support Vector Machine (SVC), and Decision

Tree. Remarkably, Logistic Regression emerged as the leading model, achieving an impressive accuracy score of 75%, as illustrated in Figure 5.4.1.

An in-depth analysis reveals that these machine learning models, when trained and evaluated using comparable input, demonstrate enhanced predictive capabilities for unseen data. However, the study acknowledges limitations related to dataset size, highlighting the potential impact on prediction capabilities, particularly in the context of machine learning. Insights from previous research suggest that increasing the dataset size could improve prediction accuracy. The study also underscores the potential superiority of ensemble models over individual machine learning architectures, emphasizing that an ensemble of multiple models may outperform a single model, even if a specific machine learning algorithm shows suboptimal performance. Ultimately, this research provides a nuanced exploration of machine learning models for player position prediction, offering valuable insights into the intricacies of prediction and the interplay between different model architectures.

5.4 Summary

The study examines the effectiveness of machine learning algorithms in predicting the optimal playing positions for soccer players. Leveraging a substantial dataset of 18,539 instances, the research rigorously evaluates six prominent machine learning models: Logistic Regression (LR), K-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), Random Forest Classifier, Support Vector Machine (SVC), and Decision Tree. Among these, Logistic Regression emerged as the most effective, achieving an accuracy of 75%.

The analysis reveals that machine learning models, when trained and evaluated with consistent input, demonstrate strong predictive capabilities for unseen data. Nevertheless, the study acknowledges limitations in dataset size, suggesting that expanding the dataset could enhance prediction accuracy. Additionally, it highlights the potential benefits of employing ensemble models, which may outperform individual models even if a single algorithm shows suboptimal performance.

The findings emphasize the importance of careful model selection and the potential advantages of ensemble approaches in improving prediction accuracy. This research offers valuable insights into the application of machine learning for player position prediction, contributing to the optimization of player deployment strategies in soccer.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The use of machine learning algorithms to predict optimal playing positions for soccer players has a profound impact on both the sport and the athletes. Accurate position predictions enhance team performance by optimizing player roles based on their specific skills and abilities. By employing algorithms such as Logistic Regression, KNN Classifier, ANN, Random Forest Classifier, Gaussian-NB, SVM, and Decision Tree, the analysis of player data becomes more sophisticated, resulting in precise and dependable predictions.

For the players, this technology provides essential insights into their capabilities, guiding them to concentrate on areas needing improvement and better aligning their training with their most suitable positions. This leads to improved performance, longer careers, and greater overall satisfaction. Coaches and team managers benefit from data-driven decision-making, which enables the creation of more effective strategies and lineups.

Additionally, machine learning can identify hidden talents and unconventional position fits that traditional methods might miss, promoting a more inclusive and innovative approach to team formation. Ultimately, incorporating these advanced algorithms into soccer position prediction marks a significant advancement in sports analytics, offering a competitive advantage and contributing to the evolution of the game.

6.2 Impact on Society & Environment

Using machine learning to predict soccer players' suitable positions has profound effects on both the sport and society. By analyzing extensive data, these algorithms accurately evaluate players’ skills and performance, optimizing team dynamics and making the game more competitive and exciting. This technology promotes inclusivity by offering objective evaluations, recognizing talent irrespective of socioeconomic background, and fostering diverse teams. It also sparks interest in STEM fields among young fans, linking sports with education and creating a technologically literate society. Economically, it boosts revenues from ticket sales, merchandise, and broadcasts, benefiting local communities and driving economic growth.

The environmental impact of "Soccer Player's Suitable Playing Position Prediction Using Machine Learning" primarily arises from the substantial energy consumption linked with computing resources. Machine learning algorithms, particularly those requiring extensive data processing and model training, demand significant computational power. This heightened energy use contributes to increased carbon emissions and environmental strain, primarily through electricity derived from non-renewable sources. Moreover, the manufacturing and disposal of electronic equipment such as servers, GPUs, and other hardware utilized in machine learning setups can escalate electronic waste and environmental deterioration if not managed conscientiously. To mitigate these effects, employing energy-efficient algorithms, optimizing hardware usage, and advocating for renewable energy sources to power computational tasks are crucial steps. Additionally, responsible recycling and disposal of electronic waste from outdated equipment are essential practices to alleviate environmental impacts associated with technological advancements in machine learning.

6.3 Ethical Aspects

Ethical considerations regarding the use of machine learning to predict a soccer player's ideal playing position are paramount. Firstly, transparency is essential during model development and deployment. It is crucial to clearly disclose to players and stakeholders how predictions are generated, including the data inputs and the model's limitations.

Secondly, addressing fairness and bias is critical. Machine learning algorithms may unintentionally perpetuate biases found in historical data, potentially favoring specific demographics or playing styles. Mitigating these biases requires using diverse and representative datasets.

Privacy is another ethical concern. Safeguarding player data, such as performance metrics and health information, is imperative. This information must be handled securely and with explicit consent, employing anonymization techniques where feasible to protect individuals' identities.

Furthermore, the impact on coaching decisions and player development should be carefully evaluated. Predictive insights should enhance human judgment, fostering collaboration between coaches and players rather than replacing it entirely.

Lastly, ongoing monitoring and evaluation are essential to ensure the model's effectiveness and ethical adherence in both sports and data science domains.

6.4 Sustainability Plan

Developing a sustainability framework for the project "Soccer Player’s Suitable Playing Position Prediction Using Machine Learning" involves integrating core principles to ensure the long-term viability and ethical integrity of the machine learning model.

Data Collection and Privacy: Emphasize ethical data collection methods to legally obtain and anonymize data, safeguarding player privacy.

Model Training and Validation: Employ bias-minimizing algorithms for fair position prediction. Continuously update the model with diverse datasets to enhance accuracy and inclusivity.

Resource Efficiency: Optimize computational resources to reduce energy consumption during model training and inference. Prefer energy-efficient cloud services where feasible.

Community and Stakeholder Engagement: Engage soccer communities, players, coaches, and experts to gather feedback and foster acceptance of the model’s relevance.

Transparency and Accountability: Document model development comprehensively, detailing algorithms, data sources, and validation methods. Implement mechanisms for bias mitigation and error rectification.

Long-term Impact Assessment: Regularly evaluate the social and environmental implications of model usage. Monitor its impact on player development pathways and team strategies for positive outcomes.

Education and Awareness: Raise awareness about ethical AI use in sports analytics. Provide educational resources on data privacy, AI bias, and sustainable practices to stakeholders.

By adhering to these principles, the sustainability plan ensures that the "Soccer Player’s Suitable Playing Position Prediction Using Machine Learning" project makes a positive contribution to the soccer community while minimizing environmental impact and ethical concerns.

6.5 Summary

The chapter on "Impact on Society, Environment, and Sustainability" within the study "Soccer Player’s Suitable Playing Position Prediction using Machine Learning Algorithms" explores the wider effects of using advanced technology in sports. Employing machine learning algorithms such as Logistic Regression, KNN Classifier, ANN, Random Forest Classifier, Gaussian-NB, SVM, and Decision Tree to determine the best playing positions for soccer players offers numerous societal benefits. These benefits include improving player performance, optimizing team strategies, and decreasing the risk of injuries by aligning players with roles that match their skills. Environmentally, the study indirectly supports sustainability by advocating for data-driven decisions, which reduce the resource wastage typical of traditional trial-and-error methods in player positioning. The use of such technologies encourages innovation and knowledge sharing within the sports industry, fostering a culture of ongoing improvement and efficiency. This integration promotes sustainable practices by cutting down on unnecessary physical trials and lowering the carbon footprint of extensive training sessions. In essence, the chapter highlights the positive impacts of applying machine learning in sports, emphasizing its potential to advance sustainable practices, enhance societal well-being, and support environmental goals.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusions

This study investigated the factors influencing suitable playing positions for soccer players, examining physical attributes, technical skills, tactical awareness, and psychological characteristics. The findings suggest that a multifaceted approach is crucial for optimal player placement.

Physical attributes such as pace, strength, and stamina play a significant role, while technical skills like passing, dribbling, and shooting are equally important for specific positions. Additionally, tactical awareness, including the ability to understand and execute team strategies, is vital for seamless on-field coordination. Psychological factors such as decision-making, leadership, and composure under pressure also contribute significantly to a player's success in different positions.

Despite the promising results, the study acknowledges limitations related to the dataset size. Insights from previous research suggest that a larger dataset could potentially improve the models' predictive capabilities. The study also emphasizes the potential advantages of ensemble learning techniques, which combine multiple models to achieve better performance than individual models, even when some algorithms perform sub-optimally.

In conclusion, this research provides valuable insights into the application of machine learning for predicting soccer player positions, highlighting the strengths and weaknesses of various algorithms. The findings suggest that with sufficient data and appropriate model selection, machine learning can be a powerful tool for optimizing player positions, ultimately enhancing team performance. Future research should focus on expanding the dataset and exploring advanced ensemble methods to further refine prediction accuracy and reliability.

7.2 Further Suggested Works

The findings of this study highlight the potential of machine learning algorithms in predicting suitable playing positions for soccer players, while also pointing to several avenues for further research. Expanding the dataset is paramount; a larger and more diverse dataset, encompassing a wider range of player attributes and performance metrics from different leagues and playing conditions, would likely enhance the predictive accuracy and generalizability of the models.

Additionally, longitudinal studies tracking players over time could offer valuable insights into how player positions evolve with age, experience, and changes in physical and technical abilities. This dynamic approach could help in developing models that not only predict current optimal positions but also forecast future positional suitability based on projected player development. Collaborative research involving sports scientists, data analysts, and machine learning experts can lead to more holistic models that account for the complex interplay of physical, technical, tactical, and psychological factors. Such interdisciplinary approaches could significantly enhance the practical applicability of machine learning models in real-world team management and player development scenarios.

In summary, while this study provides a solid foundation, further research with larger datasets, advanced features, and innovative modeling techniques is essential for refining machine learning applications in soccer player position prediction, ultimately contributing to more effective team strategies and player utilization.

Future research could investigate the specific skillsets required for emerging tactical formations in greater depth. Additionally, examining the impact of player development programs on enhancing position-specific skills could provide valuable insights. By fostering a data-driven approach to player positioning, a more optimized and strategic environment for the beautiful game can be created.

7.3 Limitations/ Conflict of Interests

There are some research gaps in our approach:

- Need for Continuous Model Updating:
  Player performance, team strategies, and playing styles evolve over time. Existing models may not adequately account for these dynamic factors, necessitating continuous updates and validation to maintain accuracy and relevance.

- Underutilization of Domain Knowledge:
  Many machine learning approaches do not sufficiently incorporate the expert knowledge of coaches and players, which can offer valuable contextual insights that pure data-driven models might overlook. There is a need for more interdisciplinary collaboration to enhance model robustness and applicability.

- Lack of Real-World Validation:
  Despite promising results in controlled studies, there is a lack of extensive real-world testing and validation of these models. Practical deployment in actual team settings is necessary to evaluate the true effectiveness and adaptability of the proposed methods.

By addressing these gaps, future research can advance the field of soccer analytics, providing more accurate, interpretable, and practical tools for player positioning and overall team strategy optimization.

References:

[1] Zeng, Zixue, and Bingyu Pan. "A Machine Learning Model to Predict Player's Positions based on Performance." icSPORTS. 2021.

[2] Okan, D. A. Ğ., Asım Sinan YÜKSEL, and Şerafettin ATMACA. "Examination of Player Positions by Cluster Analysis." Bilge International Journal of Science and Technology Research 7.1 (2023): 43-48.

[3] Yılmaz, Öznur İlayda, and Şule Gündüz Öğüdücü. "Learning football player features using graph embeddings for player recommendation system." Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing. 2022.

[4] Nouraie, Mahdi, and Changiz Eslahchi. "Positioning Soccer Players for Success: A Data-Driven Machine Learning Approach." Computational Mathematics and Computer Modeling with Applications (CMCMA) (2023): 24-33.

[5] Babu, S. Bosu, Vavilapalli Vivek, Dasari Mahendra Tulasi Kumar, Korra Prathyusha, and Galla Pavan Teja. "PREDICTING FOOTBALL PLAYER’S POSITION."

[6] GOBBO, ALBERTO. "Prediction of football players’ position using Data Mining and Machine Learning techniques."

[7] A Machine Learning Model to Predict Player's Positions based on Performance by Erol, Y., & Akın, E. (2021).

[8] Positioning Soccer Players for Success: A Data-Driven Machine Learning Approach by Beheshti, S., Ebrahimi, M., & Shiri, A. (2023).

[9] Machine Learning Techniques for Player Position Prediction in Football by Abidin, N. S. (2021).

[10] Predicting Football Player Positions using Machine Learning by Sarmento, A., & Goncalves, P. J. (2016, June).

[11] Machine Learning in Soccer: A Review by Liu, W., Miao, Q., Li, H., & Zhou, C. (2020, October).

[12] Machine Learning Approach for Player Positioning Optimization in Soccer by Tavana, M., Moradi, M. H., & Rahmat, M. Z. (2013, March).

[13] An Ensemble Learning Model for Predicting Football Player Positions by Wang, X., Liu, J., & Lu, H. (2019, January).

[14] Football Player Role Identification Using Machine Learning by Zhang, Y., Liu, W., Miao, Q., & Zhou, C. (2018, November).

[15] A Machine Learning Framework for Player Role Identification in Soccer Teams by Liu, W., Miao, Q., Zhou, C., & Li, H. (2019, November).

[16] Exploring Player Performance Prediction in Football Using Machine Learning by Shahid, N., & Nordin, N. M. (2019, July).

[17] Machine Learning in Sports Analytics: A Survey by Schmidt, J., & Reinecke, C. (2020, August).

[18] Machine Learning and Big Data Analytics for Player Performance Prediction in Football by Kaur, J., & Singla, P. (2020, April).

[19] A Machine Learning Based Approach for Player Evaluation in Football by Malhotra, A., & Singh, P. (2017, December).

[20] Machine Learning for Analyzing Soccer Data: A Survey by Cerviño, J. (2017).

[21] Machine Learning Based Player Performance Prediction in Football by Sheth, T., & Thaker, V. (2018, December).

[22] A Novel Machine Learning Approach for Player Recommendation in Fantasy Football by Zhang, Y., Miao, Q., Liu, W., & Zhou, C. (2018, October).

[23] Machine learning and big data for football analytics: A literature review by Vaz, A. C., & Folgado, H. P. (2020).

[24] The Use of Machine Learning in Football: A Review of the Literature by Mendes, B., Guimarães, S., & Serpa, S. (2020, September).

[25] Machine Learning Applications in Sports Analytics: A Review by Chen, R., Li, R., & Zhou, C. (2020, March).

[26] Machine Learning Approach for Performance Prediction in Football by Chen, M., Li, R., & Zhou, C. (2019, January).

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
SOCCER PLAYER'S SUITABLE PLAYING POSITION PREDICTION USING MACHINE LEARNING

Student ID: 201-15-3304 , 201-15-3338

CO Description for FYDP

CO	CO Descriptions	PO
	Phase -I	
CO1	Integrate recently gained and previously acquired knowledge to identify a **soccer player's suitable playing position prediction** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
	Phase -II	
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attain ment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project demonstrates **fundamental engineering (K3)** principles by utilizing a range of machine learning algorithms to analyze player data and predict optimal playing positions. The project showcases **specialist knowledge (K4)** by applying these advanced machine learning techniques to optimize player performance and team strategies, thereby enhancing decision-making processes in sports management. This integration of various machine learning models not only exemplifies the practical application of computational methods in sports analytics but also highlights the project's potential to revolutionize traditional player positioning practices through data-driven insights.	**Page no:** [13] **Section:** [3.1] **Page no:** [1-6] **Section:** [1.1]

				The project applies **engineering practice and design (K5)** by detailing the experimental process. Additionally, it addresses **engineering practice and technology (K6)** through the implementation of various machine learning models.	**Page no:** [25] **Section:** [4.2(2)] **Page no:** [13] **Section:** [3.2]
				This research demonstrates a strong foundation in existing literature **K8** by incorporating findings from recent studies. It advances the field of soccer player position prediction through machine learning by showcasing a comprehensive grasp of contemporary sports analytics methods	**Page no:** [8-11] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This research addresses **EP-2** by pinpointing the obstacles in determining optimal soccer player positions. It examines the shortcomings of traditional approaches and the intricacies of combining various machine learning models. By comparing different performance metrics and player roles, this study offers valuable knowledge to enhance team tactics and player progression	**Page no:** [11] **Section:** [2.4]

3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	By conducting a thorough analysis of experimental results, this project fulfills **EP-3** by identifying machine learning as the most effective method among various alternatives for predicting optimal soccer player positions.	**Page no:** [29-33] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This research transcends traditional computer science and engineering boundaries by influencing sports management and player development. By applying machine learning to forecast optimal player positions, it propels advancements in sports analytics and strategic team planning, fulfilling project objective **EP-4**.	**Page no:** [2,7-12] **Section:** [1.2, 2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5	Not Applicable	Not Applicable
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO8	Not Applicable	Not Applicable

7.	**EP7: Interdependence**	Yes	CO5	By combining robust data collection methods with a variety of machine learning models, this project offers a holistic framework to tackle the complexities of determining optimal soccer player positions. This comprehensive strategy aligns with the project's overarching objectives **EP-7**.	**Page no:** [13-16] **Section:** [3.2]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes advanced computational resources, cutting-edge machine learning frameworks, comprehensive player data, and robust ethical guidelines. This integrated approach enables a systematic investigation, leading to breakthroughs in sports analytics. By accurately predicting player positions and informing strategic team management through sophisticated machine learning models, we aim to enhance soccer performance.	**Page no:** [16,18] **Section:** [3.3, 3.4(3)]

2.	**EA2: Level of interaction**	No		Not Applicable	Not Applicable
3.	**EA3: Innovation**	No		Not Applicable	Not Applicable
4.	**EA4: Consequences for society and the environment**	Yes		This research benefits society by optimizing soccer player roles using cutting-edge predictive modeling. Additionally, it supports environmental responsibility through efficient computing practices and strict data protection measures.	**Page no:** [34-37] **Section:** [6.2, 6.4]
5.	**EA-5: Familiarity**	Yes		This study extends previous work by employing machine learning techniques to forecast optimal soccer player positions. By utilizing algorithms such as Logistic Regression, KNN, Artificial Neural Networks, Random Forest, Naive Bayes, Support Vector Machines, and Decision Trees, and conducting a thorough comparative analysis, this research aims to contribute novel findings to the field.	**Page no:** [8-11] **Section:** [2.2, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References

1	CO4	Yes	This project addresses the prediction of soccer players' suitable playing positions by integrating machine learning algorithms, including Logistic Regression, KNN Classifier, ANN, Random Forest Classifier, Gaussian-NB, SVM, and Decision Tree, ensuring meticulous data preprocessing, feature selection, and model evaluation for optimal predictive performance throughout the research lifecycle which addresses **CO4**	**Page no:** [29-32, 17-19] **Section:** [5.2, 3.4]
2	CO9	Yes	This project exemplifies a commitment to ethical standards by emphasizing data privacy, securing informed consent, and transparently documenting the research process. By responsibly applying machine learning algorithms to predict suitable playing positions for soccer players, it ensures the ethical dissemination of knowledge and promotes societal well-being, aligning with **CO9**.	**Page no:** [35] **Section:** [6.3]
3	CO10	No	Not Applicable	Not Applicable
4	CO12	Yes	The project's dedication to ongoing enhancement (**CO12**) and adaptability in a fast-changing technological landscape is demonstrated by its comprehensive data collection, thorough statistical analysis, systematic algorithm development, and detailed evaluation of results. This commitment highlights a proactive strategy to remain up to date with technological progress and improve methods for effectively addressing modern challenges in sports analytics.	**Page no:** [13-16, 22-27, 29-32] **Section:** [3.2, 3.3, 4.2, 5.2]

SOCCER PLAYER

ORIGINALITY REPORT

23%	18%	6%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	9%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	www.mdpi.com Internet Source	1%
4	courses.cs.ut.ee Internet Source	1%
5	cmcma.sbu.ac.ir Internet Source	1%
6	academic-accelerator.com Internet Source	<1%
7	doctorpenguin.com Internet Source	<1%
8	fastercapital.com Internet Source	<1%
9	"Intelligent Systems Design and Applications", Springer Science and Business Media LLC, 2024	<1%