A Project Report On

Cricket Concussion Replacement Recommendation System

Submitted in partial fulfillment of the requirements

for the degree of

Bachelor of Engineering

by

Patel Mohammed Rehan Irfan - 412043

Sayyed Affan Abdullah - 412050

Farooqui Mohammad Bilal Mohammad Alim - 412015

Khan Jahangir Alam Faizan - 412028

Supervisor:

Prof. Javed Taili

Computer Science and Engineering
Artificial Intelligence and Machine Learning

Anjuman I Islam's
M.H. Saboo Siddik College of Engineering

2023-24

Cricket Concussion Replacement Recommendation System

Submitted in partial fulfillment of the requirements

for the degree of

Bachelor of Engineering

by

Patel Mohammed Rehan Irfan - 412043

Sayyed Affan Abdullah - 412050

Farooqui Mohammad Bilal Mohammad Alim - 412015

Khan Jahangir Alam Faizan - 412028

Supervisor:

Prof. Javed Taili

Computer Science and Engineering
Artificial Intelligence and Machine Learning

Anjuman I Islam's
M.H. Saboo Siddik College of Engineering

2023-24

CERTIFICATE

This is to certify that the project entitled **"Cricket Concussion Replacement Recommendation System"** is a bonafide work partly done by

"Patel Mohammed Rehan Irfan (412043)"

is submitted in partial fulfillment of the requirement for the award of the degree of

"Undergraduate" in **"B.E Computer Science and Engineering (Artificial Intelligence and Machine Learning)"**.

to the

University of Mumbai

Prof. Javed Taili Supervisor/Guide	(Name and sign) Examiner

Prof. Tarannum Shaikh Project Coordinator	Dr. Jilani Sayyad Coordinator (CSE-AIML)	Dr. Jawed Habib I/C Principal

Project Report Approval for B. E.

This project report entitled ***Cricket Concussion Replacement Recommendation System***

partly done by

Patel Mohammed Rehan Irfan

is approved for the degree of **B.E Computer Science and Engineering (Artificial Intelligence and Machine Learning)**.

Examiners

1. Prof. Javed Taili _____

2. External _____

Date:

Place:

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Patel Mohammed Rehan Irfan)
(412043)

Date:

Abstract:

Imagine the thrilling last few overs of a nail-biting cricket contest. Now let say that a critical player sustains a concussion which completely throws off the team's game plan. According to the ICC, athletes who sustain concussions during a game may be substituted as opposed to their team's initial line-up. The rules, however, place a strong focus on substituting a like-for-like player for the injured player and that aspect is still up in the air. In the event that a team wishes to substitute a player during a game, the match officials mainly the Match Referee will have a major say. There arises a need for a way to base these critical substitutions on hard, cold data and make them faster and more accurate.

This is the role that our project plays. We are developing a recommendation system that will revolutionize cricket player replacements by leveraging the power of machine learning. Every choice made in the fast-paced, high-stress environment of this sport has the potential to determine the outcome. Tactics changes, exhaustion, and injuries can all throw off a team's flow. By giving match officials the access to real-time information on the ideal replacement player, our technology seeks to reduce these risks.

This project goes beyond practicality. It has to do with reducing risk and making sure that a team doesn't perform worse when faced with such adversities. Consider the scenario where the best batter on the squad is out due to a concussion injury. Using data analysis and machine learning our system can identify the player with the most similar scoring rate and batting style based on their historical data among a group of possible replacements. This lessens the impact of the injury by assuring that the team's batting order is not significantly disrupted due to severe injuries like concussion.

We have collected a dataset that includes statistics such as batting averages, bowling economy rate, etc. We are essentially teaching the system to spot patterns and identify players with comparable strengths and weaknesses by putting this data into a machine learning algorithm called K Nearest Neighbours. When a player suffers a concussion injury, the algorithm evaluates the remaining players based on their historical performance and suggests the player whose skill set most closely matches the injured player's.

Like many other sports, cricket is now utilizing data analytics to its advantage. This trend is hence demonstrated by our project which aims to give the management authorities the practical insights they require to justify and ensure that the replacement made is as similar to the currently injured player as possible. Our intelligent recommendation system promises to transform player substitutions improving the game strategically and bring a new era of data-driven brilliance for cricket.

In a world where every wicket and every run matters, our project is a paradigm change. We are heading towards a future where data-driven, well-informed decisions take cricket to new heights so as to surpass mere guesswork and intuition. This encourages the development of a smarter cricket culture in which appropriate choices based on actual data are the norm. It is more than a technological advancement. It is about using information to make cricket a more entertaining and strategic game for players, supporters and the cricket community as a whole.

Contents:

List of Figures

Abbreviation Notation and Nomenclature:

ML	Machine Learning
NN	Neural network
AI	Artificial Intelligence
mTBI	Mild Traumatic Brain Injury
KNN	K-Nearest Neighbours

Organization of the Project Report

Organization of topics in the project report is as follows.

Chapter 1 gives the introduction of the motivation and objective of our project

Chapter 2 explains the literature survey of existing systems done for the project

Chapter 3 explains the proposed system or solution in detail

Chapter 4 explains the actual outputs achieved from the system and the future scope of the project.

Chapter 1

Introduction

1.1 Motivation

A concussion is categorized as a mild traumatic brain injury (mTBI) which is essentially a blow to the head that causes a disruption in normal brain function. Direct impacts such as being struck by a cricket ball or indirect impacts such as the head whipping back and forth from an abrupt movement can both generate this impact. Although it looks harmless, the force causes the complex connections in the brain to be disrupted which results in a brief change in the chemical composition of the brain. There are other symptoms that might accompany this disturbance such as headaches, nausea, dizziness and disorientation. Memory gaps are also prevalent especially with regard to the event itself that caused the concussion.

The degree of discomfort experienced can differ significantly. Some players may only have minor discomfort while others may need a longer recovery period with more incapacitating consequences. Even though most concussions go away in a few weeks with enough rest, it is imperative to consult a doctor as soon as possible after suffering a brain injury. For the team to stay on track and remain cohesive strategically, it becomes critical to replace an injured player with someone who plays to their strengths as well to make sure that the replacement does not give an unfair to any of the two teams.

1.2 Objective

This is where our recommendation system comes into play. Through the analysis of an extensive collection of player information which includes bowling economy rate, batting averages, etc., the system is able to find a player that shares a very comparable attributes with the injured player. This like-for-like approach ensures that the team's strategy is less likely to be disrupted due to concussion injuries.

One thing we have to note is that the human factor is still very important. Match referees have a unique awareness of the dynamics of the game along with the unique pressures of each situation and the intangible characteristics that distinguish genuinely suitable players. However, they can be empowered in making the decision in regards to concussion replacement as instructed by the ICC with a recommended list of objectively strong prospects which are filtered out from the pool of available players using machine learning technique known as K-Nearest Neighbours. This enables the referee to make a well-informed choice that takes into account both their own professional judgment and the data-driven insights.

Through our system we aim to leverage this combination of machine learning and human intuition to produces a more sophisticated and nuanced method for concussion replacement which assures that the flow of the match will not be hampered by concussion injuries in particular.

Chapter 2

Literature Survey

Sr. No.	Year, Journal, Publisher	Authors and Title of Paper	Results Reported	Research Gap Identified
1	2019, Journal of Sports Analytics 5, IOS Press.	J.R. Khan et al., A quantitative approach to influential factors in One Day International cricket: Analysis based on Bangladesh.	1. Employed a modified Poisson approach and logistic regression to analyze 314 One Day International (ODI) match results. 2. Bangladesh showed a higher percentage of wins in day matches (62.4%) compared to day-night (37.6%) during the last 20 years.	1. Logistic model produced more extreme outcomes with wider confidence intervals. 2. Only focuses on Bangladesh Team.
2	2018, Journal of Sports Analytics 4, IOS Press.	K.P. Jayalath, A machine learning approach to analyze ODI cricket predictors.	1. Introduced a CART (Classification and Regression Trees) based modeling strategy as an alternative to logistic regression. 2. The effects of predictors were analyzed separately for day and day-night games.	1. The paper introduces CART-based modeling against logistic regression but it does not provide a comparative analysis of the two modeling approaches. 2. No Practical Implementation done

Chapter 3

Proposed System

3.1 Data Acquisition

The system’s foundation lies in a robust dataset encompassing a wide range of player statistics. This data acquisition process is the first step as illustrated in the flowchart. Here, the system harvests data on various player attributes and performance metrics from Howstat.com, a popular cricket website. It accomplishes this feat through as Excel Web Connection which is essentially an automated method of extracting data from web sources into Excel spreadsheets.

3.2 Data Cleansing and Preprocessing

Once the raw data is collected, the sytem indicates a crucial stage of data cleansing and preprocessing. This step ensures the data is accurate and usable for the machine learning algorithms. In this phase, any unwanted characters or inconsistencies within the data are meticulously identified and removed. For instance, there are be unwanted characters or unwanted datatypes in the data that needed to be eliminated.

3.3 Data Normalization

After the data is cleansed, We move onto data normalization. Normalization ensures that all the data points are on an equal footing, essentially speaking the same language. This is particularly important when dealing with attributes measured on different scales. Imagine comparing apples and oranges. It wouldn't be an accurate comparison. In the context of cricket, this involves normalizing the attributes of the players to their experience bias by normalizing them per

matches and innings played. By normalizing the data, the system ensures that the machine learning algorithms can analyze all the attributes effectively without any bias towards a particular metric.

3.4 Feature Engineering

Our system then incorporates feature engineering. This stage involves creating new attributes, or features, from the existing data that might prove to be more informative for the machine learning algorithms. It's akin to crafting new tools from the ones you already have. For instance, the system might create a new feature that combines a batsman's average runs scored per match, etc., providing a more comprehensive understanding of their overall batting effectiveness.

3.5 Data Separation

Following feature engineering, the data is strategically separated into two distinct categories, as depicted in the flowchart. We first filter out all the features by incorporates all the player data except for the player names or IDs. These labels essentially act as identifiers for each player. We then separate the batting and bowling features so as to make separate recommendation for Batting and Bowling. This separation is a crucial step before the training process as without the separation the model might generate faulty recommendation due confusion in weighing the parameters.

3.6 K-Nearest Neighbours

The core of the system revolves around a machine learning model, specifically a K-Nearest Neighbours (KNN) algorithm. KNN is a type of supervised learning algorithm, which means it learns from labeled data. In this case, the labeled data

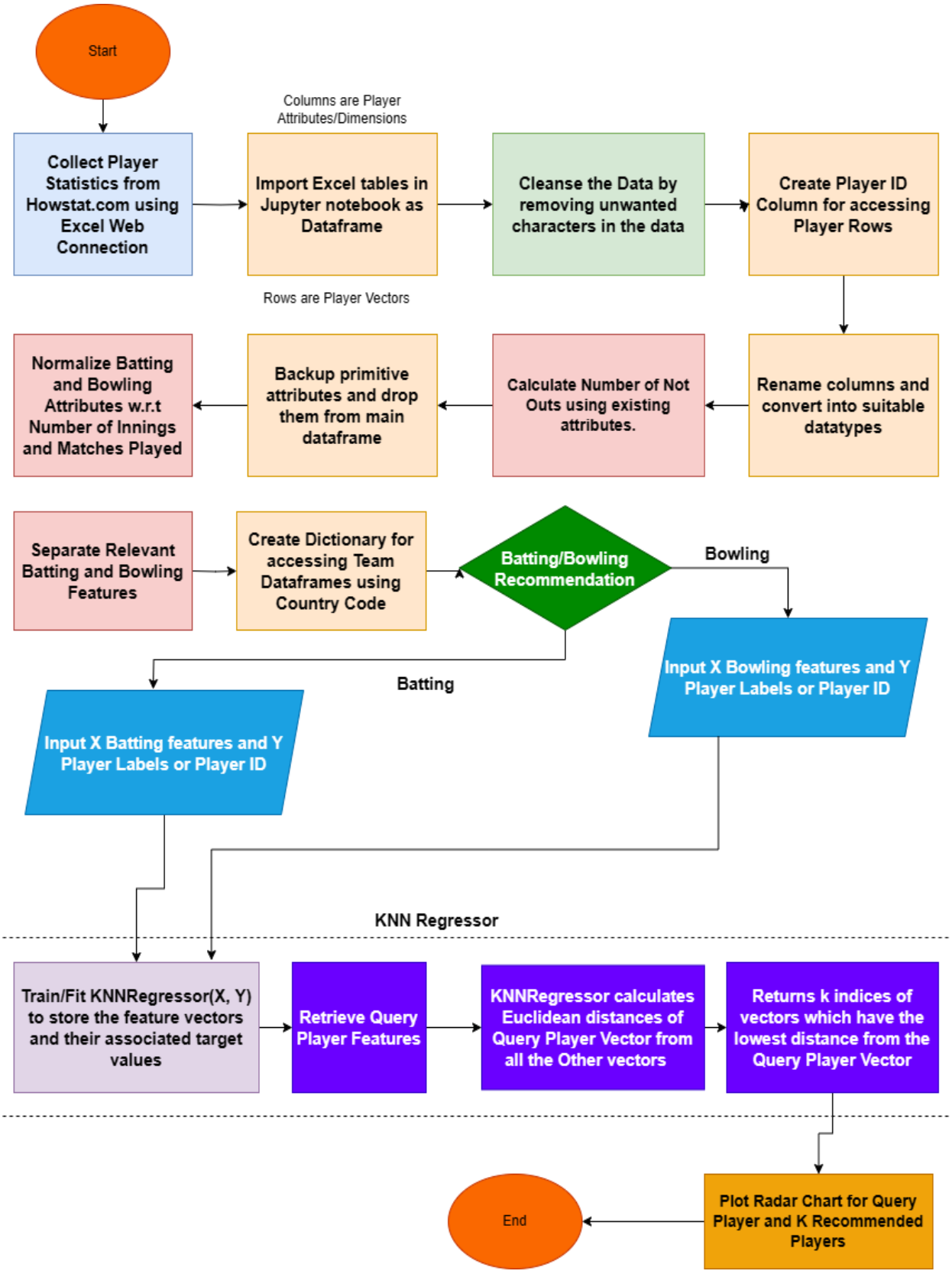
is the dataset we meticulously prepared in the earlier stages. KNN functions by identifying the ‘k’ closest data points, or ‘neighbours’ to a new query point (the injured player) based on Euclidean distance. These neighbours are then used to make predictions about the query point. In the context of the system, the query point would be the injured player, and the k nearest neighbours would be the players in the dataset most similar to the injured player in terms of their playing style and attributes.

3.7 Model Training

Our system then incorporates a stage known as model training. Here, the KNN model is meticulously trained using the prepared dataset. During training, the model ingests the player data, excluding the player labels or IDs, and learns to recognize the patterns and relationships between the various attributes. This empowers the model to identify the k closest neighbours to any new data point it encounters.

3.8 Making Recommendations

Once the model is trained, the system is ready to provide recommendations during a match scenario. As illustrated in the flowchart, when a player suffers a concussion or injury, the system retrieves the injured player’s data, encompassing their key performance metrics. This data acts as the query point for the KNN model. The model then leverages its learned knowledge to identify the k closest neighbors within the dataset, the players whose attributes are similar to original player resulting in a like-for-like replacement.


Start
Collect Player Statistics from Howstat.com using Excel Web Connection
Columns are Player Attributes/Dimensions
Import Excel tables in Jupyter notebook as Dataframe
Cleanse the Data by removing unwanted characters in the data
Create Player ID Column for accessing Player Rows
Rows are Player Vectors
Rename columns and convert into suitable datatypes
Calculate Number of Not Outs using existing attributes.
Backup primitive attributes and drop them from main dataframe
Normalize Batting and Bowling Attributes w.r.t Number of Innings and Matches Played
Separate Relevant Batting and Bowling Features
Create Dictionary for accessing Team Dataframes using Country Code
Batting/Bowling Recommendation
Bowling
Batting
Input X Bowling features and Y Player Labels or Player ID
Input X Batting features and Y Player Labels or Player ID
KNN Regressor
Train/Fit KNNRegressor(X, Y) to store the feature vectors and their associated target values
Retrieve Query Player Features
KNNRegressor calculates Euclidean distances of Query Player Vector from all the Other vectors
Returns k indices of vectors which have the lowest distance from the Query Player Vector
Plot Radar Chart for Query Player and K Recommended Players
End

Chapter 4

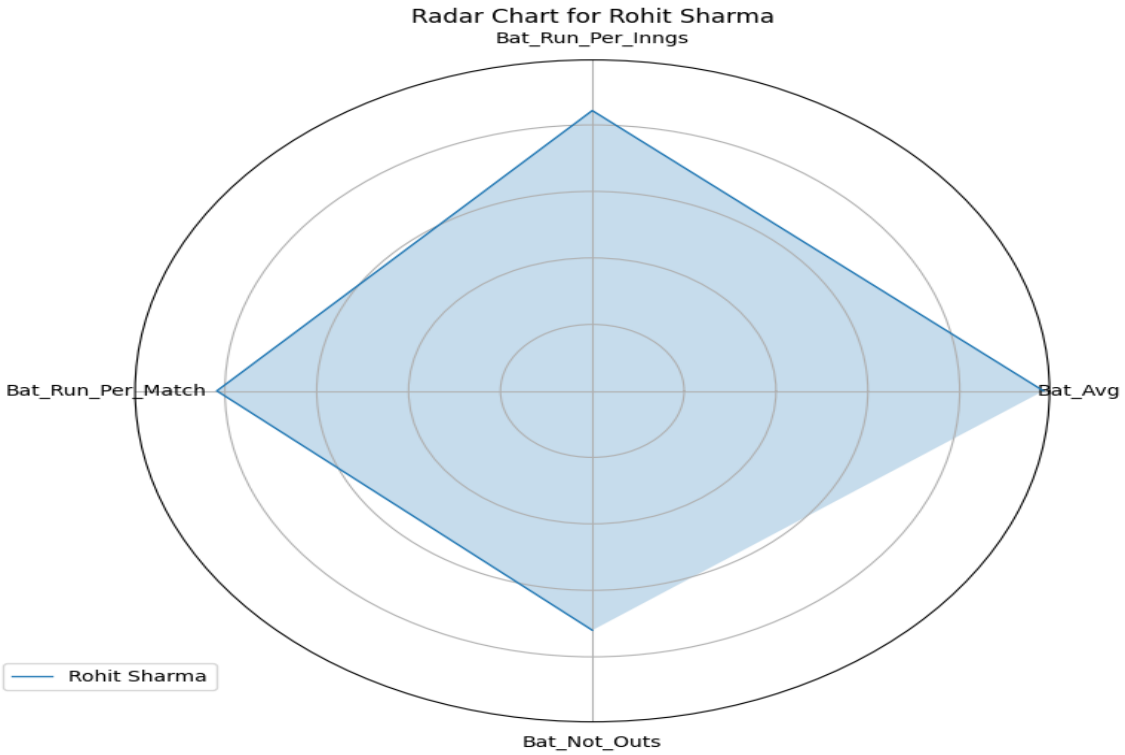
Results and Discussions

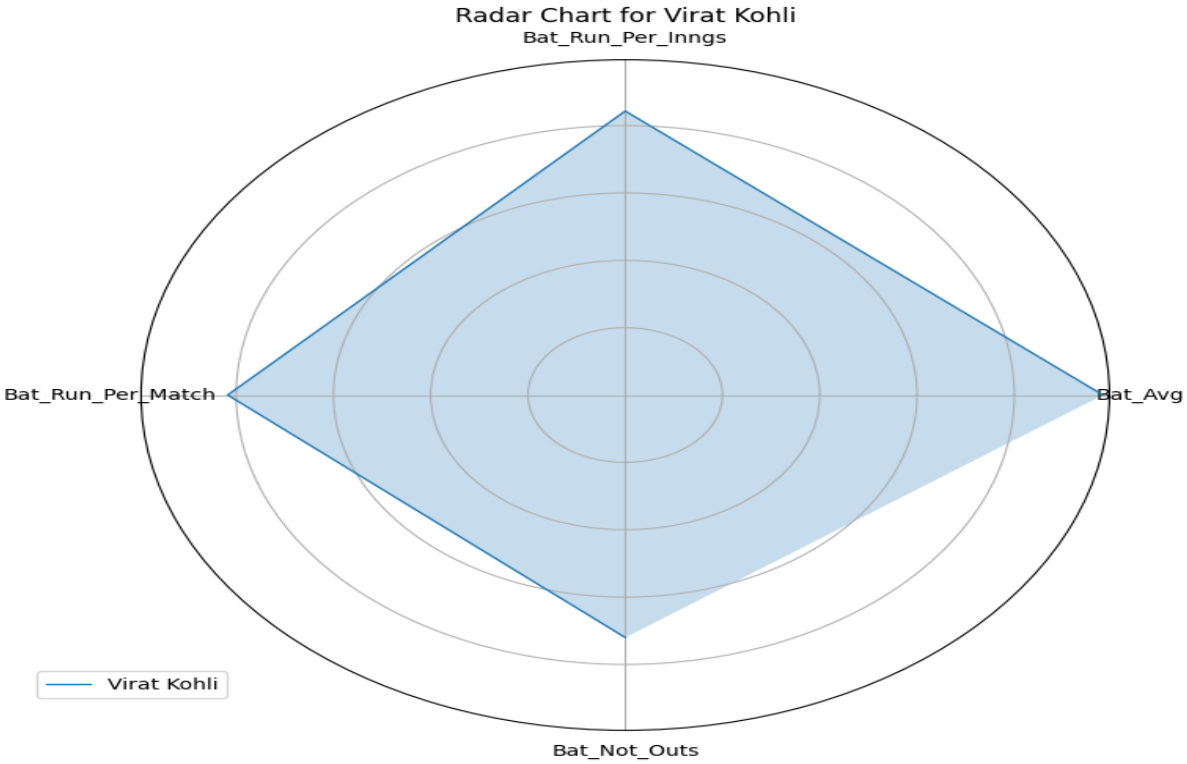
4.1 Radar Charts

Our project incorporates radar charts also known as spider charts to visually represent the multifaceted skillsets of cricket players. Imagine a spiderweb, with each spoke radiating outward from a central point. In our system, these spokes represent various player attributes like batting average, bowling strike rate, etc. The distance along each spoke corresponds to the player's performance in that specific category. High values are further away from the center while lower values reside closer. By connecting these data points on each spoke, a polygon is formed, creating a unique visual signature for each player. This chart offers a quick and clear comparison of a player's strengths and weaknesses across various cricketing disciplines. It empowers match referees, particularly when using our recommendation system to effectively assess the fit of potential replacements based on their skillset similarity to the injured player. Essentially, the radar chart acts as a visual translator, transforming complex data into an easily digestible format that aids informed decision-making.

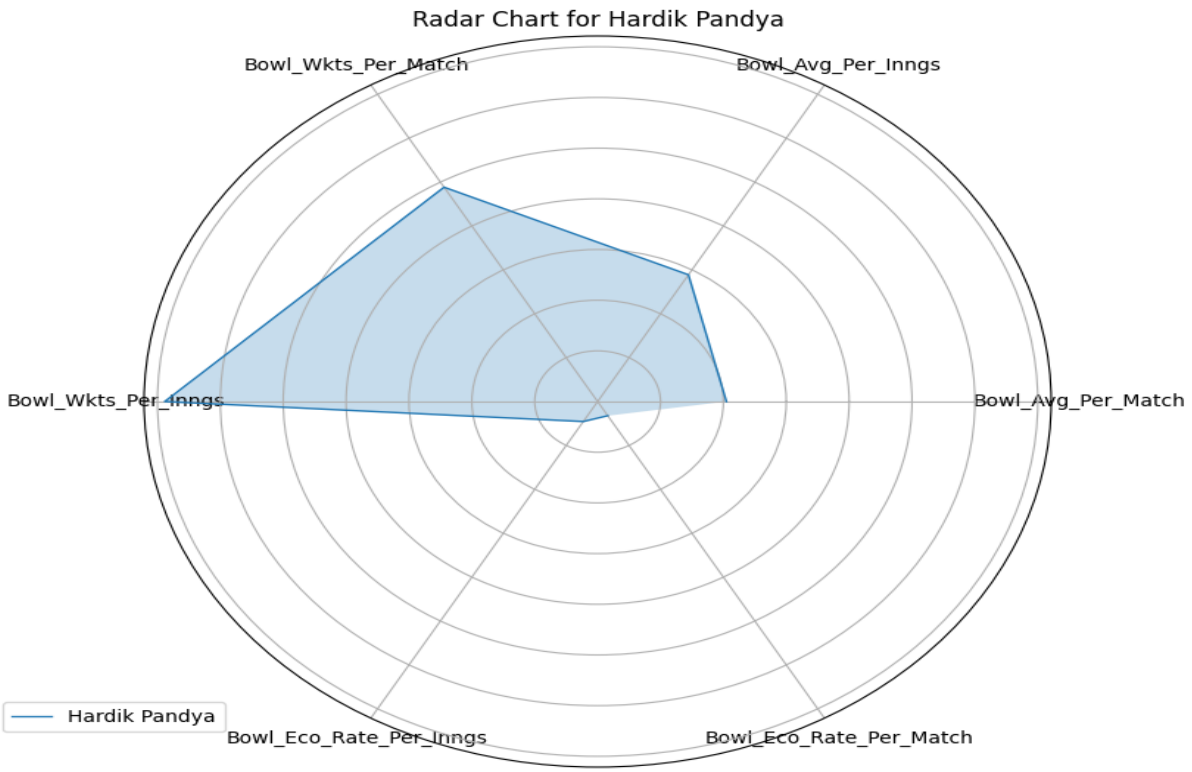
4.2 Like-for-like Batting Recommendation

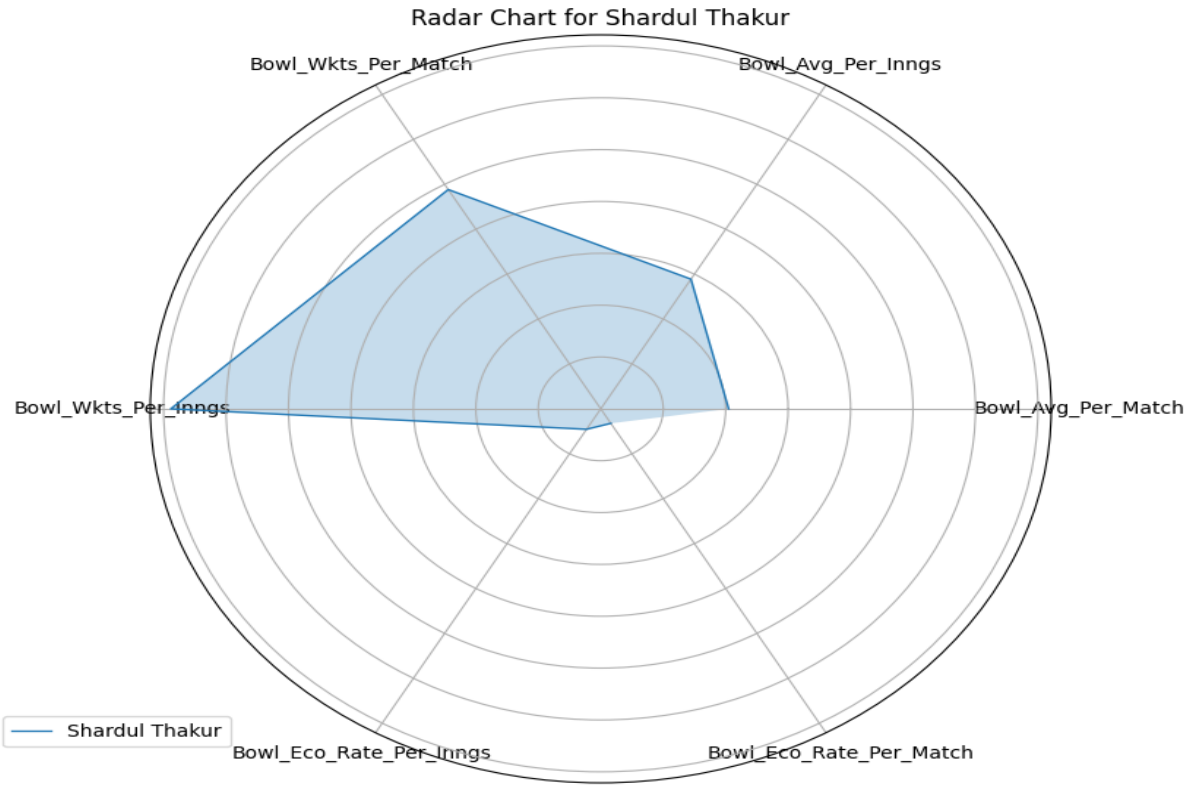
From the Radar chart of 2 players first the supposedly injured player, Rohit Sharma and the second player who is also the first recommendation from our model, Virat Kohli, can both be observed as similar in regards to their playing attributes just by observing the similarity in their Radar Chart. Our system hence correctly recommended Virat Kohli as a like-for-like replacement of Rohit Sharma.





Radar Chart Analysis for Batting Recommendation





Radar Chart Analysis for Bowling Recommendation

4.3 Like-for-like Bowling Recommendation

For example, from the Radar chart of 2 players first the supposedly injured player, Hardik Pandya and the second player who is also the last recommendation from our model, Shardul Thakur, can both be observed as similar in regards to their playing attributes just by observing the similarity in their Radar Chart. Our system hence correctly recommended Shardul Thakur as a like-for-like replacement of Hardik Pandya taking into consideration that all the more similar players were Spinners or dedicated fast bowlers whereas Hardik Pandya is a medium fast bowler.

4.4 Future Scope

The future of our "Like-for-Like" Cricket Player Recommendation System is brimming with exciting possibilities. A System where there are more segregation in the types of players like fast bowler, spinners, etc can be beneficial. We envision expanding the system's capabilities beyond concussions, encompassing a wider range of injuries and strategic substitutions. Additionally, the system could integrate real-time data on player fitness and form further refining its recommendations. Furthermore, incorporating fan preferences and sentiment analysis could introduce a fascinating layer, allowing the system to consider not just player fit but also crowd favorites. Ultimately, this project has the potential to evolve into a comprehensive cricketing brain by transforming the way player selection and strategy are approached in the dynamic world of cricket.

Chapter 5

Conclusion

In conclusion, our project has unveiled a revolutionary approach to player substitutions in cricket particularly in the face of concussion injuries. By harnessing the power of machine learning and a meticulously crafted "Like-for-Like Cricket Player Recommendation System', we have strived to bridge the gap between intuition and data-driven decision-making. The system fueled by a K-Nearest Neighbors algorithm meticulously analyzes a comprehensive dataset of player attributes, encompassing not just batting but also bowling prowess. This nuanced understanding empowers match referees to make informed substitutions for minimizing disruption to team strategy and maximizing the cricket fandom. Our project transcends mere technological innovation; it fosters a new era of informed decision-making, enhanced player safety and a more objective cricketing landscape. Ultimately, this system represents a significant leap forward, ushering in a future where cricket strategy is as much about algorithms and data analysis as it is about experience and intuition.

Chapter 6

References

IEEE standard

Journal Paper,

[1] J.R. Khan et al, "A quantitative approach to influential factors in One Day International cricket: Analysis based on Bangladesh" *Journal of Sports Analytics 5, IOS Press*, 2019.

[2] K.P. Jayalath, "A machine learning approach to analyze ODI cricket predictors" *Journal of Sports Analytics 4, IOS Press*, 2018.

Online References,

[1] About ICC Cricket | International Cricket Council (icc-cricket.com)