STARC '23 Abstract

Spectoclava

Sport:	**Cricket**
Leader:	**P Sai Charan (PES2UG21CS364)**
Members:	• **Muhammad Ashar Reza (PES2UG21CS306)** • **Pranit Prasant Pai (PES2UG21CS388)** • **Saksham Alok (PES2UG21CS461)** • **P Sai Charan (PES2UG21CS364)** • **Pratham (PES2UG21CS392)**

Project Details

Background

In the era of data-driven decision-making, the need for effective matchup analysis has become crucial in various fields. Cricket is a popular sport played internationally, and effective matchup analysis plays a vital role in understanding team performance, player statistics, and strategic decision-making. This abstract report outlines the development of a data-based matchup analyzing dashboard for the game of cricket.

Sometimes players end up performing better against certain types of players and this often has a pattern that can be predicted using concepts of machine learning. This dashboard will utilize web scraping techniques to extract relevant data. Then the extracted data will be visualized on a website for the user to use and help make decisions

Goals, Objectives and Scope

Goal: Develop a comprehensive matchup analysis dashboard for cricket that utilizes data-driven techniques to enhance decision-making.

- Objective 1: Create a web scraping module to collect matchup data from diverse cricket sources.
- Objective 2: Develop data processing algorithms to clean, preprocess, and analyze the extracted data.
- Objective 3: Design and implement a visually appealing and user-friendly website for data visualization.
- Objective 4: Implement query functionality to allow users to retrieve specific matchup data.

Goal: Enable data-driven insights and predictions for team performance and player statistics in cricket.

- Objective 1: Utilize machine learning techniques to identify patterns and trends in matchup data, enabling accurate predictions of player performance against specific opponents.
- Objective 2: Provide statistical analysis and visual representations of player and team performance metrics to support strategic decision-making.
- Objective 3: Develop algorithms to uncover hidden correlations and dependencies between different variables, helping teams identify key matchups and exploit potential advantages.

Goal: Enhance the accessibility and usability of matchup analysis in cricket.

- Objective 1: Build a user-friendly interface that allows easy navigation and interaction with the matchup analysis dashboard.
- Objective 2: Incorporate responsive design principles to ensure the website functions well across various devices and screen sizes.
- Objective 3: Provide intuitive query functionality that enables users to quickly retrieve specific matchup data and make informed decisions based on their requirements.

Goal: Ensure the accuracy, reliability, and robustness of the matchup analysis dashboard.

- Objective 1: Implement robust web scraping techniques to retrieve accurate and up-to-date matchup data from reliable sources.
- Objective 2: Apply rigorous data processing and validation algorithms to ensure the cleanliness and consistency of the extracted data.
- Objective 3: Conduct thorough testing and validation of the dashboard's functionalities to identify and fix any potential bugs or issues.

Applications:

1. Team Creation:

The matchup analysis dashboard for cricket aids team creation by providing insights into player performance and team dynamics. It offers detailed statistics on player performance against specific opponents, helping selectors assemble a balanced team that maximizes success. The dashboard provides historical data on team performance in specific conditions, aiding strategic decisions.

2. Match Prediction:

The dashboard utilizes machine learning and historical matchup data to predict match outcomes. It estimates probabilities based on player performance, team form, pitch conditions, and head-to-head records. Match predictions support decision-making for cricket enthusiasts, analysts, and sports bettors. Visualizations and statistics enhance understanding of influencing factors.

3. Fantasy Team Formation:

The dashboard assists fantasy cricket team managers by providing comprehensive player statistics, historical performance data, and matchup analysis. It aids in optimal player selections and captaincy decisions. Analyzing player performance against specific opponents and conditions gives users an edge. Visualization features aid in comparing player statistics and identifying potential differentiators.

4. Cricket Writing and Journalism:

The dashboard empowers cricket writers and journalists with accurate statistics to support analysis and storytelling. It offers insights into player performance trends, team dynamics, and matchup details. Historical data, head-to-head records, and player statistics add depth and credibility to cricket writing, enhancing journalistic coverage quality.

5. Fan Engagement:
The matchup analysis dashboard enhances fan engagement by providing comprehensive

matchup data, player statistics, and visualizations. Fans can explore historical performance, compare player metrics, and participate in discussions. The interactive platform fosters a sense of community, allowing fans to connect and share their perspectives. Social media integration features encourage interaction and sharing, enhancing the overall fan experience and involvement in the sport.

Deliverables

1. Web scraping module:

- Develop a Python module using Beautiful Soup (bs) that effectively retrieves matchup datafrom various cricket-related sources such as websites, statistical databases, and official cricket APIs.

- Implement robust error handling and data validation mechanisms within the web scraping module to ensure accurate and reliable data extraction.

2. Data processing algorithms:

- Design and implement algorithms to clean and preprocess the extracted matchup data, including data normalization, outlier detection, and handling missing values.

- Utilize machine learning techniques such as clustering and classification algorithms to identify patterns and relationships within the data, enabling the identification of key matchups and player performance trends.

3. Visualization website:

- Create a visually appealing and intuitive website using Next.js, a React framework, to display the extracted matchup data interactively.

- Incorporate responsive design principles to ensure the website is accessible and user-friendly across different devices and screen sizes.

4. Query functionality:

- Develop a user interface that allows users to input specific matchups, such as player vs. player or team vs. team, and retrieve relevant matchup data.

- Implement an efficient search mechanism that enables users to filter and sort the data based on various parameters such as player performance metrics, team statistics, and historical trends.

Potential Obstacles

Several obstacles may arise during the development of this project, including:

- Data availability: Ensuring that the required data for matchup analysis is accessible and obtainable through web scraping.
- Data quality and consistency: Addressing challenges related to inconsistent data formats, missing values, or inaccuracies in the scraped data.
- Web scraping limitations: Overcoming potential roadblocks in web scraping, such as website layout changes, anti-scraping measures, or dynamic content loading.

Project Approval

Suggestions

-

Approved by

POSITION

DD/MM/YYYY

Approved by:

POSITION

DD/MM/YYYY

Week 1

Meetings

- First meeting with Dr Bharathi R on 07/06/2023.
- Two review meetings online with team members.

Progress

- We have successfully managed to come up with a Python code to scrape the player data required for the project using the API provided by ESPN Cricinfo.
- Performed a small study on recommender system including content based and collaborative filtering which will come handy in the upcoming few weeks.

Work to be done for next week

- Perform literature review of related papers and understand the methodology that has been used.
- Build a sound understanding on Bayesian priors and implement a small program on a subset of the dataset obtained in order to make predictions and eventually extend the solution to the entire dataset in future.

Next review meeting with mentor

- Tuesday, 13/06/2023
- Thursday, 15/06/2023

Week 2

Meetings

- 2 Meetings with Dr Bharathi R on 13th and 15th June
- Two review meetings online with team members.

Progress

- We came up with a Python code to scrape the Team format specific data
- Performed literature review of related papers and understood the methodology that has been used.

Work to be done for next week

- We aim to refer a few more suggested papers by Bharathi Ma'am to

understand the mathematics behind it.

- We would like to use the same method they've used in the research papers on our dataset.

Next review meeting with mentor

Wednesday, 21/06/23

Week 3

Meetings

- Offline Meeting with Dr Bharathi R on 21st June.
- Two online meetings with team members.

Progress

- Went through a few research papers to formulate the logic and parameters to be incorporated to our model.
- Successfully secured a ball-to-ball dataset to be used.
- Discussed the various machine learning models that can be utilised.

Work to be done for next week

- Come up with a final formula to be used in our model.
- Make progress on our working solution based on the dataset we procured.

Next review meeting

- Dr Bharathi R, Wednesday, 28/06/23
- Dr Sandesh, (Biweekly review) Friday, 30/06/23