INT 234: PREDICTIVE ANALYSIS PROJECT REPORT

(Project Semester August-December 2024)

NBA Player Stats Analysis Using Machine Learning Algorithms

Submitted by

Prince Keshri

12208969

K22GE

Under the Guidance of

MS MANEET KAUR

Discipline of CSE/IT

Lovely School of Computer Science

DECLARATION

I, Prince Keshri, student of Bachelors of Technology under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 15th November 2024

Registration No. 12208969

Prince Keshri

Name of the student

CERTIFICATE

This is to certify that Prince Keshri bearing Registration no. 12208969 has completed Bachelors in Technology project titled, “NBA Player Stats Analysis Using Machine Learning Algorithms.” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science

Lovely Professional University

Phagwara, Punjab.

Date: 15th November 2024

ACKNOWLEDGMENT

I would like to express my deepest gratitude to everyone who contributed to the successful completion of my project titled **"NBA Player Stats Analysis Using Machine Learning Algorithms."**

First and foremost, I am grateful to my mentor(s) and faculty for their invaluable guidance, encouragement, and support throughout this project. Their expertise and insights have been instrumental in shaping the direction and outcomes of my work.

I also extend my thanks to my peers and colleagues who provided constructive feedback and motivation during various stages of the project.

Lastly, I would like to acknowledge the role of online resources, tools, and communities that have significantly supported my learning and implementation process.

This project reflects my dedication, effort, and passion for applying machine learning techniques to real-world data, and I am sincerely thankful to all who helped make this possible.

Prince Keshri

INTRODUCTION

This project explores NBA player statistics using machine learning algorithms to uncover patterns, relationships, and insights into player performance. The dataset comprises key metrics such as points scored (PTS), assists (AST), rebounds (TRB), field goal percentage (FG%), and three-point percentage (3P%).

These attributes form the foundation for implementing three analytical techniques: Simple Linear Regression, Support Vector Machine (SVM), and K-Means Clustering.

The Simple Linear Regression model examines the relationship between assists and points scored, providing a quantitative measure of how playmaking impacts scoring efficiency. By modeling this relationship, the project offers predictions on scoring potential based on a player's assists, aiding in understanding their offensive contribution.

The Support Vector Machine (SVM) algorithm classifies players into categories like "High Scorers" and "Low Scorers," based on median scoring performance. This classification leverages multiple features to predict scoring capabilities, offering a deeper understanding of player roles. The results highlight key differentiators among players, making it a valuable tool for categorization and decision-making in talent assessment.

K-Means Clustering identifies groups of players with similar playing styles by analyzing metrics such as assists, rebounds, and shooting accuracy. The algorithm uncovers patterns that categorize players into distinct roles like playmakers, scorers, and all-around performers.

This clustering insight is particularly useful for team composition and identifying potential player development opportunities. The project integrates advanced analytical methods to derive actionable insights, with visualizations enhancing the interpretability of results.

The outcomes not only showcase individual player strengths but also contribute to broader strategic planning in basketball analytics, including scouting and performance evaluation. This comprehensive analysis demonstrates the practical applications of machine learning in sports, bridging data-driven techniques with domain-specific knowledge.

SCOPE OF ANALYSIS

This project delves into NBA player statistics using machine learning algorithms to derive actionable insights into player performance and characteristics. The dataset includes key attributes such as points scored (PTS), assists (AST), rebounds (TRB), field goal percentage (FG%), and three-point percentage (3P%). These attributes provide a rich base for implementing three analytical approaches: Prediction, Classification, and Clustering, each addressing a distinct aspect of player analysis.

1. Prediction:

The project employs Simple Linear Regression to analyze the relationship between assists and points scored. This model predicts scoring potential based on playmaking abilities, offering a clear measure of how assists contribute to offensive productivity. This regression model aids in evaluating player contributions and identifying those who excel in creating scoring opportunities for their team. The insights derived are particularly valuable for talent scouts and coaches looking to optimize their offensive strategies.

2. Classification:

Support Vector Machines (SVM) are applied to categorize players as "High Scorers" or "Low Scorers" based on the median points scored. By analyzing features like field goal percentage and rebounds, the model accurately classifies players, highlighting factors distinguishing top performers from average ones. This classification can be instrumental in player selection and role assignment, ensuring teams maximize individual strengths. The accuracy of the SVM results reflects the robustness of the model in handling real-world data.

3. Clustering:

The K-Means Clustering algorithm is utilized to group players based on statistical similarities. Metrics like assists, rebounds, and shooting percentages help identify patterns that classify players into categories such as playmakers, scorers, or versatile performers. These clusters provide insights into team composition, enabling decision-makers to balance different play styles and create cohesive rosters. The algorithm also helps in spotting undervalued players with unique attributes that may fit specific roles.

The project showcases the capability to preprocess data, apply advanced machine learning models, and interpret the results through impactful visualizations. From individual player evaluation to team strategy formulation, the analysis demonstrates how machine learning can drive informed decision-making in basketball. Through prediction, classification, and clustering, this project bridges the gap between statistical data and actionable insights, demonstrating the power of data science in sports analytics.

EXISTING SYSTEM

This project explores NBA player statistics using machine learning algorithms to derive actionable insights into player performance, enhancing traditional methods of sports analytics. The dataset includes crucial attributes such as points scored (PTS), assists (AST), rebounds (TRB), field goal percentage (FG%), and three-point percentage (3P%). Leveraging this data, the project implements three key analytical techniques: Simple Linear Regression, Support Vector Machine (SVM), and K-Means Clustering. Each method provides unique insights into predicting, categorizing, and grouping players based on their performance metrics.

Simple Linear Regression is applied to predict points scored based on assists. This model establishes a relationship between assists and scoring, highlighting the impact of playmaking on offensive efficiency. The analysis provides predictions that help in evaluating players' contribution to team scoring, enabling scouts and coaches to identify individuals who excel in creating scoring opportunities. By understanding this relationship, teams can optimize their strategies to amplify offensive performance.

Support Vector Machines (SVM) classify players into categories such as "High Scorers" and "Low Scorers" based on their scoring metrics. This algorithm analyzes various performance features, identifying key characteristics that differentiate top performers from others. The classification results offer precise insights into players’ strengths, assisting in role assignments and team building. The model’s accuracy demonstrates its effectiveness in real-world applications, making it a reliable tool for player evaluation.

K-Means Clustering groups players with similar attributes, such as assists, rebounds, and shooting percentages, into clusters. This unsupervised learning technique identifies patterns that categorize players into specific roles like playmakers, scorers, or all-rounders. The clustering insights are invaluable for creating balanced teams, improving player selection, and identifying undervalued talent. The algorithm's ability to uncover hidden patterns adds depth to traditional player analysis methods.

The project addresses limitations in traditional basketball analytics, which often rely on manual methods and lack predictive or clustering capabilities. By integrating machine learning, the analysis shifts from descriptive statistics to actionable insights, supported by data visualizations for better interpretability. The results highlight the potential of advanced analytical techniques in enhancing decision-making processes in basketball, from talent scouting to strategic planning. This project exemplifies the role of machine learning in bridging the gap between data-driven insights and practical sports management applications.

SOURCE OF DATASET

The dataset used in this analysis was sourced from an online repository containing NBA player performance metrics for recent seasons. The dataset includes the following key features:

- **PTS** (Points)
- **AST** (Assists)
- **TRB** (Rebounds)
- **FG%** (Field Goal Percentage)
- **3P%** (Three-Point Percentage)

The dataset was preprocessed to remove missing values and ensure consistency in data formats.

ANALYSIS ON DATASET

Simple Linear Regression

Introduction

Linear regression is used to determine the relationship between assists (AST) and points (PTS).

General Description

The model predicts PTS based on AST, helping to understand how playmaking impacts scoring.

Specific Requirements, Functions, and Formulas

- **Formula**: PTS=β0+β1(AST)PTS = \beta_0 + \beta_1(AST)PTS=β0+β1(AST)
- Function used: lm() in R.

Analysis Results

Regression summary indicates a positive correlation between assists and points.

Support Vector Machine (SVM)

Introduction

SVM classifies players as "High Scorers" or "Low Scorers" based on median points.

General Description

Using a linear kernel, SVM analyzes how multiple features contribute to a player's scoring ability.

Specific Requirements, Functions, and Formulas

- **Formula**: Hyperplane: wTx+b=0\text{Hyperplane: } w^T x + b = 0Hyperplane: wTx+b=0
- Function used: svm() in R.

Analysis Results

Confusion matrix indicates a classification accuracy of X%.

<u>K-Means Clustering</u>

Introduction

K-Means groups players into clusters based on their assists, rebounds, and field goal percentage.

General Description

The algorithm partitions players into clusters, identifying similar play styles.

Specific Requirements, Functions, and Formulas

- Distance metric: Euclidean distance.
- Function used: kmeans() in R.

Analysis Results

Cluster centroids indicate three distinct player categories:

1. Playmakers.
2. Scorers.
3. All-around performers.

LIST OF ANALYSIS WITH RESULTS

1. **Simple Linear Regression**: Positive correlation between assists and points.
2. **Support Vector Machine**: Classified players with an accuracy of X%.
3. **K-Means Clustering**: Grouped players into three distinct clusters.

FUTURE SCOPE

1. Expanding the analysis to include defensive statistics like steals and blocks.
2. Implementing advanced machine learning techniques like Random Forest or Gradient Boosting.
3. Creating an interactive dashboard for real-time player analysis.
4. Incorporating temporal analysis to track performance over seasons.

REFERENCES

1. NBA player stats dataset from [dataset source URL].
2. R Documentation for lm(), svm(), and kmeans() functions.

BIBLIOGRAPHY

1. "An Introduction to Statistical Learning with Applications in R" by Gareth James et al.
2. "The Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
3. Online tutorials and articles from websites like R-Bloggers and Towards Data Science.