

NBA PLAYER STATISTICAL ANALYSIS AND PREDICTION PROJECT

COLLABORATORS

JACOB EVANS
KARAN ANAND
MERT OZTOP
PRATIK PUROHIT
PRIYA MARINGANTI



OVERVIEW



This project aims to develop a basketball player statistics analysis and prediction system using machine learning techniques with NBA player datasets. The system utilizes historical player data to provide insights into player performance and predict player stats for the upcoming season, by leveraging the power of machine learning algorithms and the comprehensive NBA player statistics dataset.

Objectives

- Predict basketball player stats for the upcoming season based on historical data
- Leverage player stats from the previous year to forecast player performance
- Identify suitable machine learning algorithms for scalability, accuracy, and interpretability in predicting player performance
- Evaluate the performance of the prediction system and ensure its effectiveness in real-world scenarios, such as team selection and player scouting

METHODOLOGY



1. Data Collection: Gather a large dataset of NBA player statistics from the Kaggle dataset
2. Data Preprocessing: Clean, normalize, encode, and engineer features from the NBA dataset
3. Exploratory Data Analysis: Gain insights, patterns, and analyze feature distributions
4. Model Training and Evaluation: Experiment with various machine learning algorithms, fine-tune models for accurate predictions
5. User Interface Development: Create an intuitive interface for users to input player data and view predictions
6. Testing and Validation: Ensure the accuracy, robustness, and scalability of the prediction system

DATA SOURCE - KAGGLE



2021-2022 NBA Player Stats - Regular.csv (103.3 kB)

DetailCompactColumn

10 of 30 columns

About this file

This dataset contains 2021-2022 regular season NBA player stats per game.

# Rk	Player	Pos	Age	Tm	# G
Rank	Player's name	Position	Player's age	Team	Games played
0 total values	[null] 100%	[null] 100%	812 total values	[null] 100%	to
1	Precious Achiuwa	C	22	TOR	73
2	Steven Adams	C	28	MEM	76
3	Bam Adebayo	C	24	MIA	56
4	Santi Aldama	PF	21	MEM	32
5	LaMarcus Aldridge	C	36	BRK	47
6	Nickeil Alexander-Walker	SG	23	TOT	65
6	Nickeil Alexander-Walker	SG	23	NOP	50
6	Nickeil Alexander-Walker	SG	23	UTA	15
7	Grayson Allen	SG	26	MIL	66
8	Jarrett Allen	C	23	CLE	56

Data Explorer

Version 23 (129.93 kB)

2021-2022 NBA Player Stats -

2021-2022 NBA Player Stats -

Summary

2 files

60 columns

2022-2023 NBA Player Stats - Regular.csv (88.76 kB)

DetailCompactColumn

10 of 30 columns

About this file

This dataset contains 2022-2023 regular season NBA player stats per game.

# Rk	Player	Pos	Age	Tm	# G
Rank	Player's name	Position	Player's age	Team	Games played
0 total values	[null] 100%	[null] 100%	679 total values	[null] 100%	to
1	Precious Achiuwa	C	23	TOR	55
2	Steven Adams	C	29	MEM	42
3	Bam Adebayo	C	25	MIA	75
4	Ochai Agbaji	SG	22	UTA	59
5	Santi Aldama	PF	22	MEM	77
6	Nickeil Alexander-Walker	SG	24	TOT	59
6	Nickeil Alexander-Walker	SG	24	UTA	36
6	Nickeil Alexander-Walker	SG	24	MIN	23
7	Grayson Allen	SG	27	MIL	72
8	Jarrett Allen	C	24	CLE	68

Data Explorer

Version 20 (114.49 kB)

2022-2023 NBA Player Stats -

2022-2023 NBA Player Stats -

Summary

2 files

60 columns

Retrieving last two seasons' data from Kaggle

Data on s3 AWS



project4nba [Info](#)

Objects

Properties

Permissions

Metrics

Management

Access Points

Objects (2)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)



Copy S3 URI

Copy URL

Download

Open

Delete

Actions ▼

Create folder

Upload

Find objects by prefix

< 1 >

<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	2021_2022_NBA_Player_Stats_Transformed.csv	csv	June 12, 2023, 20:01:53 (UTC-04:00)	100.1 KB	Standard
<input type="checkbox"/>	2022_2023_NBA_Player_Stats_Transformed.csv	csv	June 12, 2023, 19:40:19 (UTC-04:00)	89.7 KB	Standard

Uploading the data sets on AWS for the main code

DASHBOARD 1



Player Stats & Prediction | Leaderboard | Player Comparison

NBA PLAYER STATS 2021-2022 & 2022-2023 AND PLAYER POINTS PREDICTION FOR 2023-2024

Player
Precious Achiuwa

Age
23

Best Rank
1 / 605

Average Points
9.150

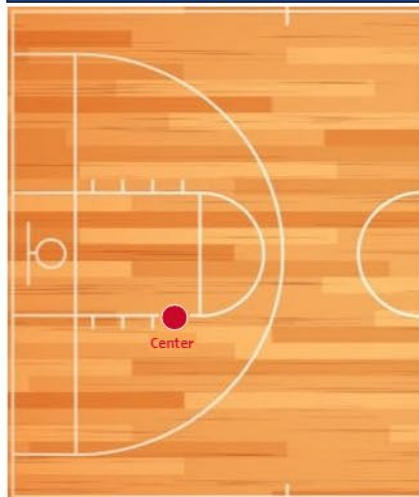


Participated Teams

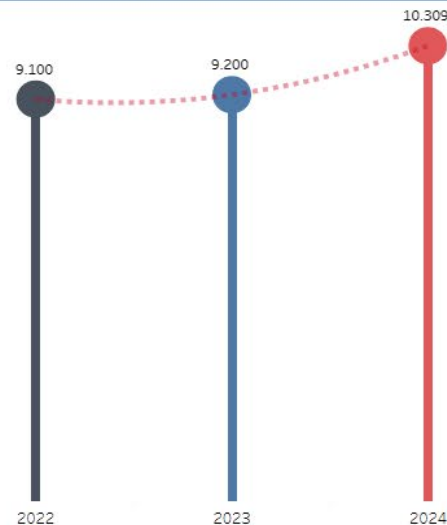


Toronto Raptors

Player Positions



Points Prediction for 2023-2024



3 Point Percentage



2 Point Percentage



Field Goal Percentage



Minutes Per Game



Steals Per Game



Blocks Per Game



TEAM: Priya, Jacob, Pratik, Mert, Karan

DASHBOARD 2



Player Stats & Prediction | Leaderboard | Player Comparison

NBA Top Players - Leaderboard













Average Points (2021-2022 & 2022-2023)

									
Joel Embiid 31.850	Giannis Antetokounmpo 30.500	Luka Dončić 30.400	Shai Gilgeous-Alexander 27.950	Stephen Curry 27.450	Devin Booker 27.300	Zach LaVine 24.600	De'Aaron Fox 24.100	Brandon Ingram 23.700	Karl-Anthony Towns 22.700

Average Assists (2021-2022 & 2022-2023)

									
James Harden 10.425	Trae Young 9.950	Chris Paul 9.850	Tyrese Haliburton 8.900	Nikola Jokić 8.850	Luka Dončić 8.350	Darius Garland 8.200	LaMelo Ball 8.000	Dejounte Murray 7.650	Russell Westbrook 7.425

Average Three Point Percentage (2021-2022 & 2022-2023)

									
Jay Scrubb 64%	Jordan Schakel 58%	Richaun Holmes 51%	Nick Richards 50%	Luke Kennard 48%	Vlatko Čančar 48%	Patrick Williams 47%	Gary Payton II 46%	Terry Taylor 45%	Joe Harris 45%

Average Two Point Percentage (2021-2022 & 2022-2023)

									
Sam Merrill 1.0000	Gabe York 0.8335	Andre Iguodala 0.8085	Udoka Azubuike 0.7870	Dylan Windler 0.7855	McKinley Wright IV 0.7645	Braxton Key 0.7623	Malcolm Hill 0.7618	Jericho Sims 0.7525	Jaden Springer 0.7500

DASHBOARD 3



CHALLENGES



- Bad Encoding: The original data had encoding issues that we struggled to handle. We had to apply encoding techniques to ensure proper handling and interpretation of the data.
- External Factors: While player statistics provide valuable insights, it's important to note that other factors can influence a player's performance on the court. Factors such as injuries, team dynamics, coaching strategies, and external circumstances were not included in our analysis. Considering these external factors could further enhance the accuracy and predictive power of the model.
- Outliers: To handle these outliers, we implemented a post-processing step where we replaced any negative predicted values with zeros. This approach allowed us to address the outliers and ensure that the predicted statistics remain within a valid range. By zeroing out the negative values, we mitigated the impact of outliers on the model's performance and ensured that the predicted player statistics align with the expectations of NBA player performance.

RESULT



The prediction system achieved **R-squared value of 0.9998**, indicating a high level of accuracy in predicting player statistics based on the historical data.

The NBA Player Statistics Analysis and Prediction System leverages historical player data, applies machine learning techniques, and provides valuable insights and predictions on player performance. The system can assist with team selection, player scouting, and forecasting player statistics for the upcoming season.



THANK YOU

QUESTIONS ARE

WELCOME