Predicting NBA Playoff Participation and Championship Wins

Overview

This project aims to predict NBA playoff participation and championship wins based on regular season statistics. By leveraging historical game data, we have developed machine learning models that provide insights into a team's chances of making the playoffs and winning the championship.

Data Collection and Preparation

I collected game data from the NBA for both regular season and playoff games. The data includes various performance metrics such as field goal percentage, three-point percentage, free throw percentage, points scored, and more.

1. **Data Source**: NBA game data stored in a SQLite database obtained from Kaggle.com (NBA Database from Wyatt Walsh)

2. **Data Processing**: The data was fetched, cleaned, and prepared for analysis using Python.

Methodology

1. **Data Extraction:**
    - o Regular season and playoff game data were extracted from the database.
    - o A dataset was created with labels indicating playoff participation and championship wins.

2. **Data Visualization:**
    - o The data parsing and visualization script parsed data from the tables of average statistics.
    - o Informative plot graphs were created to visualize team performance and clustering based on regular season statistics using PCA (Principal Component Analysis).
    - o Clusters of teams were identified, and the plots were annotated with team names to provide a clear visualization of their performance.

3. **Feature Engineering:**
    - o Features such as average field goal percentage, three-point percentage, free throw percentage, points scored, and other metrics were calculated for each team.

4. **Model Training:**
    - o We trained two classification models: one to predict playoff participation and another to predict championship wins.
    - o We used Logistic Regression for playoff prediction and Random Forest for championship prediction.

- To address class imbalance, we applied the SMOTE (Synthetic Minority Over-sampling Technique) to oversample the minority classes.

5. **Model Evaluation:**
   - The models were evaluated using accuracy and classification reports.
   - We ensured that undefined precision and recall values were handled appropriately using the `zero_division` parameter.

Results

The models demonstrated good performance in predicting playoff participation and championship wins. Below are the key metrics from the model evaluation:

1. **Playoff Participation Model:**
   - **Accuracy**: 68.25%
   - **Classification Report**:

	Precision	Recall	F-1 Score	Support
0	0.01	0.67	0.02	48
1	1.00	0.68	0.81	8038
Accuracy			0.68	8086
Macro Avg	0.50	0.67	0.42	8086
Weighted Avg	0.99	0.68	0.81	8086

2. **Championship Win Model:**
   - **Accuracy**: 98.33%
   - **Classification Report**:

 | | Precision | Recall | F-1 Score | Support |
     |---|---|---|---|---|
     | **0** | 0.00 | 0.00 | 0.00 | 130 |
     | **1** | 0.99 | 0.68 | 0.81 | 8006 |
     | | | | | |
     | **Accuracy** | | | 0.98 | 8086 |
     | **Macro Avg** | 0.49 | 0.50 | 0.50 | 8086 |
     | **Weighted Avg** | 0.97 | 0.98 | 0.98 | 8086 |

Insights and Predictions

Using the trained models, we can predict the probability of playoff participation and championship wins for teams based on their regular season statistics. For example, given the regular season stats for a new team, the models can output:

- **Playoff Probability**: 99.99%
- **Championship Win Probability**: 41.00%

Visualization

To visualize the clustering of teams based on their regular season stats, we performed PCA (Principal Component Analysis) and plotted the results. Teams were clustered into groups, and the plot annotated with team names provides a clear visualization of their performance. *Blue names denote participants in the playoffs and red name denotes the winner of the playoffs*.
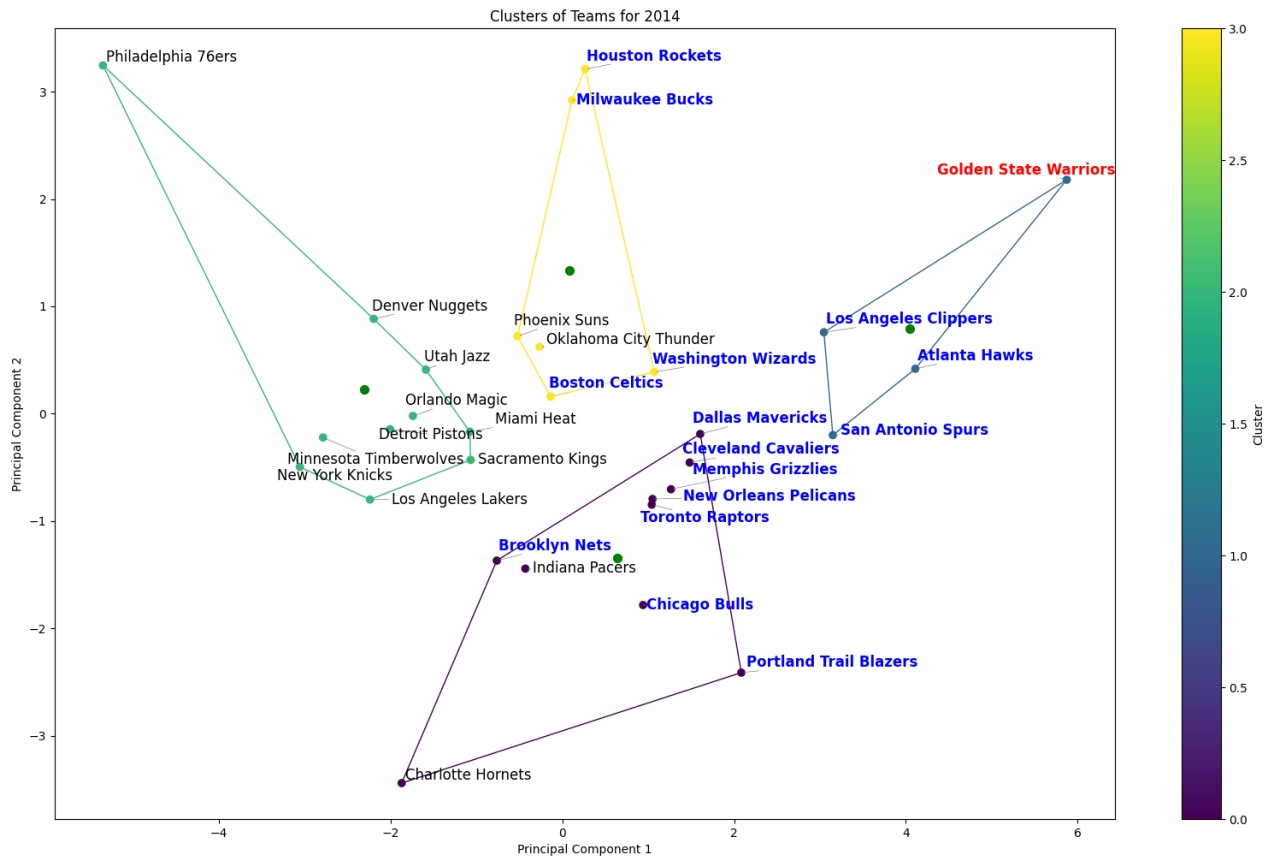
Code Implementation

The project involved three scripts:

1. **Data Extraction and Preparation Script:**
   - Fetches game data from the database.
   - Creates tables with average statistics for regular season and playoff teams.
2. **Data Parsing and Visualization Script:**
   - Parses data from the tables of average statistics created in the data extraction and preparation script.
   - Creates informative plot graphs to visualize team performance and clustering based on regular season statistics.
3. **Model Training and Prediction Script:**
   - Trains classification models to predict playoff participation and championship wins.
   - Evaluates model performance.
   - Provides a function to predict probabilities for new data.

Conclusion

This project successfully demonstrates the use of machine learning models to predict NBA playoff participation and championship wins based on regular season statistics. The insights and predictions generated can be valuable for teams, analysts, and fans alike.

Code Repository

The full code for this project, including data extraction, model training, and prediction scripts, is available in my GitHub here. Feel free to explore the code and use it as a reference for your own data analysis projects.

Contact

For more information or to discuss this project, please contact me at rmnesbitt@gmail.com