Project Hoopster

This project aims to leverage advanced data analytics and Artificial Intelligence (AI) to predict various events and outcomes within the National Basketball Association (NBA). We will focus on developing predictive models for diverse scenarios, including game outcomes, individual player matchups, shot success, and the progression and results of playoff series and championships. The insights generated from this project are expected to be highly valuable for applications in sports betting, strategic coaching decisions, and in-depth media analytics. We have secured access to comprehensive historical NBA datasets to build robust and realistic models.

Challenging Features:

The core problem we seek to address is: **"Can Artificial Intelligence accurately predict specific NBA events such as game outcomes, individual player performance in head-to-head scenarios (1v1), shot-making probability, and the overall progression and winner of playoff series and championships?"**

Traditional sports analysis often relies on human intuition, basic statistical averages, and expert opinions. Our goal is to move beyond these methods by creating sophisticated AI models that can uncover deeper patterns and provide more precise and actionable predictions, offering a significant edge in various real-world applications.

A significant challenge and core aspect of this project is the development of distinct AI models tailored for each head-to-head event prediction:

- **Game Outcomes:** Predicting winners and score differentials.
- **1v1 Player Matchups:** Analyzing individual player efficiency and impact in direct confrontations.
- **Shot Making:** Predicting the probability of a shot being successful based on context (player, location, defender, game state).
- **Playoff Series Progression:** Forecasting which team will advance through each round of the playoffs.
- **Championship Winner:** Predicting the ultimate NBA champion.

Each of these events requires a unique feature set, model architecture, and evaluation methodology to achieve high accuracy.

Useful Features:

The successful implementation of these predictive models will have broad utility:

- **Sports Betting:** Providing data-driven insights and probabilities to inform betting strategies.
- **Coaching:** Offering analytical tools for game planning, opponent scouting, player rotation optimization, and real-time decision-making.
- **Media Analytics:** Enhancing sports commentary, providing deeper insights for broadcasts, and creating engaging content for fans.

Realness:

The project's feasibility is bolstered by the availability of rich, detailed, and relevant datasets. The two Kaggle datasets provide a strong historical foundation for model training and validation, covering a vast range of seasons, games, and player metrics. The SportsData.io API offers the potential for integrating real-time data, ensuring the models can be applied to current scenarios and not just historical analysis.

Dataset 1: NBA Games, Players, Box Scores, and Play-by-Play Data (from Kaggle)

- **Source:** https://www.kaggle.com/datasets/wyattowalsh/basketball?select=nba.sqlite
- **Content:** This extensive dataset provides comprehensive historical NBA information, including:
    - Data for all 30 NBA teams.
    - Profiles for over 4,800 players.
    - Records of over 65,000 games, covering every game since the inaugural 1946-47 NBA season.
    - Detailed Box Scores for over 95% of all games, offering traditional statistical aggregates.
    - Granular Play-by-Play game data, with over 13 million rows, providing sequential event information within games.
- **Utility:** This dataset will be crucial for training models on game outcomes, player performance within games, and potentially shot-making analysis due to the detailed play-by-play data.

Dataset 2: NBA Draft Players Stats (Cleaned Dataset from Kaggle)

- **Source:** https://www.kaggle.com/datasets/joelesteban/nba-draft-players-stats-clean-dataset
- **Content:** This dataset offers a complete overview of NBA Draft picks across various years. It includes:
    - Player statistics.
    - Physical attributes (e.g., height, weight).
    - Draft positions.

    - Team selections.
- **Utility:** This dataset is ideal for analyzing historical trends related to player development, scouting, and performance forecasting. It can inform features for predicting long-term player value or potential.

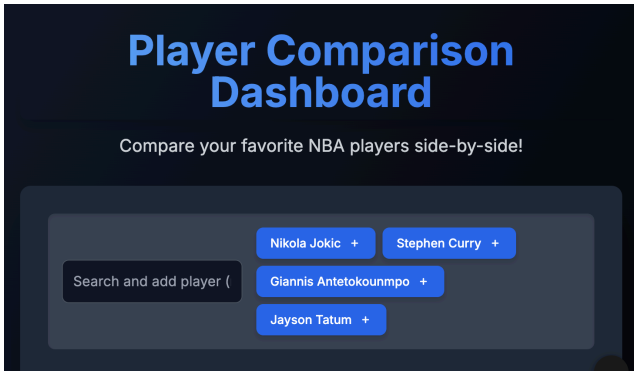
Dataset 3: SportsData.io NBA API

- **Source:** `https://sportsdata.io/nba-api`
- **Content:** Provides access to real-time and historical NBA data, including live scores, schedules, odds, and detailed statistics.
- **Utility:** While the Kaggle datasets provide static historical data, this API will be invaluable for obtaining up-to-date information for live predictions, validating models against current events, and potentially expanding features with dynamic data points.

Functionality (UI/UX Component)

A key user-facing component of this project will be an intuitive dashboard to present insights and enable user interaction. The initial focus for the UI will be a **Player Comparison Dashboard**:

- **Concept:** This dashboard will allow users to dynamically select and compare 2-5 NBA players side-by-side, offering a comprehensive view of their statistical profiles.
- **Elements:**
    - **Radar Charts:** Visual representations of players' skill profiles (e.g., shooting, passing, defense, rebounding, athleticism).
    - **Bar Charts:** Side-by-side comparisons of traditional and advanced per-game or per-36 minute statistics.
    - **Customizable Metric Selection:** Users can choose which specific statistics they want to compare.
    - **Head-to-Head Records (Future Enhancement):** Data on how selected players have performed directly against each other in past games.
- **Interaction:**
    - **Search and Add Players:** An intuitive search functionality to easily find and add players to the comparison list.
    - **Toggle Stat Types:** Ability to switch between different categories of statistics (e.g., basic box score stats, efficiency metrics, impact stats).



Work Distribution:

Ananth Ramani (ramani3):Back End/ API Management

- Design and implement backend architecture.
- Develop APIs to serve data and prediction results to frontend.
- Integrate SportsData.io API.
- Collaborate on database schema design and integration.

Devam Shah (dshah97): Back End/ Machine Learning

- Development and optimization of machine learning models for:
    - Game outcomes
    - 1v1 matchups
    - Probability of making shots
    - Playoff and championship predictions
- Support backend integration of ML models.

Sam Chadri (chadri2): Front End UI/Dev Ops (if necessary)

- **Infrastructure Provisioning:** Design, provision, and manage the cloud or on-premise infrastructure required for the project (servers, databases, networking, etc.).
- **Environment Management:** Manage development, staging, and production environments.
- **Containerization & Orchestration:** (e.g., Docker, Kubernetes) for consistent and scalable deployments.
- **UI/UX:** Collaborate on User Interface Design implemented either using React, Vue, or other Javascript frameworks.

Javen Boettcher (javencb2): Front End/UI Development

- Collaboration with Sam on frontend development with a focus on clean design, user experience, and overall polish.
- Assist in building and managing AWS deployment and integration.
- Support UI error handling, testing, and optimization.
- Assist with data preprocessing to prepare data for visual analytics.

Shared Responsibilities:

- **Database Design and Integration:** All members participate in planning and implementing a robust database.
- **Documentation and Presentation:** All members work together in writing technical and user documentation.