

# Comprehensive Guide to Building Specifications for NBA Shot Analysis Project

June 10, 2024

## 1 Executive Summary

- **Business Problem:** Incorporating advanced statistical analysis to stay competitive in the NBA by leveraging digital tracking tools to gather data on player performance, focusing on shot frequency and shooting efficiency.
- **Objectives:** Deploy a predictive model for shot success to optimize player performance and inform strategic decisions, incorporating automated maintenance, performance monitoring, and iterative improvement aligned with MLOps practices.
- **Scope:** Focus on historical shot data from the past 23 NBA seasons.

## 2 Detailed Specification

### 2.1 Business Requirements

Incorporating advanced statistical analysis into our operations is crucial for staying competitive in the realm of American sports, particularly in the NBA. As technology continues to evolve, we recognize the importance of leveraging digital tracking tools to gather comprehensive data on player performance. By focusing on shot frequency and shooting efficiency among top NBA players, as identified by ESPN, we aim to gain valuable insights that can inform our strategic decisions and enhance our team's performance on the court. This project aligns with our commitment to utilizing data-driven approaches to drive success in the highly competitive landscape of professional basketball.

#### 2.1.1 Technical Perspective:

This project entails harnessing cutting-edge digital tracking tools to capture real-time data on player movements during NBA games. We will employ sophisticated statistical analysis techniques to process the vast amount of information collected, focusing on shot frequency and shooting efficiency across various game situations and court locations. Additionally, predictive modeling techniques will be utilized to estimate the probability of shot success for each of the 20 esteemed NBA players included in the dataset. By leveraging advanced analytics and machine learning algorithms, we aim to extract actionable insights that can optimize player performance and inform strategic decision-making within our organization.

### 2.1.2 Economic Perspective:

Investing in this project holds significant economic potential for our organization within the sports industry. By leveraging data analytics to understand player performance and optimize strategic decision-making, we can gain a competitive edge on the court. Improved shot selection and efficiency among our players can lead to increased game success, fan engagement, and ultimately, revenue generation through ticket sales, merchandise, and sponsorships. Additionally, by demonstrating our commitment to data-driven approaches, we enhance our brand reputation and attract top talent, further bolstering our long-term economic viability and success in the market.

### 2.1.3 Scientific Perspective:

This project represents an opportunity to advance the scientific understanding of basketball performance through rigorous data analysis and predictive modeling. By examining shot data from top NBA players in various game contexts, we aim to uncover underlying patterns and trends that contribute to successful outcomes on the court. Through the application of statistical methods and machine learning algorithms, we can identify key factors influencing shot selection and effectiveness, contributing to the broader body of knowledge in sports analytics. Moreover, the insights gained from this research have the potential to inform coaching strategies, player development programs, and future scientific inquiries into sports performance optimization.

## 2.2 Functional and Non-Functional Requirements

### 2.2.1 Functional Requirements:

- The system should predict the likelihood of a shot being successful based on historical shot data.
- The system should allow users to input new shot data and receive predictions in real-time.

### 2.2.2 Non-Functional Requirements:

- **Accuracy:** The model should achieve an accuracy of at least 60%.
- **Performance:** The system should handle up to 1,000 concurrent users without performance degradation.
- **Usability:** The user interface should be intuitive and easy to navigate for both technical and non-technical users.
- **Reliability:** The system should have an uptime of 99.9%.
- **Scalability:** The system should be able to scale to accommodate increasing amounts of data and users.

## 2.3 Data Requirements

- **Data Storage and Management:** Ensure a robust data storage and management system capable of handling large volumes of data. This includes storing historical data for model training, as well as new incoming data for retraining.
- **Data Preprocessing Pipeline:** Develop an automated data preprocessing pipeline to prepare new incoming data for model retraining. This includes cleaning, transforming, and standardizing the data to ensure consistency and compatibility with the model.
- **Model Retraining Strategy:** Define a strategy for periodically retraining the model using new incoming data. This ensures that the model remains up-to-date and continues to provide

accurate predictions over time.

- **Model Versioning and Deployment:** Establish a system for versioning and deploying trained models to production. This enables seamless integration of updated models while maintaining stability and reliability of the production environment.

## 2.4 Model Objectives and Outputs

- **Objective:** Predict shot success.
- **Outputs:** Probability score for each shot.

## 2.5 System Architecture

- **Data Ingestion Pipeline:** Data from Kaggle (<https://www.kaggle.com/datasets/jonathangmwl/nba-shot-locations>) is collected, cleaned, transformed into a standard format, splited into training and testing sets and loaded into a separate joblib file. This ensures high-quality, consistent data for analysis.
- **Model Training and Deployment:** A random forest model is trained to predict shot success probability and deployed in Docker containers. The model processes real-time data inputs and provides predictions.
- **User Interface:** The UI, (*should be discussed with the team*), allows users to input new shot data and view real-time predictions.

## 2.6 Evaluation Metrics

- **System performance:** Response time, throughput.

[ ]: