

ShotIQ: Predicting NBA Shot Success

By Ryan Davis and Vishnuvardhan Reddy Kanchi

I pledge my Honor that I have abided by the Stevens Honor System

This project explores the intersect between sports and data analytics using machine learning in Python. To do so, a predictive model will be developed in Python to estimate the probability of a basketball shot being made based on historical NBA data. Basketball shot selection and efficiency are essential to success in the sport, and such modeling and predictions can be supportive for coaches, scouts, and fans of the sport to analyze in-game decision making.

The overall objectives of the project are to develop a machine learning model that predicts whether a shot is successful based on the player, shot location, and the frequency at which the player makes that type of shot. I aim to visualize the tendencies of player on a plot that displays a basketball court. Probabilities will be calculated based on the user's interactions with the shots chart. Users will be able to select shot locations with the interactive map and will receive a predicted shot probability from the model. Ultimately, this provides insights to users on how different players perform in different locations on the court.

The tools and methodologies used in this project all involve data science tools for data analysis. The dataset is provided publicly by Kaggle, titled NBA Shots Dataset (2000 - Present), (<https://www.kaggle.com/datasets/techbaron13/nba-shots-dataset-2001-present?resource=download>). The IDE used to implement the model will be Jupyter Notebook running Python. The essential libraries for data manipulation and creating the model includes: pandas, scikit-learn, matplotlib, seaborn, plotly, and NumPy. For modeling the dataset, I will use a Random Forest classifier to train the predictive model on the Sklearn Python Library, utilizing the shot locations (x, y), distance, player name, and shot_type data fields within the dataset to train the model. The dataset of 4,449 csv files will successfully model players' shots as there is more than enough data to predict correlations. Categorical variables like shot_type and player will allow for numerical analysis within the model. The court plot for the user will be generated by matplotlib to visual a player's makes and misses.

With this project, I expect to create this model to better understand the choices that NBA players make in different game situations. Using what I've learned in previous courses regarding machine learning models, and what I've learned in this course about Python, I want to create a meaningful model for basketball fans to access. With an interactive

basketball court interface where users can select different shot selections, and visual shot charts identifying a players' strengths and weaknesses, I hope to allow users to create potential discoveries about a player's potential successes.