**The Evolution of Shooting a Basketball in the NBA using Machine Learning Techniques**

Anthony Chan

Ryerson University, Chang School of Continuing Education

CIND820 DJ0- Big Data Analytics Project

Student Number: 500766830

Supervisor’s Name: TBA

January 23, 2022

**Abstract**

Data analytics have impacted the sporting industry greatly since we can now analyze and predict what a player’s potential can be. Many of these statistics and factors can be translated to what we can see developing in the game. A game where each team can optimize their strategies to win. In this capstone project, the theme will be revolting around predictive analysis and knowledge discovery about the evolution of shooting the basketball within the NBA (National Basketball Association). The dataset used for this project is composed of statistics of NBA players and teams from the years, 1946 to 2016. However, I will be focusing on the players within the year of 2015 to 2016 as this year was known to be one of the greatest seasons of all time. Many factors can go into winning a championship in the NBA, for example the 3-Point Field Goal Percentage, Field Goal Percentage, Free Throw Percentage, Turnovers, Offensive and Defensive Rebounds, and many more statistics.

The model that I will be creating will use the data provided to simulate the how the players in the NBA are shooting in the modern era and compare them to an older generation of basketball. This will allow us to find out the best performing model and use it as a basis for our study. Next, the model will attempt to predict how players will adapt to this new playstyle based on trends found which will help create insight to what teams and coaches should prioritize. This project would be able to investigate the problems: How important is the 3-point line? Should teams focus on teaching players to shoot 3-pointers? What factors could have led the league into this transition? Is there a relationship between high 3-point percentages and winning? I will be using techniques learned throughout this certificate such as linear regression, line of best fit, classification methods, train-test sets, clustering, decision trees, and predictive modeling. I will be implementing my analysis with software such as the Python language in our Jupyter notebook where I will utilize packages such as panda and R where I will use libraries such as ggplot to help visualize my findings.

In conclusion, this capstone project will dive into the variables that take place in every NBA basketball game and learn about how they impact the players around the organization.

**References**

My Github where I will be uploading all code can be found here:

<https://github.com/anthonyychan/cind820>

This dataset can be found on Figshare at:

Wen, R. (2017, September 17). NBA Data. figshare. Retrieved January 23, 2022, from

<https://figshare.com/articles/dataset/NBA_data/5414170>