Abe Woycke
09/26/20

**NBA Projections Project Proposal**

**Problem Statement**

Based on each team's statistics entering a game, which NBA team will win the game?

**Context**

The hypothetical client for this project is a sportsbook looking to maximize its earnings when setting its betting odds for public consumption.

**Criteria for Success**

Compared with previous research, an overall accuracy of over 70% would be very successful. To prevent overfitting, this would be checked on 20% of the data set held out as a testing set. AUROC will be a valuable metric in model selection.

**Scope of Solution Space, Deliverables**

The solution space will be limited to predicting the classification of NBA game outcomes.

Deliverables, all of which will be contained in a GitHub repo, will include:

- model of NBA game outcome predictions based on specific pre-game inputs
- project report detailing conclusions and the process that lead to those conclusions
- slide deck presenting forecast methods and results

**Stakeholders**

Sportsbook Executive Team, who will make final decisions based on my recommendations.

**Data Sources**

This Kaggle dataset contains information on NBA games over the last 15 years.
https://www.kaggle.com/nathanlauga/nba-games
This data is in CSV format.

**Outline of Problem Solving Steps**

**1. Problem Identification**

a. making this project proposal
b. identifying the specific scope of the project

**2. Data Wrangling**

a. compile individual game statistics to recreate snapshot of team statistics entering each game
b. organize data into one cohesive DataFrame
c. clean data to deal with "na" entries

d. validate data to make sure everything has been compiled correctly
e. make sure data definitions within dataset are clear

**3. Exploratory Data Analysis**

a. create plots and charts to look for relationships within dataset
b. see which features of the dataset might have surprising relationships

**4. Pre-Processing and Training Data Development**

a. standardize dataset
b. split into training data for modeling

**5. Modeling**

a. select various potential models, included neural network
b. check performance metrics of potential models

**6. Documentation**

a. create a project report detailing process and conclusions
b. create a slide deck presenting projections results and the methodology therein