Springboard: Data Science Career Track

Capstone 3: Project Ideas

Patrick Long and Co.

Summary: We have been contracted by "The client" (a basketball organization), which is interested in machine learning models they can use to predict a how their basketball team will do overall in the season. This client will like want to know if their team will win, go to the playoffs, and what player statistics will help them do so. Essentially creating a 'scouting' machine learning model.

1.**Context**: The Client is interested in the following

    a. Machine learning models they can use to estimate if their team will win X amount of games and make the playoffs. They will want to know this as soon as possible, so they can know what to do to help prepare their team.

    b. A list of variables that are deemed important to owning a winning basketball team that will make the playoffs. These variables will be deemed predictive. The modeling will mostly be used with Player stats and Team stats. To determine what player stats have helped make winning teams in the past.

    c. The client will want to own the modeling.

2. **Criteria for Success:** This is what the client is expecting from this Project

    a. A working model that shows that will predict if a Basketball team is going to make the playoffs, and how many games they can expect to win give their statistics.

    b. The services provided will be based on a general basis, however, will be deemed successful for the entirety of the season.

    c. They will be able to replicate the project for upcoming seasons, even if contract is not extended.

3. Scope of the Solution Space: The project will only use publicly accessible data. The data will use 'box score' data.

4. **Constraints of the Solution Space:**

    a. The model may need to be retrained after each season, or game. with someone manually inputting data. The model will have to

    b. There will need to be someone monitoring the data input, to make sure it is accurate.

5. Stakeholders to provide Key Insights: The Client CEO "Mr. Bossman" is the advocate for this project for this NBA team.

6. **Required Data Sources:**

    a. We will get the data from https://figshare.com/articles/dataset/NBA_data/5414170, and https://www.sports-reference.com/ as part of the service we will offer to the client for the duration of this pilot project.

    b. There may be possible other data used, which we will obtain and amend.

7. **Anticipated Data Science Approaches to be used:**

    a. This project will be modeled as a binary classification problem, to estimate the probability of a particular basketball team winning or losing a game.

    b. Important elements:

        i. Features of the model

        ii. Data to be used/combined

        iii. How to implement the project based on time period of the season

    c. Models need to be retained daily