Machine Learning Project: Classification with Python
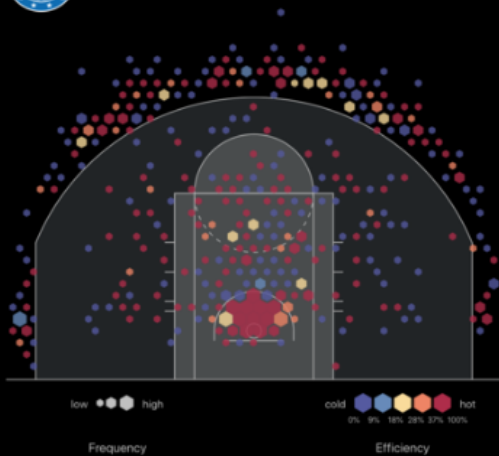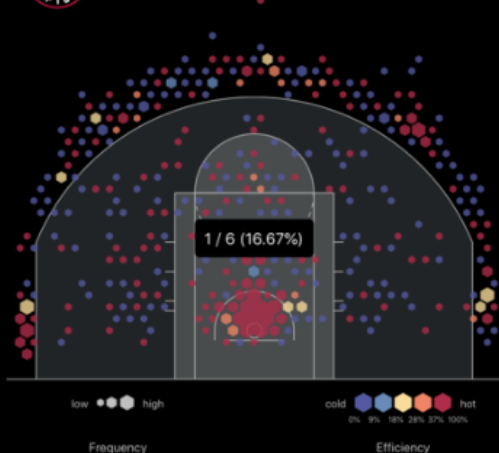

Philadelphia and Toronto playoffs frequency and efficiency 2018-2019.

Problem Statement

This is a project to determine which basketball teams are most likely to make it to the semifinal round of the College Basketball Tournament known as the Final Four.

We will look at historical data to see which team metrics (individually or in combination) make a team more likely to make it into the Final Four. For example, if a team is more efficient defensively, does this have a direct relationship to their ability to get into the Final Four? What about defensively efficiency along with overall wins? We will figure out if there is a combination of metrics that give a team more of a chance of making it into this tournament.

We will load a historical data set from previous seasons, clean the data, and apply different classification algorithms to the data.

Models Algorithms

- K-Nearest Neighbour
- Decision Tree
- Support Vector Machine
- Logistic Regression

Metrics for Classification Accuracy

- Jaccard Index
- F1-Score
- Accuracy

Languages: Python

Platforms: IBM cloud, Watson Studio

Fields: Data Science, Machine Learning