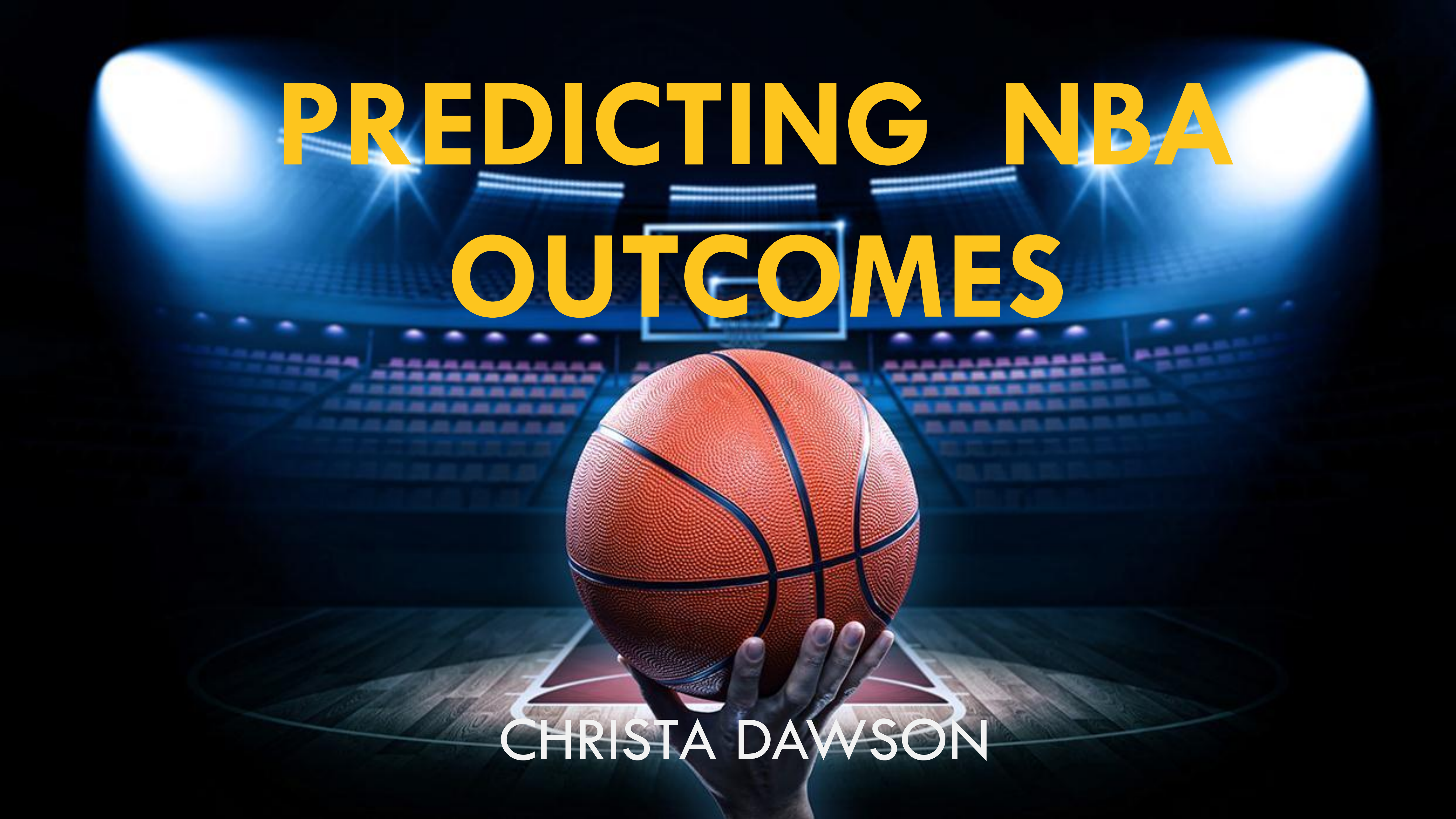


PREDICTING NBA OUTCOMES

CHRISTA DAWSON

DIRECTORY

1. Business Problem
2. Approach
3. Models & Results
4. Next Steps

Overview

What are we looking to solve?

As the sport of basketball has becoming increasingly popular, data has come to the forefront for analyzing player and team performance.

What if we had a model that could predict the outcome of any given game?

Approach

Data Science Process Steps

Data Mergers & Acquisitions

Acquired game data for each team over the past 18 seasons through basketball-reference.com.

Cleaning & Feature Engineering

Cleaned our data and created new variables such as team disparity, win-percentage at home, and rolling statistics of each teams past 10 games.

Modeling

Implemented six different classification models in order interpret which model would work best on our data.

Results

Accuracy was our primary evaluation metric for this project as we were trying to maximize the how well we can predict a given outcome

Source

basketball-reference.com

The aforementioned website is a great place for any avid NBA fan to find data relating to Game/Team/Player stats.

Engineering

Predictive Modeling

We needed to create data that was predictive rather than data from a current game (i.e. we needed data from before the game was going to be played in order to accurately predict).

Binary Classification

Multiple Classifiers

We exhausted our resources and used multiple classification models in order (Logistic Regression, KNN, Decision Tree, Random Forest, Adaboost and Gradient Boosting.

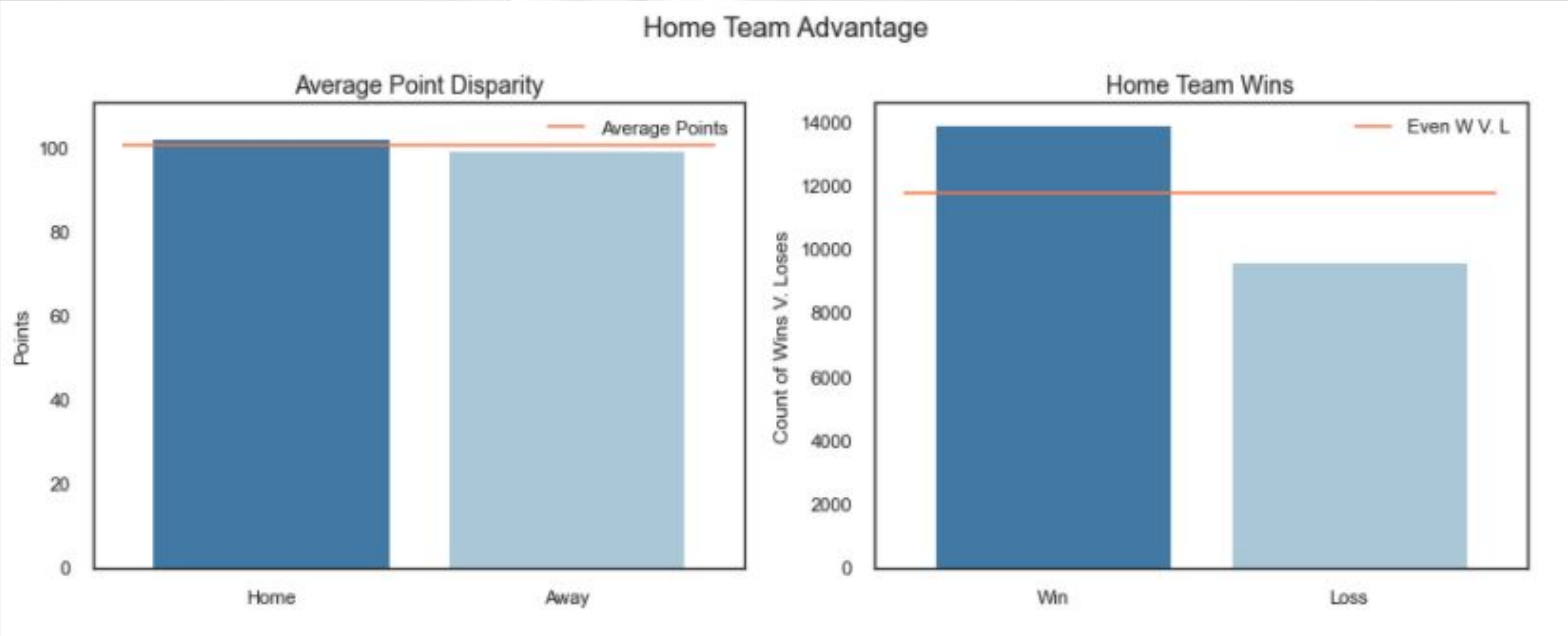
Results

Evaluation Metrics

Our evaluation metric of accuracy is important as we only care about how many games we predicted right; we do not care about false positives or false negatives.

Understanding Underlying Trends

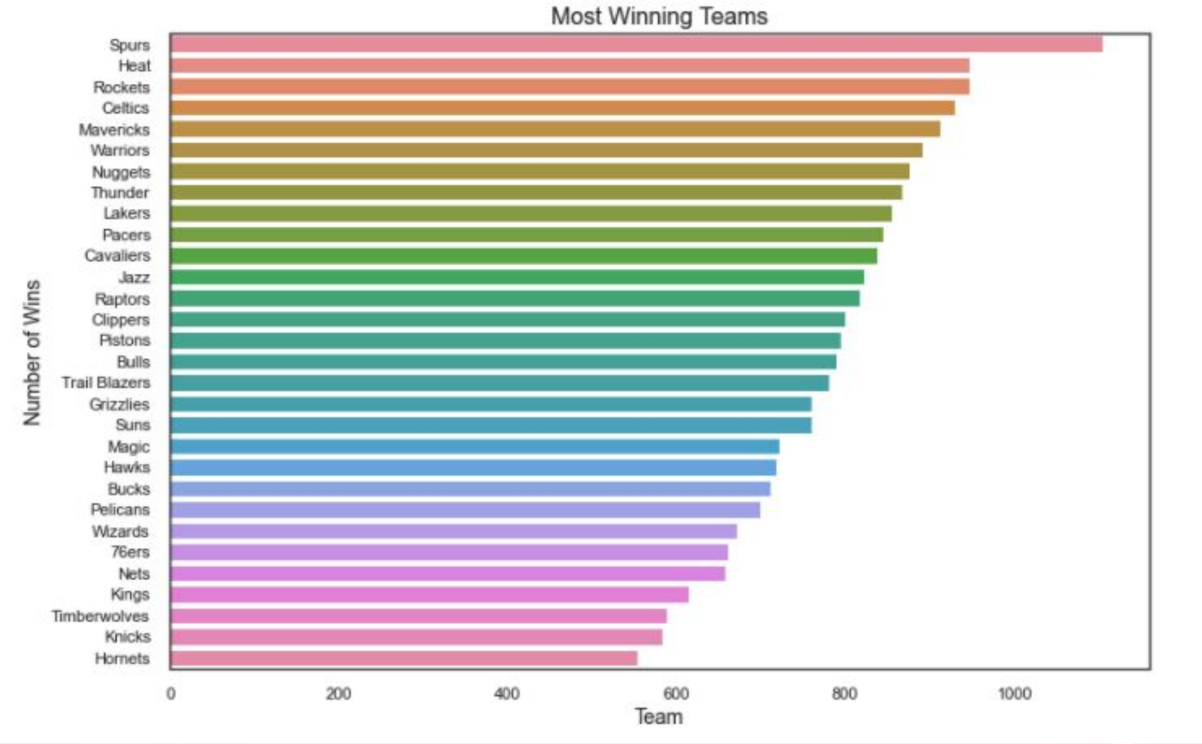
Home Team Advantage



www.github.com/dawsonc96
dawsoncr96@gmail.com

Home team advantage also shows the disparity between our classes. Due to this, we decided to resample in order to make our classes (home team wins/home team losses) more balanced.

Do certain franchises win more?

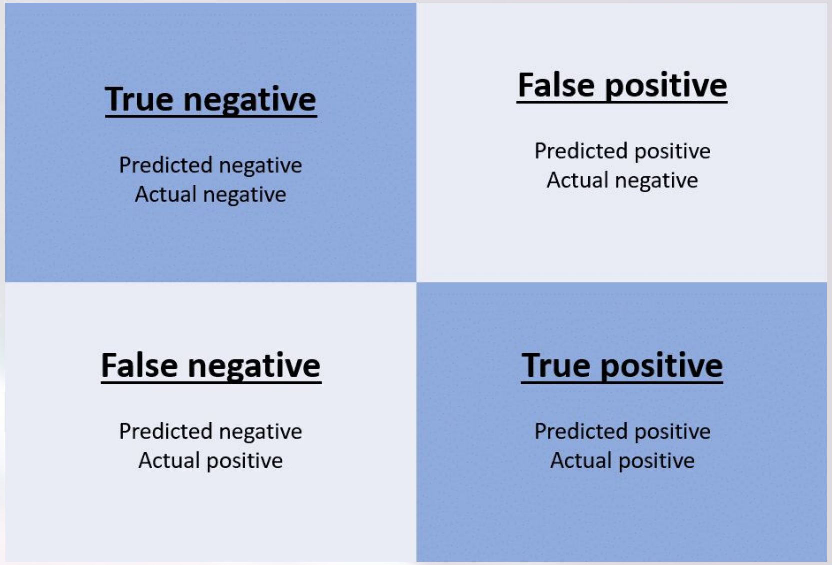


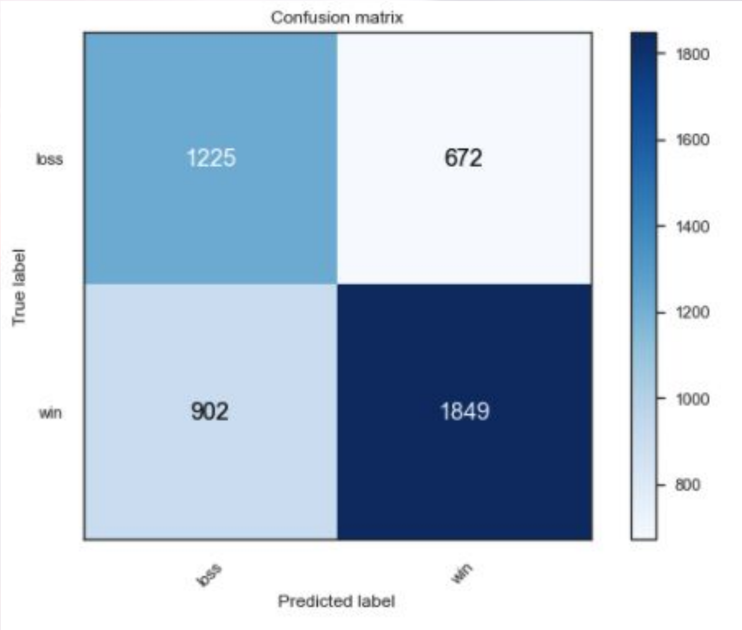
Although most teams should have similar amounts of wins over the course of 18 years, team wins are not normally distributed and therefore franchise pedigree is something we need to account for.

Evaluation Metrics

Classification Evaluation (Confusion matrix metrics):

- **Accuracy**: how often are we correct at predicting wins and losses for the home team for a given game?
- Precision: when we predict the home team to win, how often is that prediction correct?
- Recall (Sensitivity): what proportion of wins were identified correctly?
- F1 score: harmonic average of precision and recall metrics.





Models Used

Six different classification model to evaluate our data

Logistic Regression

KNN Classifier

Decision Tree

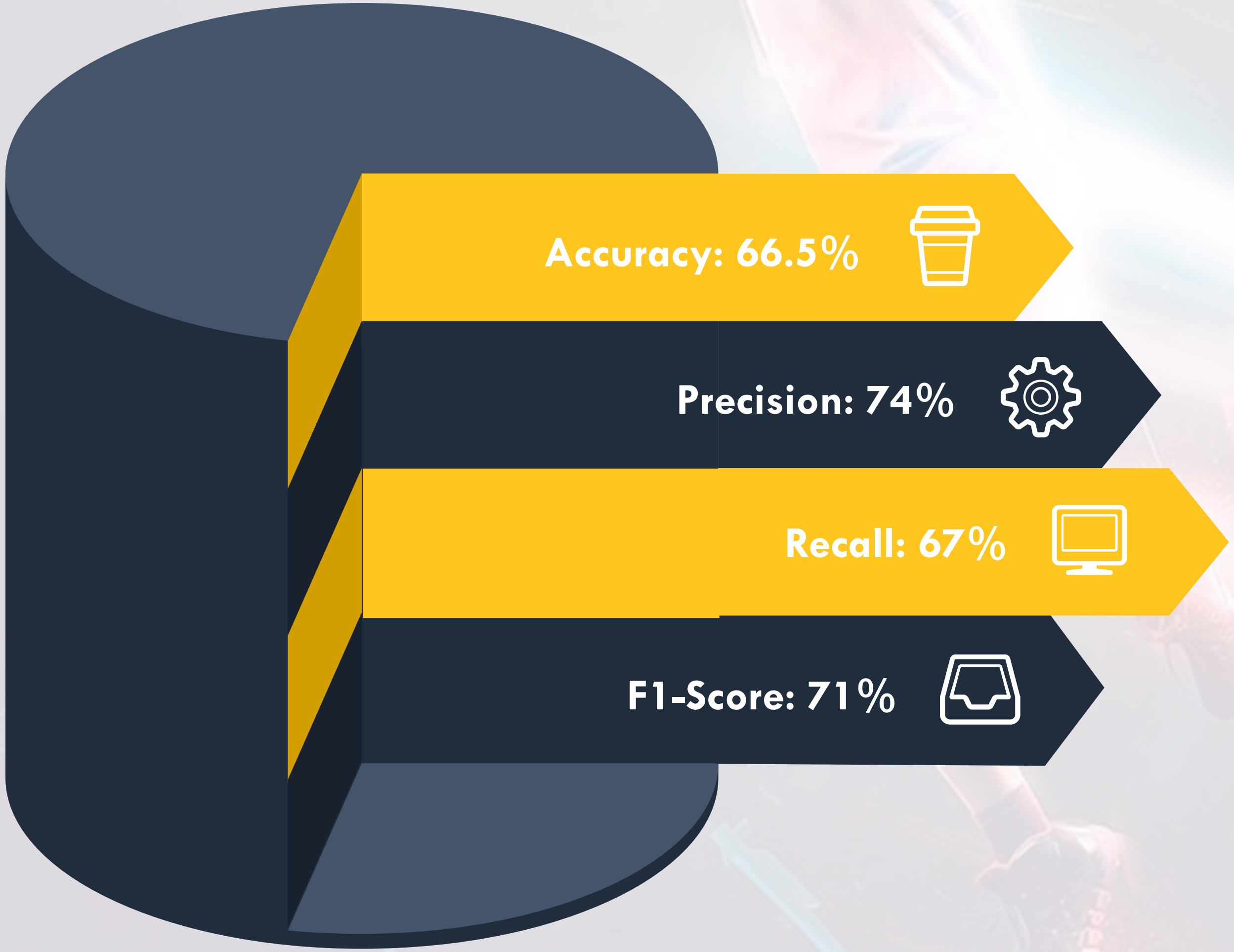
Random Forest

Adaboost

Gradient Boosting

Best Model

Which took home the finals trophy?



ADABOOST

AdaBoost - weak learners are tweaked in favor of those instances misclassified by previous classifiers.

Can be less susceptible to the overfitting problem than other learning algorithms (which was a problem in our model).

Conclusion

Machine learning was extremely useful in helping us predict the outcome of NBA games, going from a 50% baseline score to an accuracy score of 66.5%

Next Steps

Grid Search

For each model, run a couple of more grid searches so that we can ensure we are using the best model and parameters for our given data set

Regression Model

Implement a new model that predicts point differential (point spread) in a given game

Incorporate Player Data

Incorporating data from players of each given team will give us a better overall model.

CONTACT INFO

dawsoncr96@gmail.com

github.com/dawsonc96

THANK YOU!!!

Christa Dawson