
Random Forest Classification Model

— NBA Win-Loss Prediction —
Seasons 2002 - 2016

Project Team:

- Dragan Bogatic
- Howard Edwards
- Jordan Gross
- Ryan Reisner

Project goal:

- Make accurate predictions about NBA game Win-Loss based on historical game statistics using Machine Learning Model

Objectives:

- Research available sources of historical NBA game data
- Download data in standardized format
- Clean the data
- Select model features
- Select prediction variable
- Transform data to format to fit machine learning model

Objectives cont'd

- Select prediction model
- Train and test the model
- Evaluate prediction performance of the model
- Fine-tune model hyper-parameters to improve model prediction
- Evaluate results
- Conclusions

Resources:

- Data source: AWS Data Exchange / Stats Perform
- Data: NBA Seasons 2002 - 2016
- Data format: CSV
- Dataset type: detailed play by play stats by game played
- Dataset size: approx. 8 million rows by 326 columns

Machine Learning Model:

- Model selected: Balanced Random Forest Classifier
- Reduces overfitting
- Proven performer in classification prediction problems
- Works with categorical and continuous values
- Uses Ensemble Learning technique (many weak learners strong together)

Model Features and Prediction:

- Total features in raw dataset: 325
- Features selected for the base model:
 - HT_POSS
 - GAME_TIME
 - HT_SCORE_DIFF
 - HT_ORBD
 - HT_DRBD
 - HT_TOV
- Prediction: OUTCOME (Win-Loss)

Model Accuracy

Balanced Accuracy Score : 0.8892515660382707

Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	107355	13065
Actual 1	19194	150662

Classification Report

	f1-score	precision	recall	support
0.0	0.869380	0.848328	0.891505	120420.000000
1.0	0.903295	0.920203	0.886998	169856.000000
accuracy	0.888868	0.888868	0.888868	0.888868
macro avg	0.886338	0.884265	0.889252	290276.000000
weighted avg	0.889226	0.890385	0.888868	290276.000000

Most Important Features

