Oracle Arena: Predicting NBA Games

Triton Eden and Ryan Peruski

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Introduction/Motivation

The goal of this project is to predict the winner of current NBA games given previous NBA games. Furthermore, we want to predict the total number of points scored. Therefore, we will first collect data from an existing NBA stats API to form a constantly-updating database. Then, we will use a classifier for a winner predictor and a regressor for a total point predictor.

Data Collection

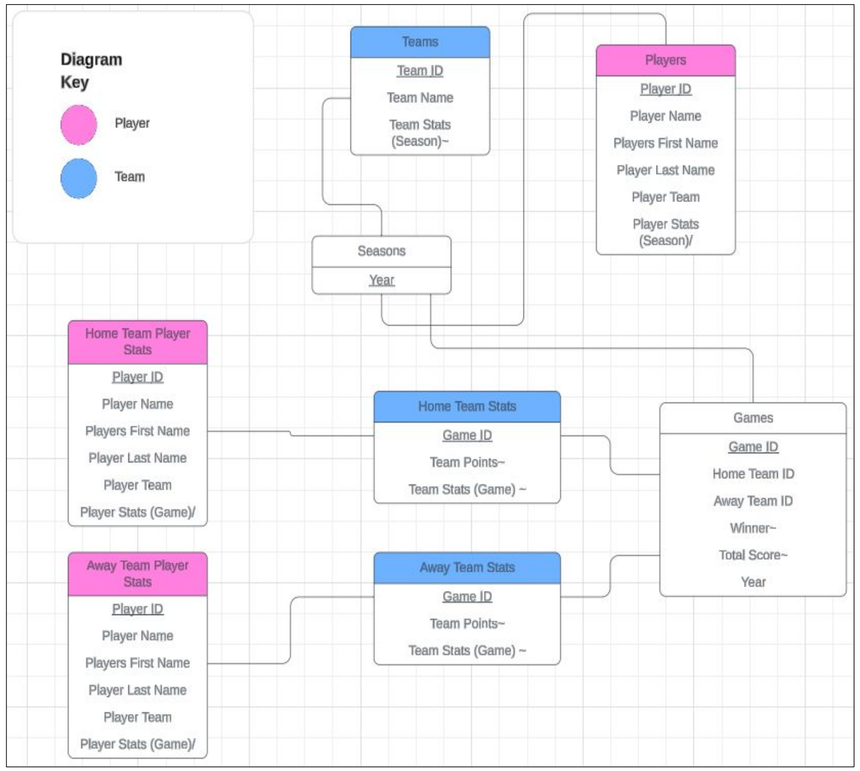
We collected data using the different API calls from the NBA stats API [1]. We call this API every day at noon when new games are available from the stats page [2]. We used an Azure PostgreSQL server [3] to store this data. From here, it can be used in a Dataframe with the Python library SQLAlchemy, and normal methodologies can be used.

Model Features

Stats Columns =

{FGM, FGA, FG3M, FG3A, FTM, FTA, OREB, DREB, AST, STL, BLK, TO, PTS, POSS, wins, losses}

Stats are from the 2018-19 NBA season to the current NBA season. Each of the stats in the stats columns are found for the home and away team per 100 possessions. Those stats are then averaged over the current season and over the past 5 games for the team and the average of their opponents in the current season.



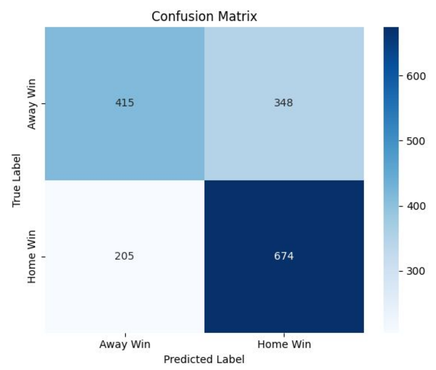
Data ERD

Results

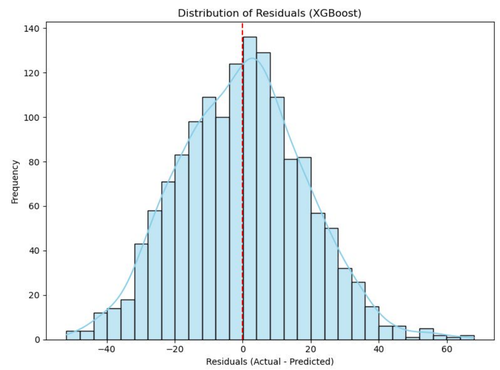
Win Prediction			Total Score Prediction		
	Regular Season	Playoffs		Regular Season	Playoffs
	Deep Feedforward Neural Network	Deep Feedforward Neural Network		Model Tuned XGBoost Regression	Ridge Regression
Accuracy	0.66	0.67	RMSE	18.60	16.67
F1 Score	0.71	0.73	MSE	346.33	277.83
Loss	0.62	0.66	R^2 Score	0.11	0.08

Model Metrics

Model Outputs



Regular Season Win Predictions



Regular Season Total Score Predictions

Discussion/Conclusion

Whenever data is pulled from an external source, it is best to incrementally test and test often. Using API's are harder than they look! Lots of data + multiple languages = lots of cleaning and normalization Depending on the size and contents of your training data, model selection and hyperparameter tuning can look very different even for the exact same prediction. If a person used our model results as a tool along with their personal intuition, that person WOULD likely successfully gain money in the long run

Citations

[1] Swar. (n.d.). *nba_api: An API client for NBA.com*. GitHub. Retrieved April 25, 2025, from https://github.com/swar/nba_api
[2] National Basketball Association. (n.d.). *NBA.com/stats*. Retrieved April 25, 2025, from https://www.nba.com/stats
[3]Microsoft Corporation. (n.d.). *Microsoft Azure*. Retrieved April 25, 2025, from https://azure.microsoft.com
[4]TritonEden. (2025). *Oracle-Arena: Senior Design Project* [Computer software]. GitHub. https://github.com/TritonEden/Oracle-Arena

Acknowledgement

Thank you to the University of Tennessee department of EECS for providing the opportunity to learn and grow in the field of computer science, as well as our professor of this current class, Dr. Jack Marquez, and our TAs.