Name(s)
Joseph P. Thomas

Project Number
J0818

Project Title
A Machine Learning Approach to Predicting NBA Rookie Potential

Abstract

Objectives/Goals
The objective of this project was to apply machine learning algorithms to predict whether an NBA rookie will become an all-star based on his first year statistics.

Methods/Materials
The data obtained from Kaggle, a data science competition platform, contained per-year statistics for every NBA player from 1950 to 2017. Another dataset containing every player who became an all-star was obtained from Wikipedia. These two datasets were combined into one with 52 columns and 3,921 rows. Python was used to trim the data by removing the years 1950 to 1982 because most statistics were not recorded during these years. The years 2006 to 2017 were also removed to ensure that every rookie had finished his career. Seventy percent of this remaining dataset was devoted to training and the rest was used for testing. The training set had the problem of class imbalance, which in this case means that there are many more non all-stars than all-stars. This problem was solved by removing random non all-star players so that the number of non all-stars was equal to the number of all-stars. The rookie dataset was then classified by four machine learning algorithms that were already trained.

Results
Each machine learning algorithm was run three times on the test set and an average was calculated for the accuracy, precision, recall, and F1 score. The Naive Bayes classifier performed the best with an average accuracy of 82%, precision of 33%, recall of 80%, and a F1 score of 47. These results established viability of machine learning as a way to overcome cognitive biases in player-trading decisions.

Conclusions/Discussion
The Gaussian Naive Bayes classifier performed much better than other algorithms because it works very well with small datasets and makes probability based decisions. The Naive Bayes classifier was correct in its predictions 33% percent of the time, which is better than a classifier that always assumes one answer. Removing the class imbalance contributed to the success of this project by avoiding the fixed answer, non all-star.

Summary Statement
Using analyzed data and the power of machine learning, I discovered that computer algorithms can assist general managers to overcome cognitive biases when making player-trading decisions.

Help Received
I applied machine learning algorithms to the data that I gathered and analyzed. The help that I received was from my science teacher, Mr. Colucci to understand the capabilities of the different machine learning algorithms.