Moses Xie & Victoria Spann-Burton

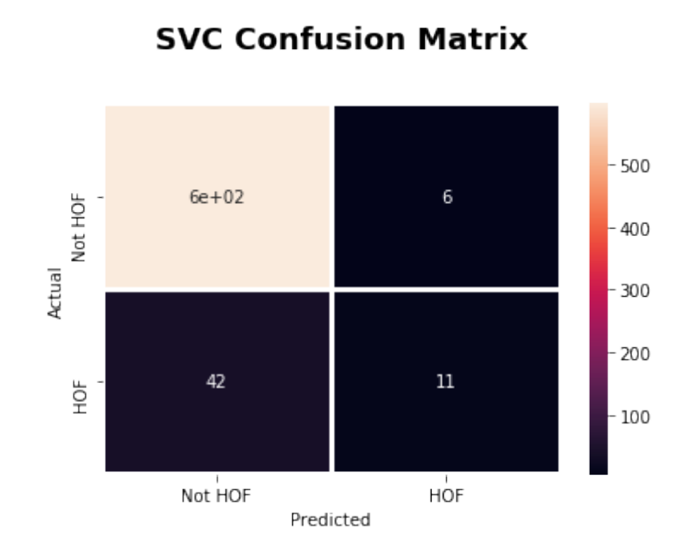
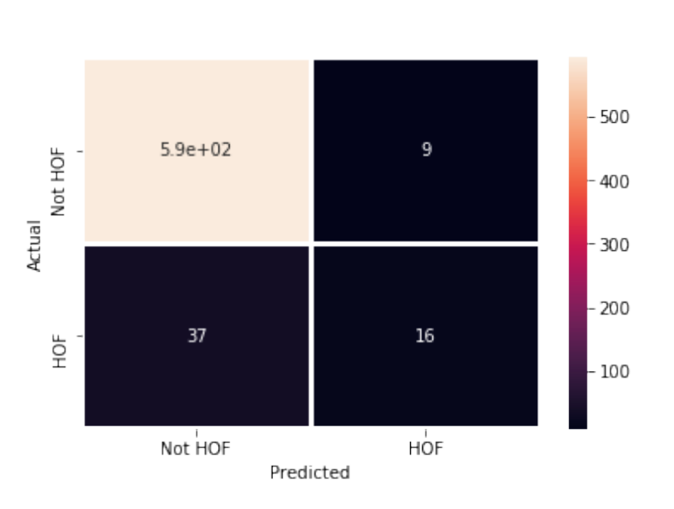
Intro to Machine Learning Final

**Predict Hall of Famers and All Stars from Rookie Year Stats in the NBA**

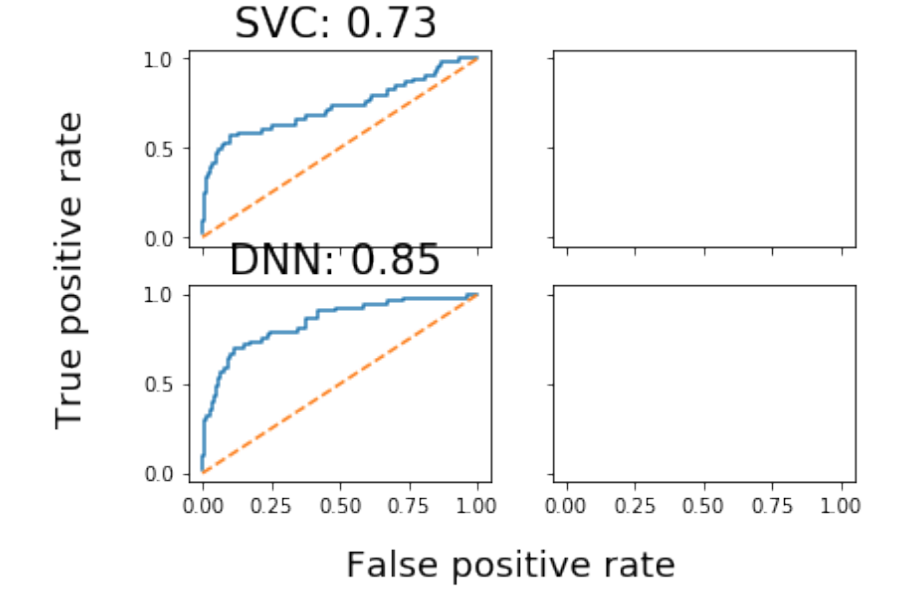
Often times, many star basketball players are followed from their college years into their rookie year of the NBA very closely. Based on performance during rookie year, many fans have high expectations for these players to become all-stars or even hall-of-famers. But, is that always the case?

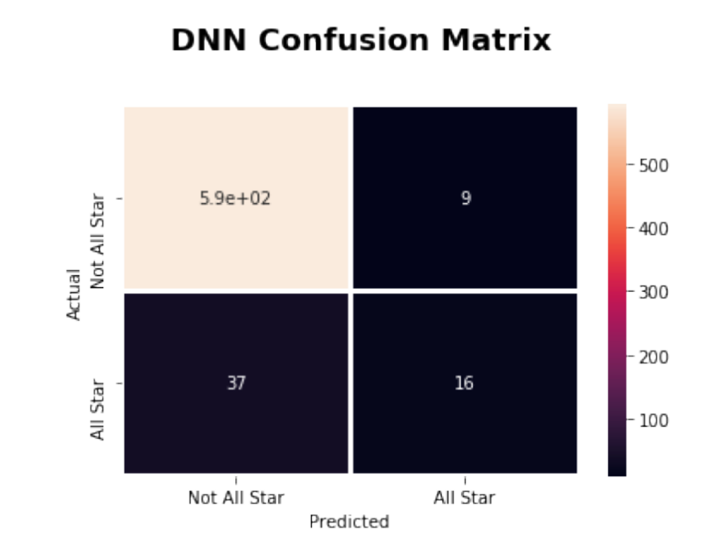
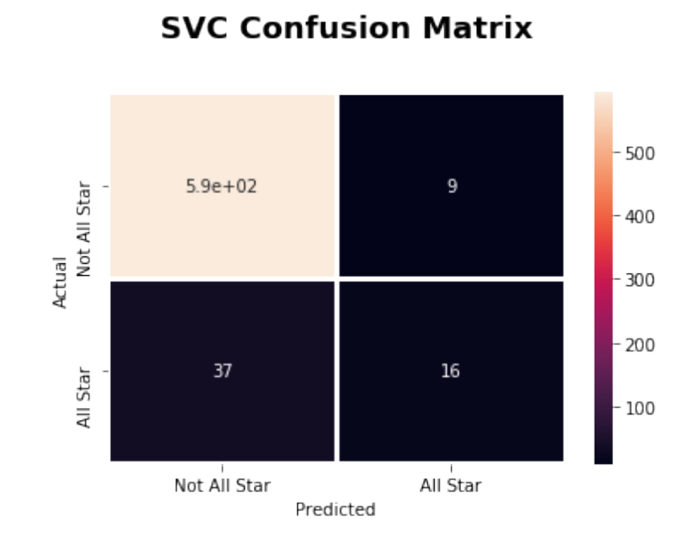
We retrieved the stats of the rookies’ players during the 1979-1980 season and of the 2017 draft class. In each model used to predict hall-of-famers, 0 was assigned to a player is not predicted to be a Hall-of-Famer or all-star and 1 is assigned to someone who is. We used support vector machine and deep neural network to try and make the predictions. Each model was created by using a train/test split with a test size of 25%. Regarding the 2017 draft class, there is a 50 game limit so the sample size for players’ stats is a decent size.

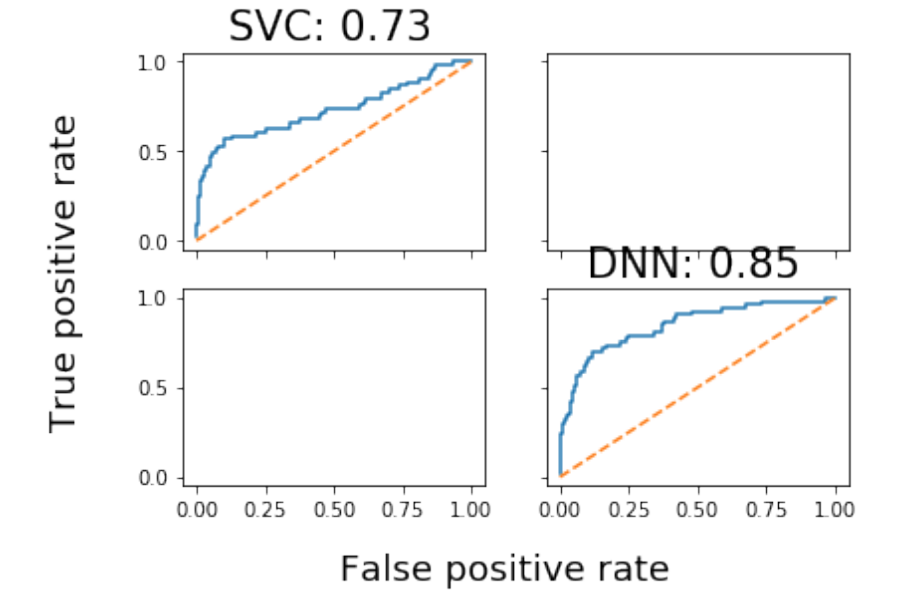
Out of of 2618 players drafted since the 1980 draft, 51 made the hall of fame and 229 were on at least one all-star team. So, 1.95% players drafted since 1980 were hall-of-famers and 8.95% were all-stars. Based on these statistics, it is not likely for many players to become all-stars and a lot less likely for them to become hall-of-famers. One of the things we realized about analyzing the models, was checking if they predict all 0s. If a model predicts almost all 0s, then it is high accuracy given the stats from the 1979-80 season. We also checked for the ROC curve to see if it would have an area under the curve that was closer to a perfect model.

**SVC Confusion Matrix, DNN Confusion Matrix for Hall of Fame**

****

****

****

When a random sample was chosen for the Hall of Fame analysis at first, there were actually more hall of famers than expected. Still, after measuring the accuracy, the SVC and deep neural network models were found to have high accuracy scores and cross-validated scores. None of the models predicted any of the rookies to be hall of famers which is not surprising. When measuring the accuracy of the models for all-stars, the scores were higher because of the greater number of all stars than hall of famers.

Many of the rookies seem promising, but both models predicted that no one would be a hall of famer from the 2017 draft class. The SVC model predicted that Ben Simmons would be an all-star and the deep neural network predicted that Ben Simmons and Donovan Mitchell would be all-stars.

Some of the shortcomings of this project is that simple logistic regression could have been used, but we wanted to explore different models. Another is that all stars are selected each season so we should have used a season model for the case to predict all stars.