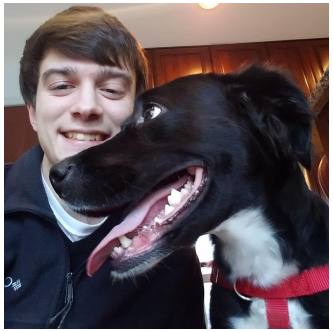
Predicting All-NBA Teams

Dan Salerno, Catherine Javadian, Brianne Trollo

Team



Dan Salerno

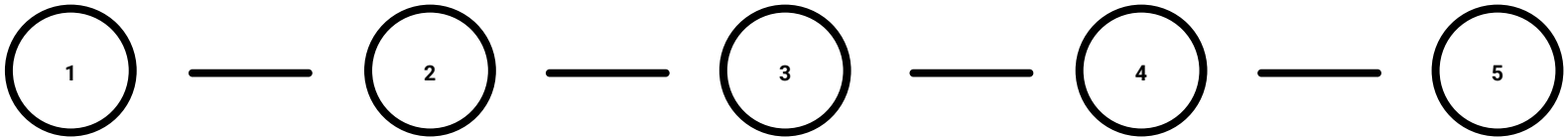
Catherine Javadian

Brianne Trollo

Background

- Every year, NBA media members vote on who is selected to the All-NBA team after the end of the regular season
  - 3 Teams (1st, 2nd, and 3rd)
  - Each team is comprised of 2 Guards, 2 Forwards, and 1 Center
- Our goal is to create a model that can predict who from the current 2018-19 NBA season will be selected for the All-NBA team

Pipeline



Scrape Data

Python Script

Merge Data

R Program to combine all data into a single CSV file

Exploratory Data Analysis

Understand the data

Classification

Build, train, and evaluate our models

Prediction

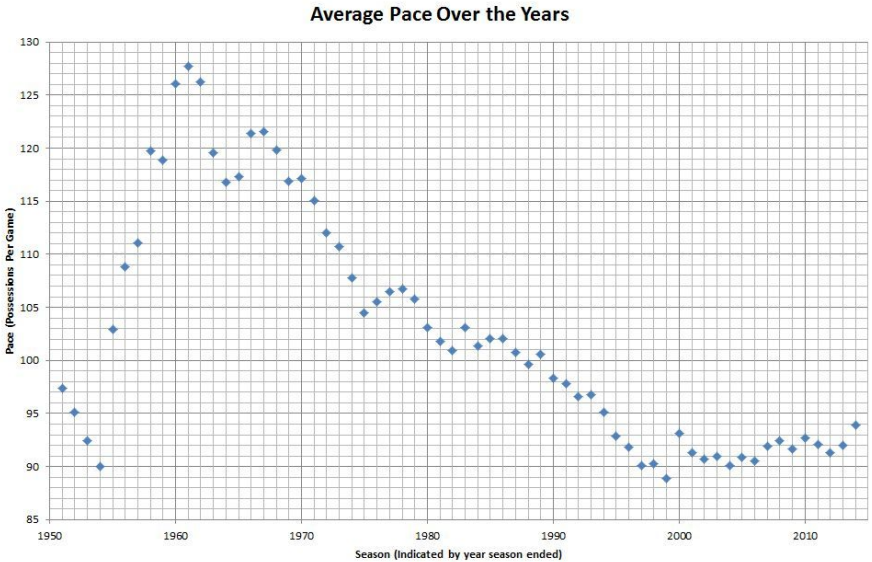
Use our best models to predict this year's All-NBA team

Data Source

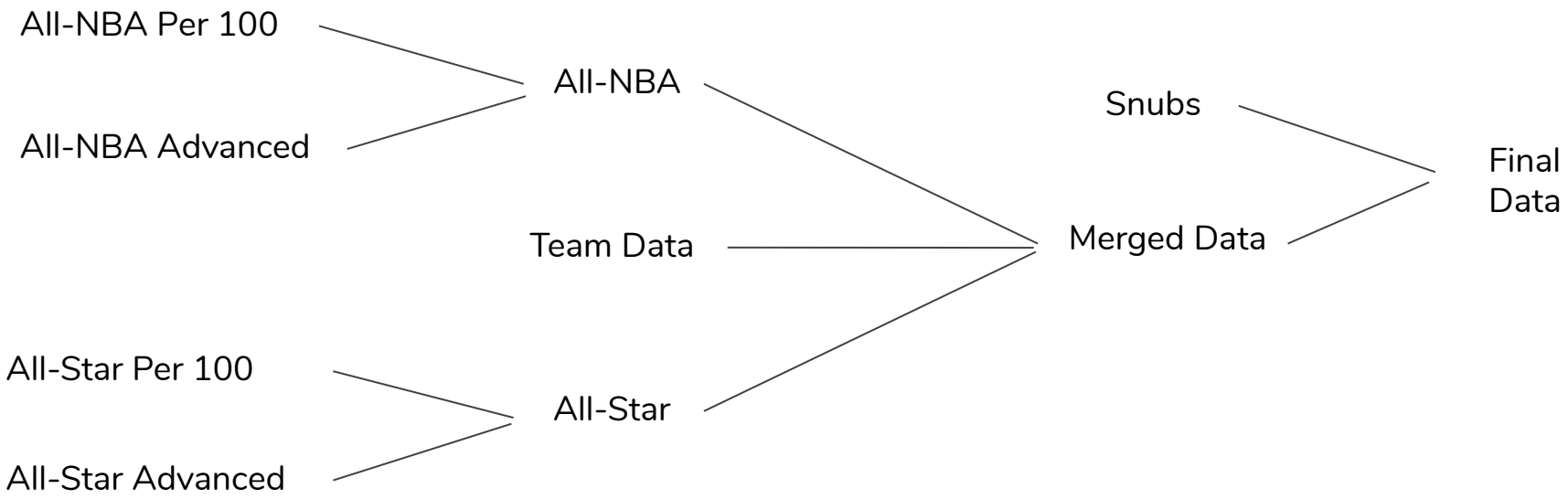
- All of our data comes from https://www.basketball-reference.com/
  - Scraped with a python script
- Data Collected
  - All data from the 3 Point Era (1979-80 - Present)
  - Two groups
    - Players who have made the All-NBA team
    - Players who have made the All-Star team
    - Optional group: All-Star snubs
  - 3 Data Sets
    - Per 100 Possession Counting Stats
    - Advanced Stats
    - Team Win/Loss Data

Data Source - Assumptions

- 3 Point Era only
  - Adding the 3 point line changed basketball so much that comparing players before and after the 3 point line is very difficult
- Using All-Star players instead of all players
  - ~500 NBA players per season, but only 15 make the All-NBA team
  - >90% of NBA players in a season have a negligible chance to make the All-NBA team
- Using Per 100 Possessions Data



Merging Data

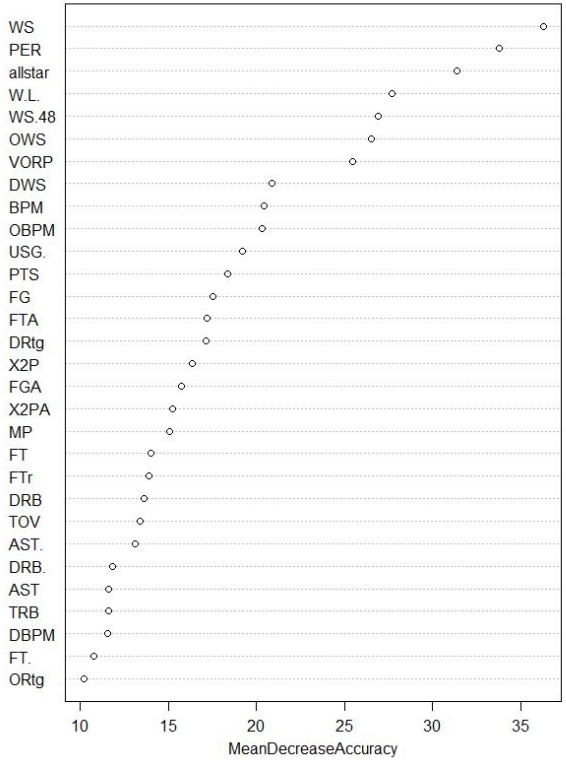


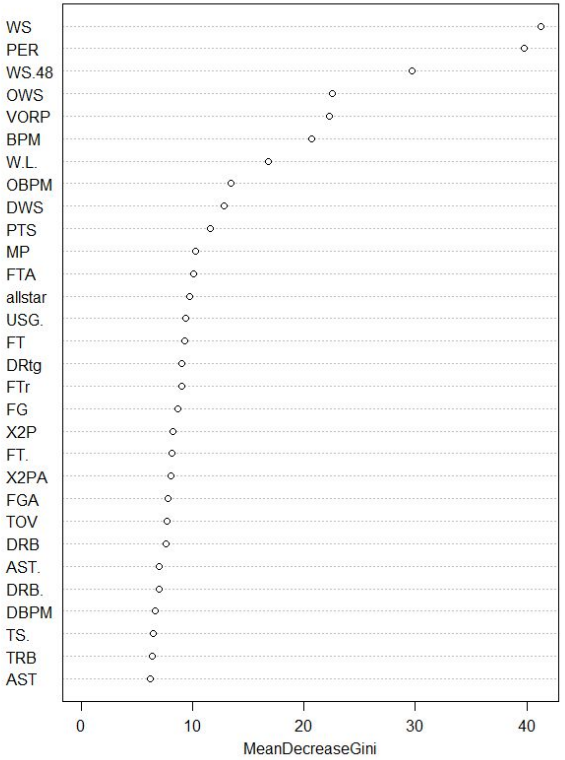
EDA

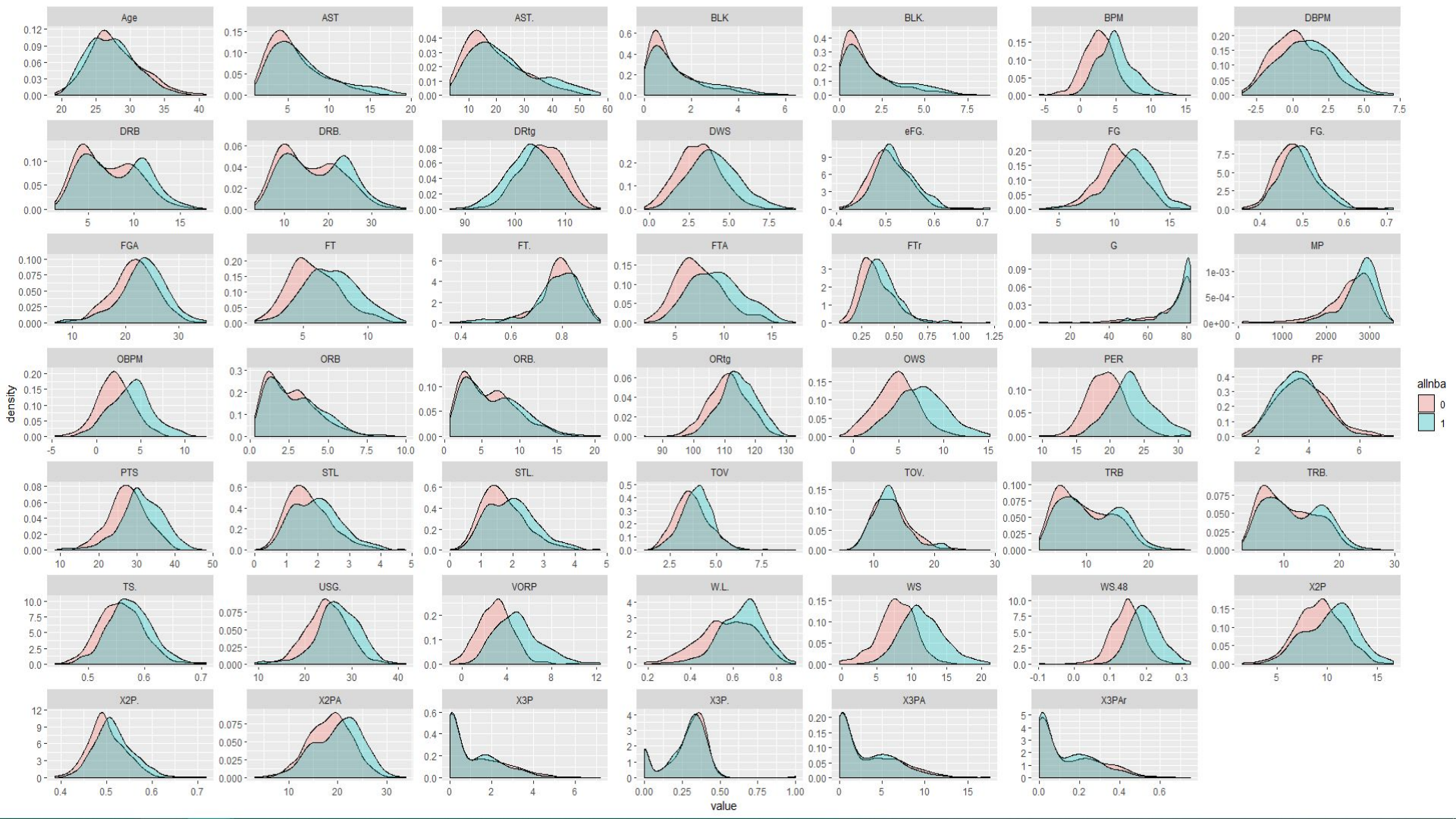
allnba	allstar
0:511	0: 51
1:545	1:1005

DRtg	
Min.	: 87.0
1st Qu.	:101.0
Median	:104.0
Mean	:104.2
3rd Qu.	:108.0
Max.	:117.0
NA's	:1

W.L.	
Min.	:0.1830
1st Qu.	:0.5240
Median	:0.6205
Mean	:0.6039
3rd Qu.	:0.6950
Max.	:0.8900
NA's	:24






Age
AST
AST.
BLK
BLK.
BPM
DBPM
DRB
DRB.
DRtg
DWS
eFG.
FG
FG.
FGA
FT
FT.
FTA
FTr
G
MP
OBPM
ORB
ORB.
ORtg
OWS
PER
PF
PTS
STL
STL.
TOV
TOV.
TRB
TRB.
TS.
USG.
VORP
W.L.
WS
WS.48
X2P
X2P.
X2PA
X3P
X3P.
X3PA
X3PAr
density
value
allnba
0
1

Classification

- We chose to test the following classifiers in this project
  - KNN
  - Naive Bayes
  - C5.0
  - Random Forest
  - ANN
  - SVM

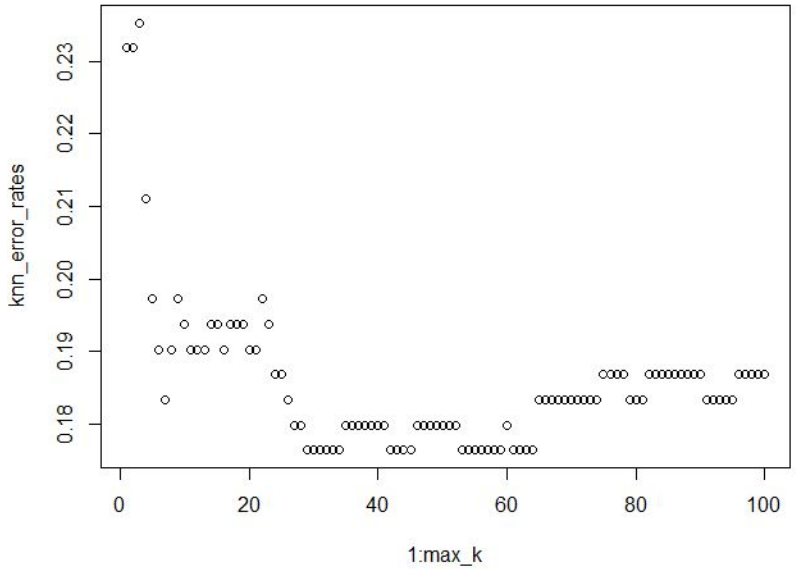
Methodology

- Normalize data using min max normalization
- Clean data of NAs (if necessary for the model)
- Split data set
  - One split to remove the data from the current year and save it for prediction later
  - Another split to create test and training data with the rest of the data
    - 70%/30% split
- Train/Evaluate models
- Predict with this years data

KNN

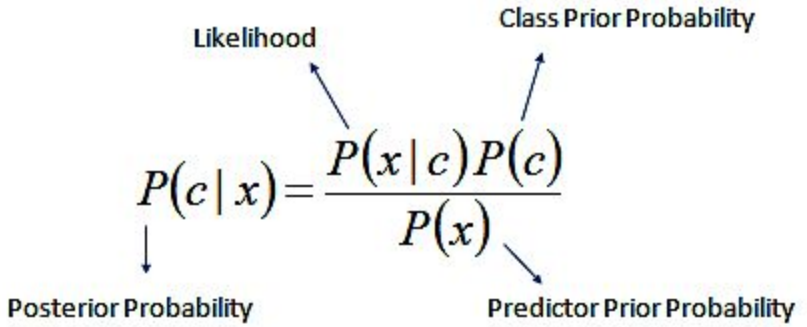
- Weighted KNN
- Value of K chosen by training the model with multiple Ks and choosing the K with the lowest error rate (K = 29)
- **Accuracy**: 0.8235

Prediction	Reference 0	1
0	105	21
1	30	133



Naive Bayes

- **Accuracy**: 0.7961



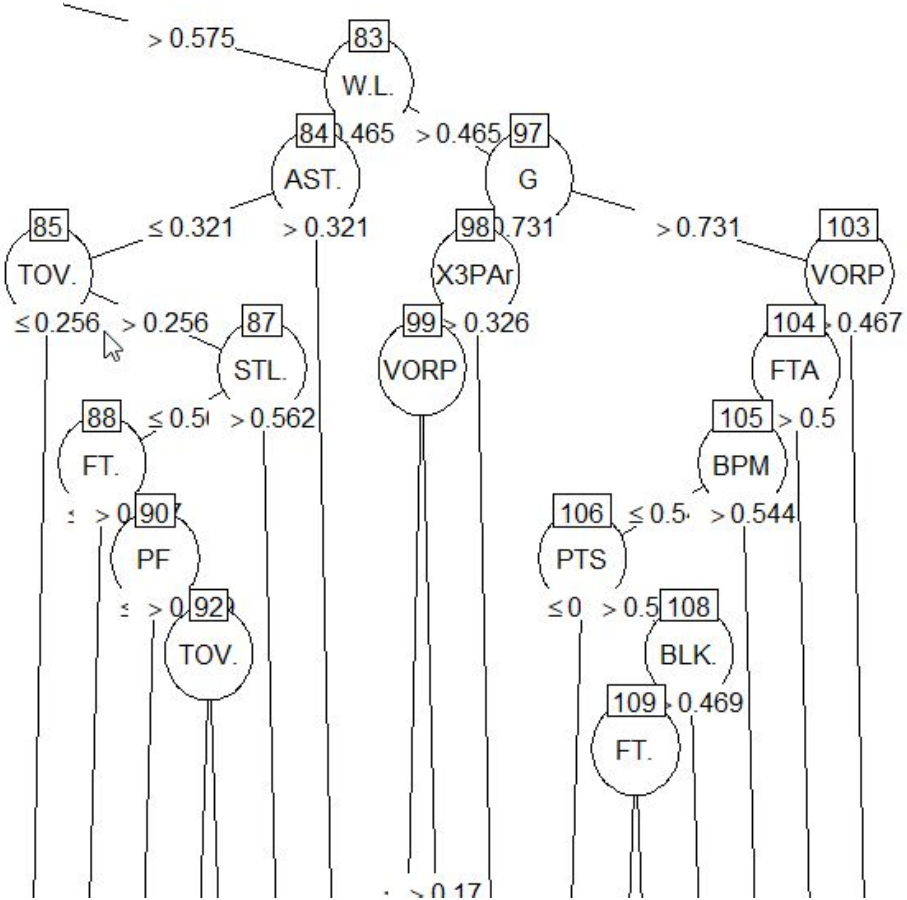
Prediction	Reference 0	1
0	121	33
1	30	125

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

C5.0

- **Accuracy**: 0.7249

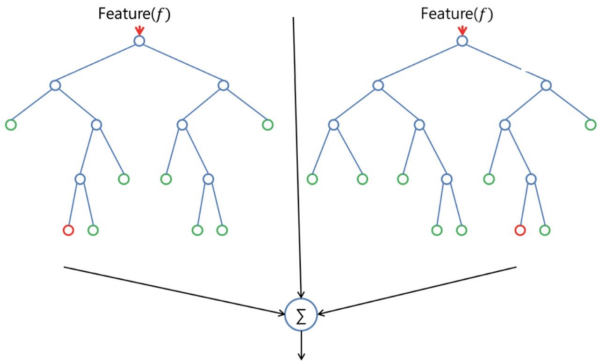
Prediction	Reference 0	1
0	105	39
1	46	119

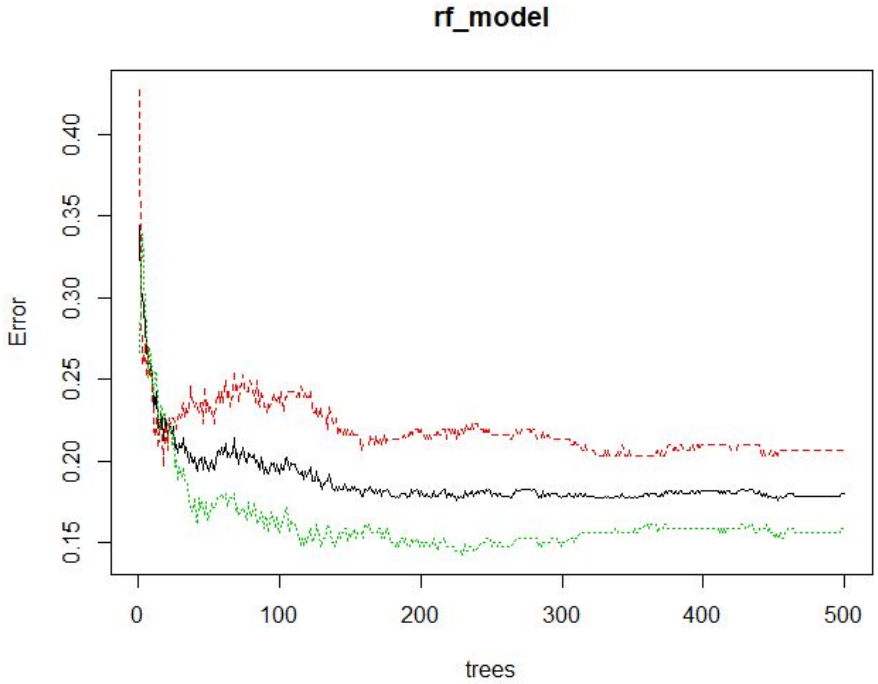


Random Forest

- **Accuracy**: 0.8304

Prediction	Reference 0	1
0	107	21
1	28	133

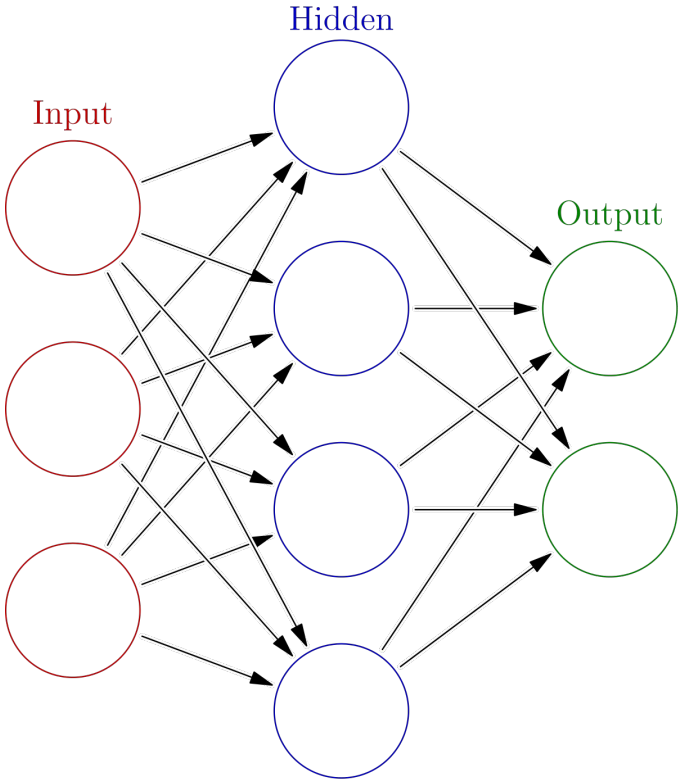




Artificial Neural Network

- **Accuracy**: 0.8235
- Only 1 hidden node

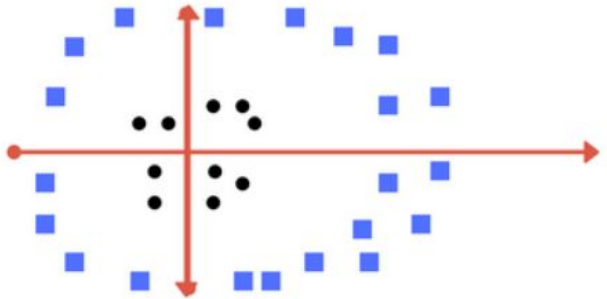
Prediction	Reference 0	1
0	109	25
1	26	129



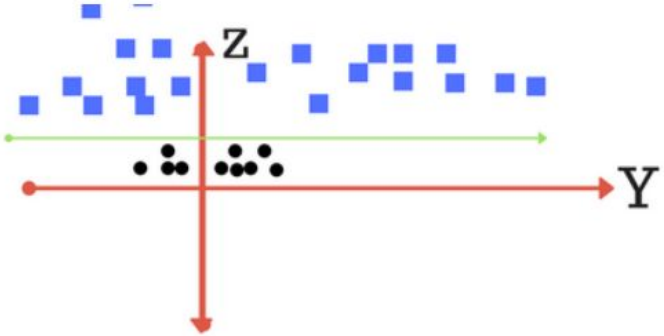
SVM

- **Accuracy**: 0.827

Prediction	Reference 0	1
0	106	21
1	29	133



Can you draw a separating line in this plane?



plot of zy axis. A separation can be made here.

Prediction

- With normal prediction with classifiers, the output is the class that each input falls under
- For this application, that does not work
  - How would you pick who is on which All-NBA team with the output?
- Instead, we predict the probability that each input would make the All-NBA team
  - Easy thanks to R

Prediction

Player	knn_allnba_prob
Giannis Antetokounmpo	1
Nikola Jokic	0.980434158
James Harden	0.92011827
Rudy Gobert	0.90103838
Karl-Anthony Towns	0.89025478
Nikola Vucevic	0.878679369
LeBron James	0.868954642
Joel Embiid	0.829010434
Paul George	0.796025288
Kevin Durant	0.780646384
Kawhi Leonard	0.767397995
Anthony Davis	0.762054713
Damian Lillard	0.664092828
Russell Westbrook	0.655863878
Stephen Curry	0.57890449
Kyrie Irving	0.431947994
Ben Simmons	0.361629306

- Example output from KNN
- To create the All-NBA team, grab the highest probability for each position
- Those are selected to the 1st team
- The next highest probability gets assigned to the 2nd team, and so on

- In order to create the final output, we average the probability for each player across all classifiers

Final Output

1st Team

James Harden (G), Damian Lillard (G), Giannis Antetokounmpo (F), Kevin Durant (F), Nikola Jokic (C)

2nd Team

Kyrie Irving (G), Steph Curry (G), Paul George (F), Kawhi Leonard (F), Rudy Gobert (C)

3rd Team

Russell Westbrook (G), Ben Simmons (G), Lebron James (F), Blake Griffin (F), Joel Embiid (C)

Possible Sources of Error

- All-NBA voting uses a point system, but data of the raw point values (not just the resultant teams) are hard to find
  - 1st team vote is 5 points, 2nd is 3, 3rd is 1
- Media members are biased
- Non-optimal parameters to R functions
- Some error due to data not being split by position
  - Not comparing only with others at their position
- Differences in basketball over the almost 40 years of data we have

Improvements/Future Work

- Try to find raw point totals of All-NBA voting and try to predict that instead
- More work testing different function parameters
- Gather position data and try splitting data by position and training/evaluating models with that data
- Do some form of feature selection
    - Some features are redundant/don’t add much value

Questions?

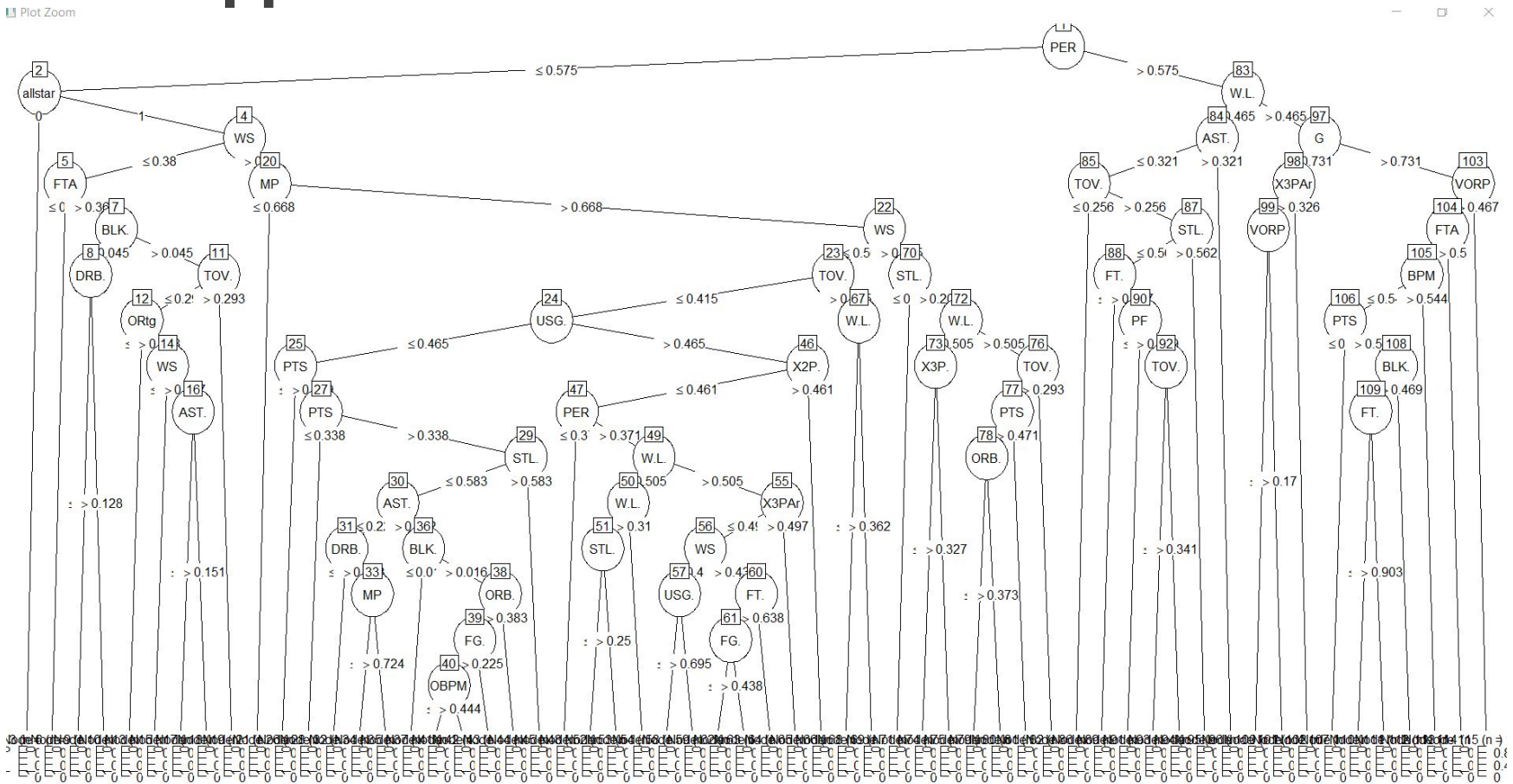
Appendix - Summary Output

```
     Season                Player          Age              G               MP             PER              X3PAr       
 1991-92: 36   Kobe Bryant     : 19   Min.   :19.00   Min.   : 4.00   Min.   :  91   Min.   : 9.60   Min.   :0.0000  
 1996-97: 30   Kevin Garnett   : 16   1st Qu.:25.00   1st Qu.:72.00   1st Qu.:2464   1st Qu.:18.90   1st Qu.:0.0100  
 2000-01: 29   Shaquille O'Neal: 16   Median :27.00   Median :78.00   Median :2781   Median :21.30   Median :0.0700  
 2006-07: 29   Tim Duncan      : 16   Mean   :27.52   Mean   :74.02   Mean   :2676   Mean   :21.45   Mean   :0.1329  
 2007-08: 29   Dirk Nowitzki   : 15   3rd Qu.:30.00   3rd Qu.:81.00   3rd Qu.:2995   3rd Qu.:23.70   3rd Qu.:0.2355  
 2009-10: 29   Karl Malone     : 15   Max.   :41.00   Max.   :82.00   Max.   :3533   Max.   :31.70   Max.   :0.7480  
 (Other):874   (Other)         :959   NA's   :1       NA's   :1       NA's   :1      NA's   :1       NA's   :1       
      FTr              ORB.             DRB.             TRB.             AST.             STL.             BLK.       
 Min.   :0.0830   Min.   : 0.700   Min.   : 3.20   Min.   : 2.60   Min.   : 3.10   Min.   :0.000   Min.   :0.00  
 1st Qu.:0.2920   1st Qu.: 2.900   1st Qu.:10.50   1st Qu.: 7.10   1st Qu.:12.10   1st Qu.:1.300   1st Qu.:0.60  
 Median :0.3640   Median : 5.600   Median :15.90   Median :10.60   Median :18.30   Median :1.700   Median :1.20  
 Mean   :0.3819   Mean   : 6.179   Mean   :16.57   Mean   :11.46   Mean   :20.92   Mean   :1.841   Mean   :1.85  
 3rd Qu.:0.4545   3rd Qu.: 8.600   3rd Qu.:22.50   3rd Qu.:15.80   3rd Qu.:27.10   3rd Qu.:2.300   3rd Qu.:2.60  
 Max.   :1.2190   Max.   :20.800   Max.   :37.80   Max.   :29.70   Max.   :57.50   Max.   :4.800   Max.   :8.80  
 NA's   :1        NA's   :1        NA's   :1       NA's   :1       NA's   :1       NA's   :1       NA's   :1     
      TOV.             USG.             ORtg             DRtg             OWS              DWS              WS        
 Min.   : 4.30   Min.   : 9.20   Min.   : 84.0   Min.   : 87.0   Min.   :-1.600   Min.   :-0.300   Min.   :-0.400  
 1st Qu.:10.75   1st Qu.:23.20   1st Qu.:109.0   1st Qu.:101.0   1st Qu.: 4.100   1st Qu.: 2.500   1st Qu.: 7.500  
 Median :12.50   Median :25.90   Median :113.0   Median :104.0   Median : 5.900   Median : 3.500   Median : 9.600  
 Mean   :12.94   Mean   :25.89   Mean   :112.8   Mean   :104.2   Mean   : 6.137   Mean   : 3.555   Mean   : 9.689  
 3rd Qu.:14.50   3rd Qu.:29.00   3rd Qu.:117.0   3rd Qu.:108.0   3rd Qu.: 8.000   3rd Qu.: 4.500   3rd Qu.:11.600  
 Max.   :29.20   Max.   :41.70   Max.   :133.0   Max.   :117.0   Max.   :15.200   Max.   : 9.100   Max.   :21.200  
 NA's   :1       NA's   :1       NA's   :1       NA's   :1       NA's   :1        NA's   :1        NA's   :1       
     WS.48             OBPM             DBPM             BPM              VORP             FG              FGA        
 Min.   :-0.0970   Min.   :-4.700   Min.   :-3.5000   Min.   :-5.900   Min.   :-1.100   Min.   : 3.20   Min.   : 6.80  
 1st Qu.: 0.1410   1st Qu.: 1.400   1st Qu.:-0.7000   1st Qu.: 2.000   1st Qu.: 2.600   1st Qu.: 9.60   1st Qu.:19.80  
 Median : 0.1720   Median : 3.000   Median : 0.6000   Median : 3.700   Median : 3.800   Median :11.00   Median :22.60  
 Mean   : 0.1722   Mean   : 3.065   Mean   : 0.7361   Mean   : 3.801   Mean   : 3.996   Mean   :10.92   Mean   :22.39  
 3rd Qu.: 0.2050   3rd Qu.: 4.700   3rd Qu.: 2.1000   3rd Qu.: 5.300   3rd Qu.: 5.200   3rd Qu.:12.30   3rd Qu.:25.30  
 Max.   : 0.3220   Max.   :12.400   Max.   : 7.0000   Max.   :15.600   Max.   :12.400   Max.   :16.80   Max.   :35.00  
 NA's   :1         NA's   :1        NA's   :1         NA's   :1        NA's   :1        NA's   :1       NA's   :1      
      X2P              X2PA             X3P              X3PA              FT              FTA              ORB        
 Min.   : 1.500   Min.   : 3.20   Min.   :0.000   Min.   : 0.000   Min.   : 1.300   Min.   : 1.800   Min.   : 0.300  
 1st Qu.: 8.000   1st Qu.:15.85   1st Qu.:0.000   1st Qu.: 0.200   1st Qu.: 5.000   1st Qu.: 6.400   1st Qu.: 1.300  
 Median : 9.900   Median :19.50   Median :0.400   Median : 1.600   Median : 6.300   Median : 8.200   Median : 2.400  
 Mean   : 9.817   Mean   :19.32   Mean   :1.096   Mean   : 3.068   Mean   : 6.539   Mean   : 8.404   Mean   : 2.715  
 3rd Qu.:11.600   3rd Qu.:22.80   3rd Qu.:1.900   3rd Qu.: 5.400   3rd Qu.: 8.000   3rd Qu.:10.100   3rd Qu.: 3.800  
 Max.   :16.600   Max.   :33.80   Max.   :7.200   Max.   :17.600   Max.   :12.900   Max.   :17.400   Max.   :10.000  
 NA's   :1        NA's   :1       NA's   :1       NA's   :1        NA's   :1        NA's   :1        NA's   :1       
```

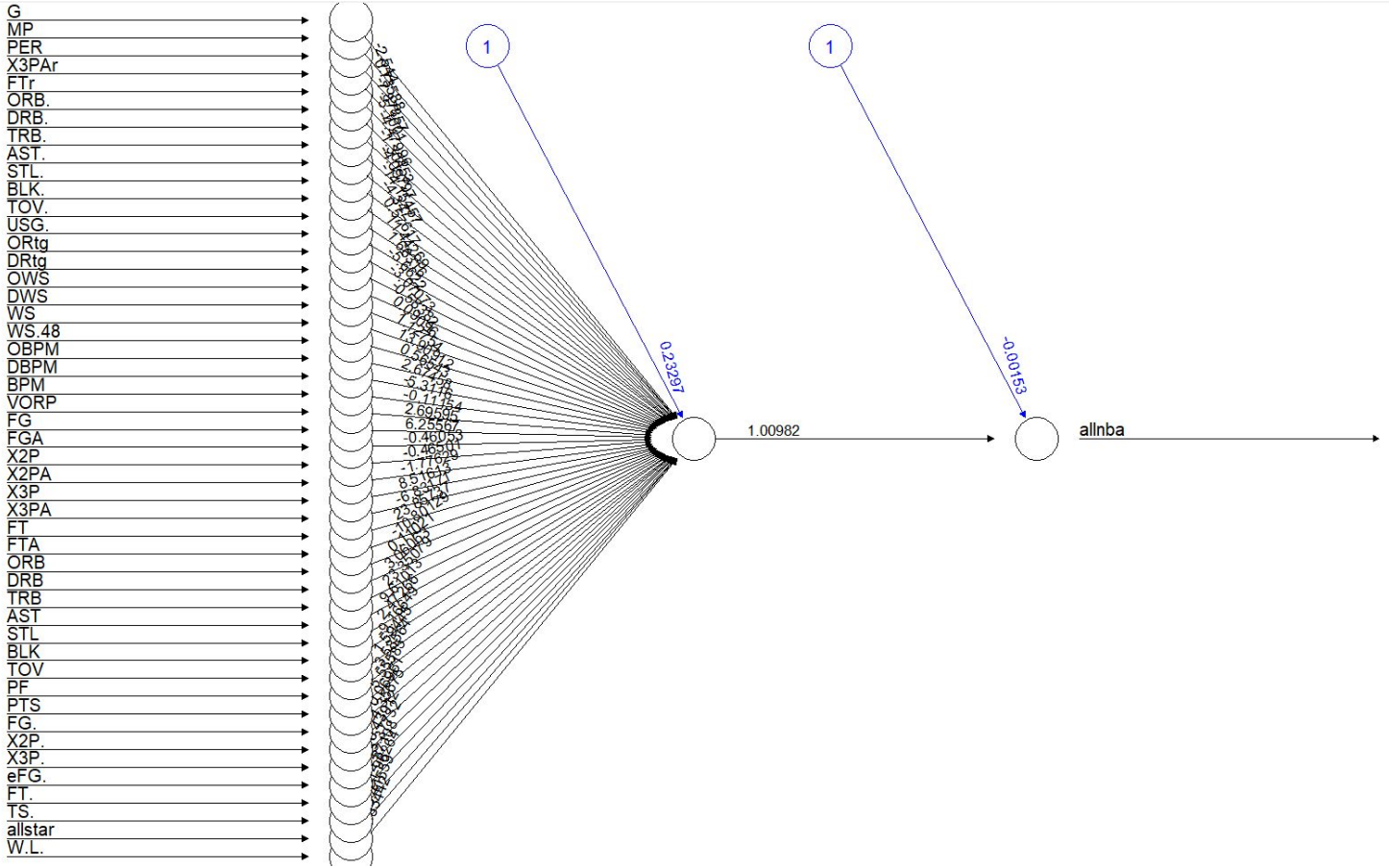
Appendix - Summary Output

```
      DRB              TRB              AST              STL              BLK              TOV              PF
 Min.   : 1.400   Min.   : 2.50   Min.   : 1.000   Min.   :0.000   Min.   :0.000   Min.   :1.100   Min.   :1.400
 1st Qu.: 4.650   1st Qu.: 6.35   1st Qu.: 3.600   1st Qu.:1.300   1st Qu.:0.400   1st Qu.:3.200   1st Qu.:3.000
 Median : 7.100   Median : 9.40   Median : 5.300   Median :1.700   Median :0.900   Median :3.800   Median :3.600
 Mean   : 7.512   Mean   :10.23   Mean   : 6.333   Mean   :1.841   Mean   :1.355   Mean   :3.808   Mean   :3.705
 3rd Qu.:10.250   3rd Qu.:14.20   3rd Qu.: 8.100   3rd Qu.:2.300   3rd Qu.:1.900   3rd Qu.:4.400   3rd Qu.:4.300
 Max.   :17.800   Max.   :26.60   Max.   :19.400   Max.   :4.800   Max.   :6.400   Max.   :9.300   Max.   :7.300
 NA's   :1        NA's   :1       NA's   :1        NA's   :1       NA's   :1       NA's   :1       NA's   :1
      PTS              FG.              X2P.             X3P.             eFG.             FT.              TS.
 Min.   : 8.50   Min.   :0.3580   Min.   :0.3870   Min.   :0.0000   Min.   :0.4050   Min.   :0.3550   Min.   :0.4410
 1st Qu.:25.90   1st Qu.:0.4580   1st Qu.:0.4800   1st Qu.:0.2000   1st Qu.:0.4860   1st Qu.:0.7475   1st Qu.:0.5380
 Median :29.50   Median :0.4870   Median :0.5040   Median :0.3095   Median :0.5100   Median :0.7940   Median :0.5640
 Mean   :29.47   Mean   :0.4906   Mean   :0.5089   Mean   :0.2723   Mean   :0.5144   Mean   :0.7821   Mean   :0.5659
 3rd Qu.:33.50   3rd Qu.:0.5190   3rd Qu.:0.5330   3rd Qu.:0.3690   3rd Qu.:0.5400   3rd Qu.:0.8410   3rd Qu.:0.5910
 Max.   :48.20   Max.   :0.7140   Max.   :0.7170   Max.   :1.0000   Max.   :0.7140   Max.   :0.9520   Max.   :0.7080
 NA's   :1       NA's   :1        NA's   :1        NA's   :40       NA's   :1        NA's   :1        NA's   :1
 allnba   allstar          W.L.
 0:511    0:  51   Min.   :0.1830
 1:545    1:1005   1st Qu.:0.5240
                   Median :0.6205
                   Mean   :0.6039
                   3rd Qu.:0.6950
                   Max.   :0.8900
                   NA's   :24
```

Appendix - C5.0 Plot



Appendix - ANN Plot



Appendix - Extra Classification Stats

KNN

Sensitivity	0.8636
Specificity	0.7778
Pos Pred Value	0.8160
Neg Pred Value	0.8333
Precision	0.8160
Recall	0.8636
F1	0.8391
Prevalence	0.5329
Detection Rate	0.4602
Detection Prevalence	0.5640
Balanced Accuracy	0.8207

Naive Bayes

Sensitivity	0.7911
Specificity	0.8013
Pos Pred Value	0.8065
Neg Pred Value	0.7857
Precision	0.8065
Recall	0.7911
F1	0.7987
Prevalence	0.5113
Detection Rate	0.4045
Detection Prevalence	0.5016
Balanced Accuracy	0.7962

C 5.0

Sensitivity	0.7532
Specificity	0.6954
Pos Pred Value	0.7212
Neg Pred Value	0.7292
Precision	0.7212
Recall	0.7532
F1	0.7368
Prevalence	0.5113
Detection Rate	0.3851
Detection Prevalence	0.5340
Balanced Accuracy	0.7243

Appendix - Extra Classification Stats

Random Forest

Sensitivity	0.8636
Specificity	0.7926
Pos Pred Value	0.8261
Neg Pred Value	0.8359
Precision	0.8261
Recall	0.8636
F1	0.8444
Prevalence	0.5329
Detection Rate	0.4602
Detection Prevalence	0.5571
Balanced Accuracy	0.8281

ANN

Sensitivity	0.8377
Specificity	0.8074
Pos Pred Value	0.8323
Neg Pred Value	0.8134
Precision	0.8323
Recall	0.8377
F1	0.8350
Prevalence	0.5329
Detection Rate	0.4464
Detection Prevalence	0.5363
Balanced Accuracy	0.8225

SVM

Sensitivity	0.8636
Specificity	0.7852
Pos Pred Value	0.8210
Neg Pred Value	0.8346
Precision	0.8210
Recall	0.8636
F1	0.8418
Prevalence	0.5329
Detection Rate	0.4602
Detection Prevalence	0.5606
Balanced Accuracy	0.8244

Appendix - Basketball Stats Glossary

https://www.basketball-reference.com/about/glossary.html