

Kobe Bryant Shot Selection

Taufer - PSTAT 231 - Spring 2016

Team: The Lakers

Colin Menz, Shaoyi Zhang, Emily Zheng

Does Kobe Make the Shot?

- Kobe Bryant attempted over 30,000 field goals attempts from 1996–2016
- Investigate which variables are most relevant to making shots
- Predict a shot's success or failure based on these findings
- Seek trends in the data
- Discuss whether or not our findings coincide with conventional wisdom

Data

- 25,697 records across 25 variables
- Each record represents one field goal attempt
- 14 categorical predictor variables, such as location on the court, season, types of actions, opponent...
- 10 continuous predictor variables, such as time remaining, position on the court, position on the globe...
- 1 dependent variable, indicator of whether or not the shot made is successful
- From Kaggle

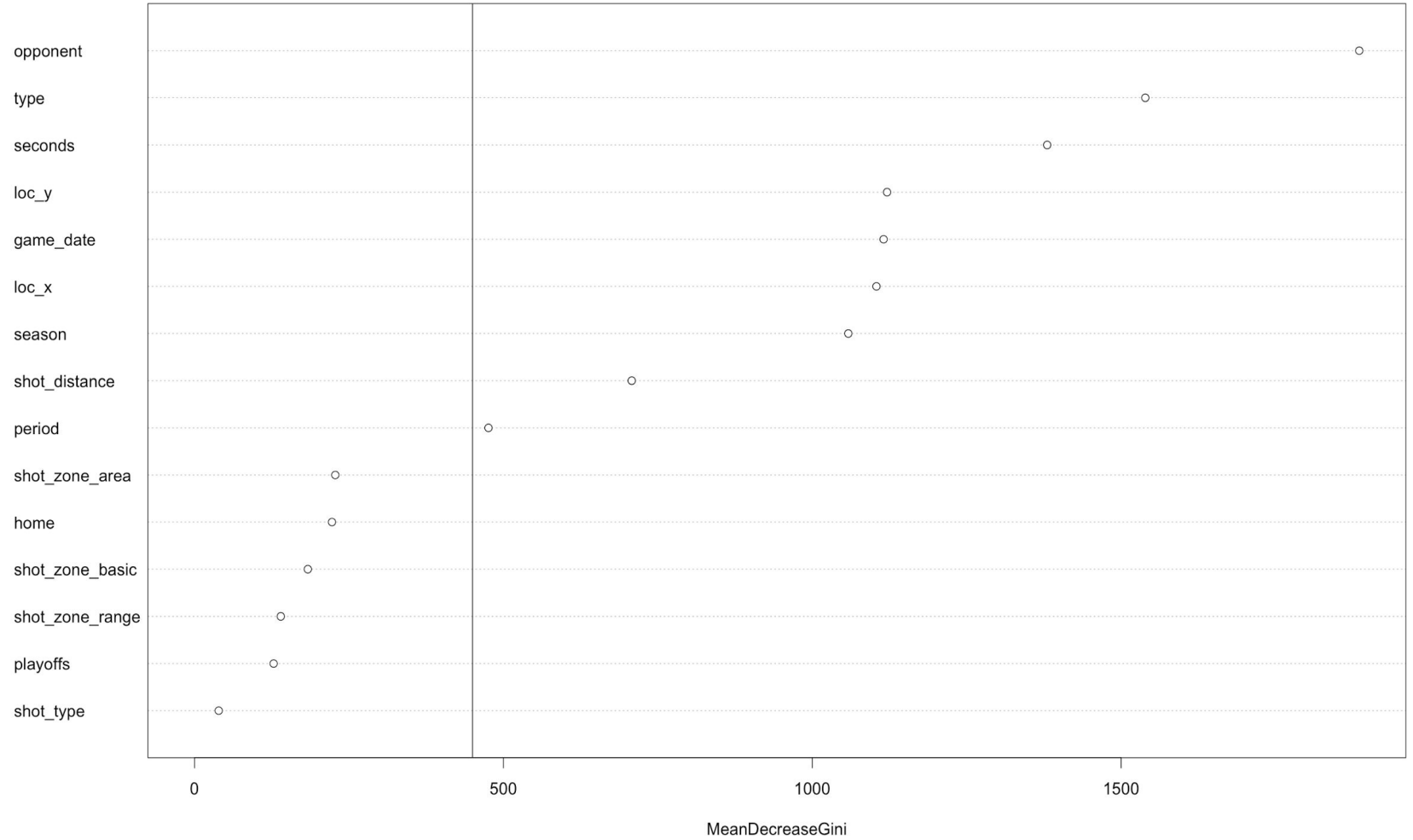
Review of Approaches

- PCA is used for dimension reduction, helpful for a dataset with 25 variables
 - Have mixed data types, won't work well with PCA
 - Are concerned with finding a probability rather simplifying data
- Classification using a single tree
 - Would be better off with bagged trees or random forest methods
- LDA to model predicted variable because we have well separated classes
 - Works best with small n & normality assumption; neither is true
- Worth noting that cluster analysis would not be particularly useful for us as our goal is to predict rather than separate

Final Approach

- Used random forest to identify relevant predictors
 - Generated 500 trees in order to measure ‘mean decrease in Gini’
 - Selected: ‘Opponent’, ‘Type’, ‘Seconds’, ‘Loc_x’, ‘Loc_y’, ‘Game_date’, ‘Season’, ‘Shot_distance’, and ‘Period’
- Dependent variable ‘Shot_made_flag’ is binary indicating multiple logistic regression
 - Want to predict probability of making the shot using equation given by MLR

Random Forest Variable Importance Plot



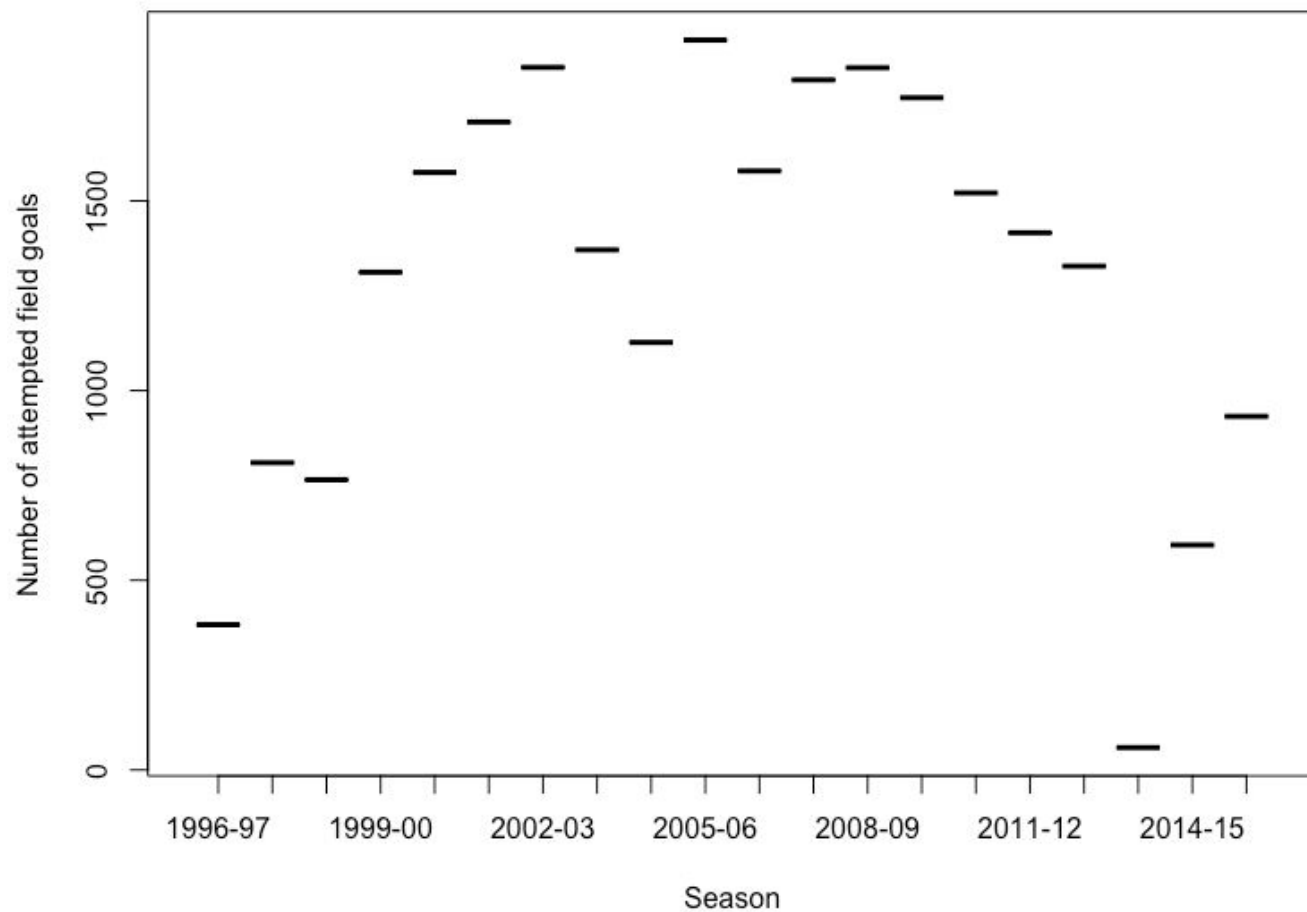
Results

- Local peak in shooting percentage from 2000–01, with career apex in 2007–08 season
 - Fairly typical of athlete's career to fall off around age 30
- The number of field goal attempts from season to season follows a similar trend as compared to shooting percentage
 - Two peaks, followed by a decline until his retirement

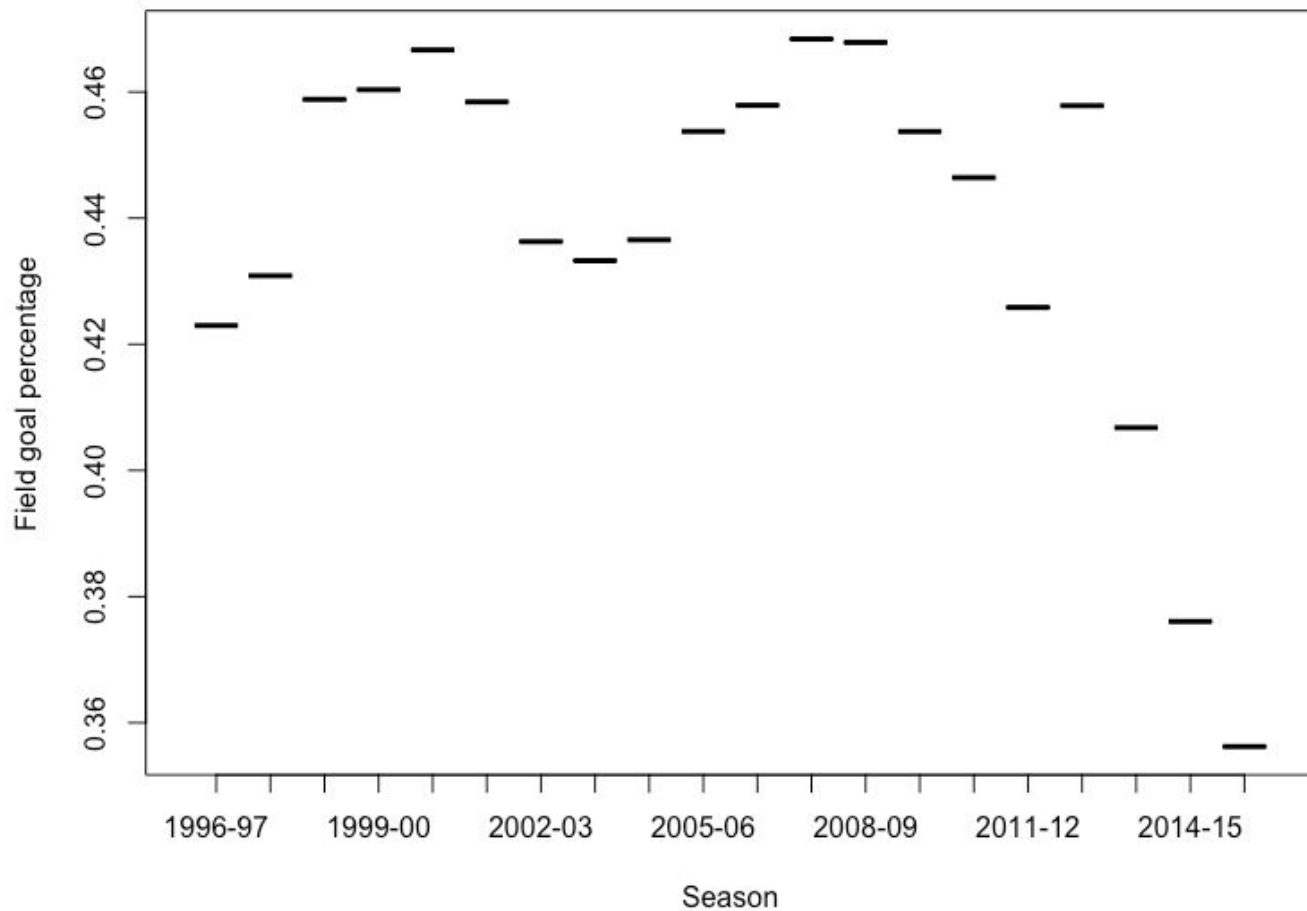
Model Results

- Logistic regression model with 9 predictors
 - Model assumes 99.7% probability of successfully making a shot if all predictors are zero
 - Type of shot very significant predictor
 - Dunks are more likely to be made than fadeaways and jump shots
 - Matchup was surprisingly significant
 - Performed worse against certain Eastern Conference teams
 - More likely to make shot in earlier periods and later seasons
 - Unclear why this is the case
 - Achieved test error rate around 30-33%

Frequency of Attempts



Percentage of Shots Made



Conclusion

- Random forest tends to overfit data
 - Achieved train error rate of 0.000476 and test error rate of 0.348
 - Attempted to increase amount of trees in random forest, did not do much to help
 - Attempted to decrease number of predictors used in each tree, we see a decrease in test error rate
- Most critical step was to identify relevant predictors
 - Many variables in original dataset were useless