

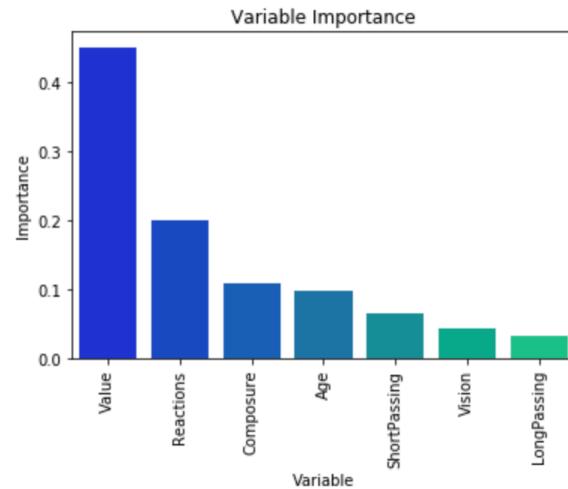
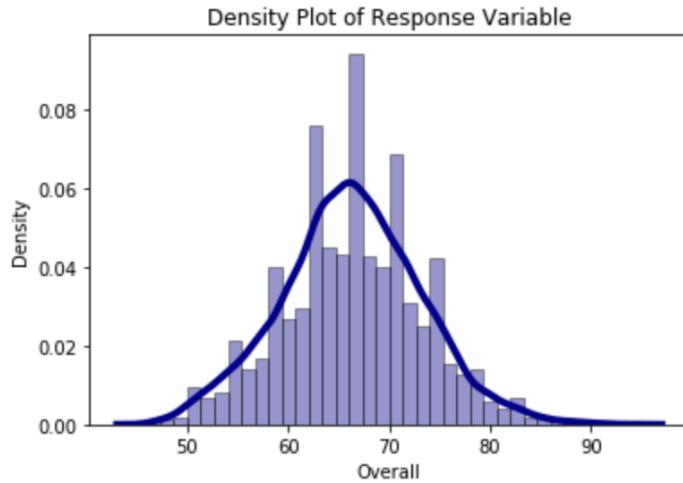
# Mid-project Review: Soccer Player Score (Class) Prediction

Si (Angela) Chen



# Highlights

- Response variable (player overall score) is normally distributed, players are divided into two classes, “good” and “bad” by median score, which is **66**.
- Seven variables are most related to the output variable: reactions, composure, vision, short passing, long passing, value and age. Their correlation is more than **0.37** respectively.
- Five models are built to classify the player class: logistic regression, random forest, decision tree, K nearest neighbors and neural network. Decision tree is the best model with misclassification rate of **0.028**.





# Review Progress

Epic	Story	Status
Exploratory Data Analysis	Data Overview	Done
Exploratory Data Analysis	Data Cleaning	Done
Model Building	Model Fitting	Done
Model Building	Model Comparison	Done

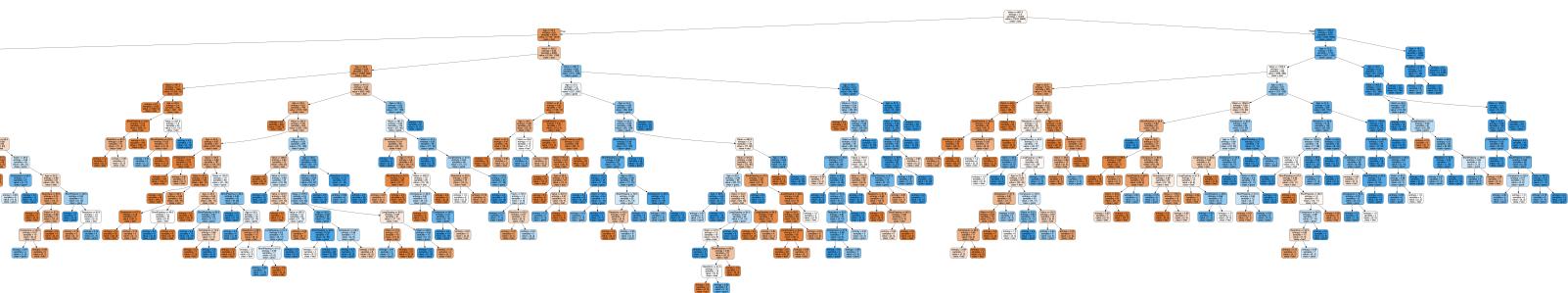


# Analysis

- Misclassification rate of all models:

Model	Misclassification Rate
Decision Tree	0.028476
Random Forest	0.033501
K Nearest Neighbors	0.061977
Logistic Regression	0.078169
Neural Network	0.111669

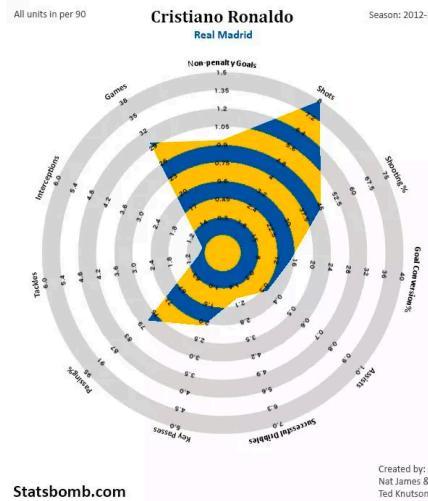
- Visualization of target decision tree:





# Lessons Learned

- The product aims to predict player class to help soccer managers decide whether to buy the player. In the future, a more comprehensive radar plot of player attributes would be more helpful.



- Although a soccer player has many attributes, only some of them are decisive to the “quality” of the player.
- Despite of predictive power, visualization is also an important part for a model. In this case, decision tree performs better.



# Recommendations

Epic	Story	Status
Product Building	Web APP Building	To be Done
Product Building	Local to AWS Transition	To be Done
Product Refinement	Product Test	To be Done
Product Refinement	Product Refinement	To be Done
Product Refinement	Product Summary	To be Done

# Thank you for watching!