

# NBA Player Positions

Jason Katz

Brown University

December 6, 2019

[github.com/jasonk33/nba-player-positions](https://github.com/jasonk33/nba-player-positions)

# Recap

- Understand difference between the 5 distinct NBA positions
- This is a classification task, predicting a player's position based on their seasonal statistics
- Data was scraped from [basketball-reference.com](https://www.basketball-reference.com)
- After preprocessing, there were 10,000 data points
- There is close to an even breakdown the positions

# Cross Validation

- Split the data into 80/20 for cross validation vs holdout set
- Repeated Stratified K Fold cross validation
  - 5 folds
  - 3 repeats
  - Stratification to keep the 5 classes balanced
- Use grid search for parameter tuning

# Cross Validation

- 5 different ML algorithms (tuned parameters)
  - Logistic regression (C)
  - Random forest (max\_depth, min\_samples\_split)
  - K nearest neighbors (n\_neighbors)
  - Gradient boosting (max\_depth, n\_estimators, learning\_rate)
  - Support vector machine (C, gamma)

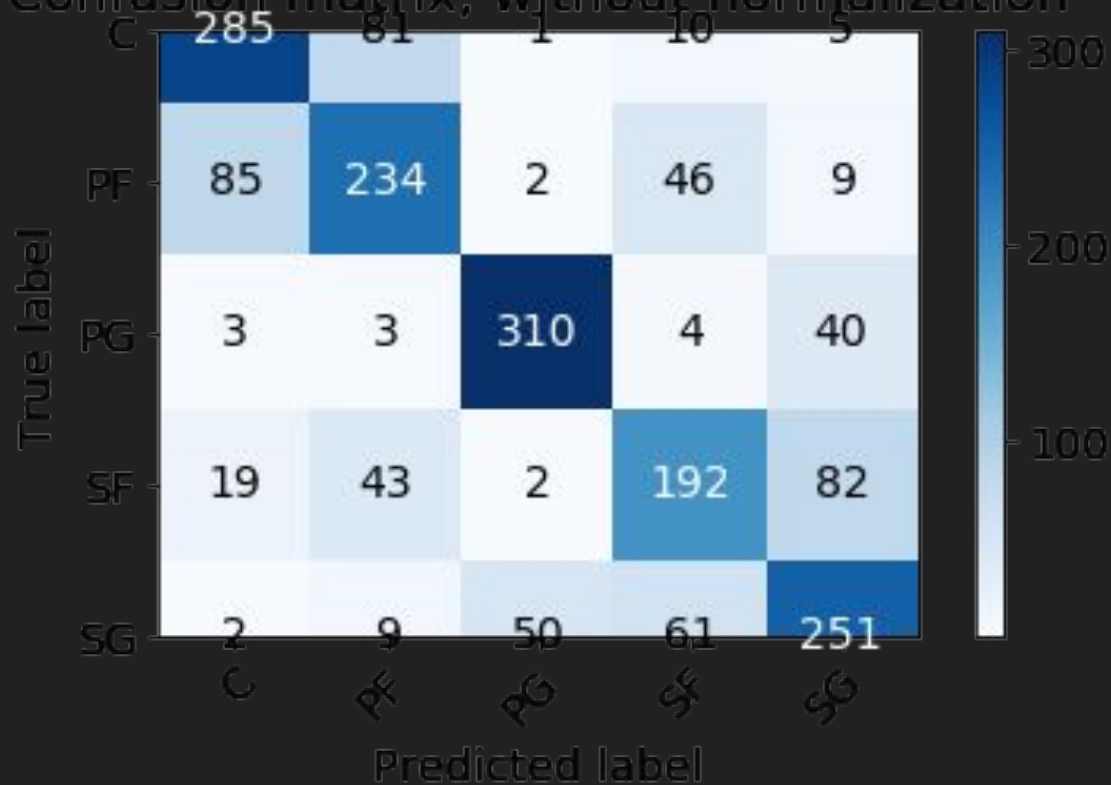
# Results

	Cross Validation Accuracy/Log Loss	Validation Accuracy
Logistic Regression	.622 / .948	.615
Random Forest	.638 / .886	.636
K Nearest Neighbors	.635 / .921	.631
Gradient Boosting	.662 / .822	.686
<b>Support Vector Machines</b>	<b>.679 / .762</b>	<b>.695</b>

Baseline model (balance) : 0.209

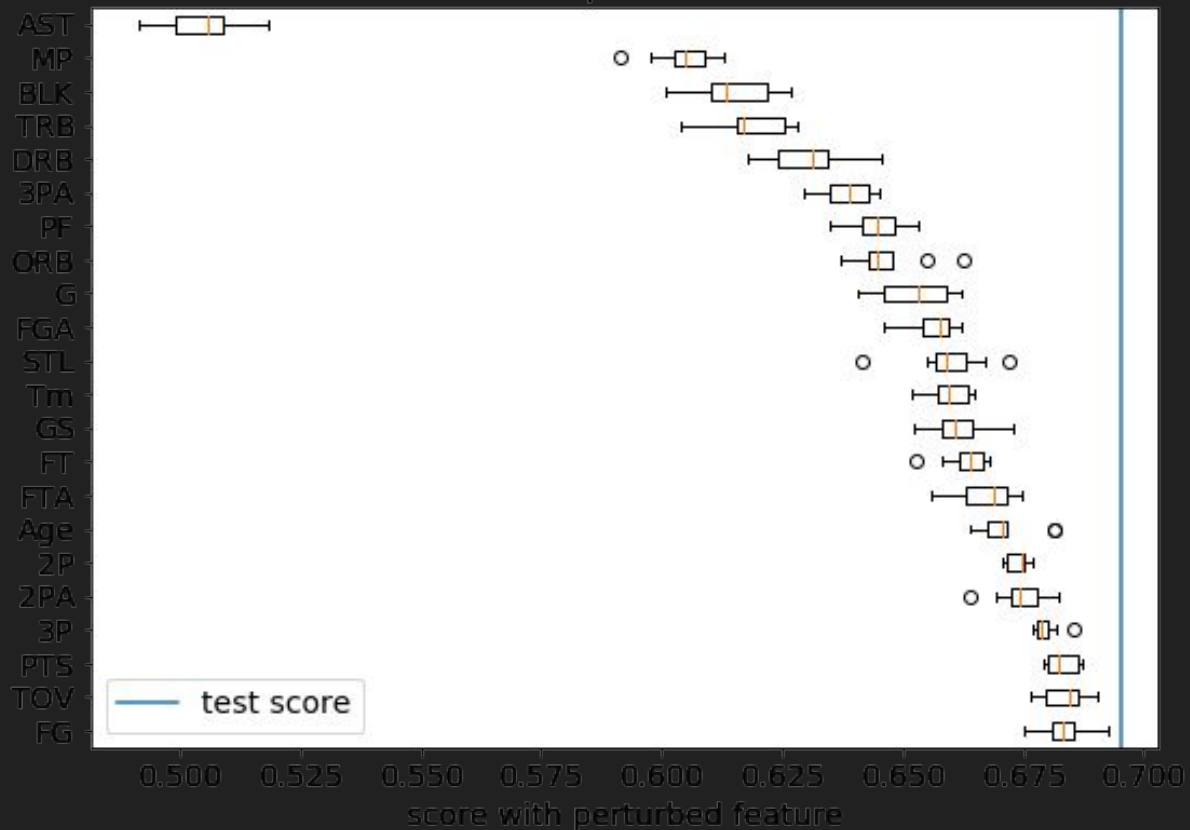
# Results

Confusion matrix, without normalization



# Results

Permutation Importances (test set)



# Results



- This figure above depicts local feature importance for an example of a center, where the model predicts it correctly (values for center)
- This figure below depicts local feature importance for an example of a center, where the model predicts it correctly (values for point guard)





# Outlook

- Experiment with other types of models, such as neural networks
- See if there are any other hyperparameters I could tune, possibly different types of kernels for the support vector machine model
- Acquire more data in addition to seasonal statistics, such as salary information or player details such as height and weight
- Experiment with different feature engineering to combine different stats together in different ways.