NBA COMBINE ANALYSIS

by Ian Stack

February 6, 2023

PROBLEM STATEMENT

01. Vision

Utilizing the stats from the NBA combine via stats.nba.com, I will construct a regression model to predict the target variable of Win Share per player based off different events in the NBA combine.

02. Mission

Model performance will be determined by RMSE (Root Mean Squared Error) in order to determine the most accurate prediction of a player's Win Share.

PROJECT ROADMAP

Data Collection & Cleaning

- NBA.com/stats
- http://basketball-reference.com/
- Address missing values

EDA

- Trends
- Patterns

Modeling

- Linear, Lasso, Ridge Regression
- Decision Trees
- Random Forest

Conclusion

- Prediction of Win Shares
- Significant Features

DEFINITIONS

RMSE

Measure of error(in terms of Win Share metric)

Win Share

“a player statistic which attempts to divvy up credit for team success to the individuals on the team.” (http://basketball-reference.com/)

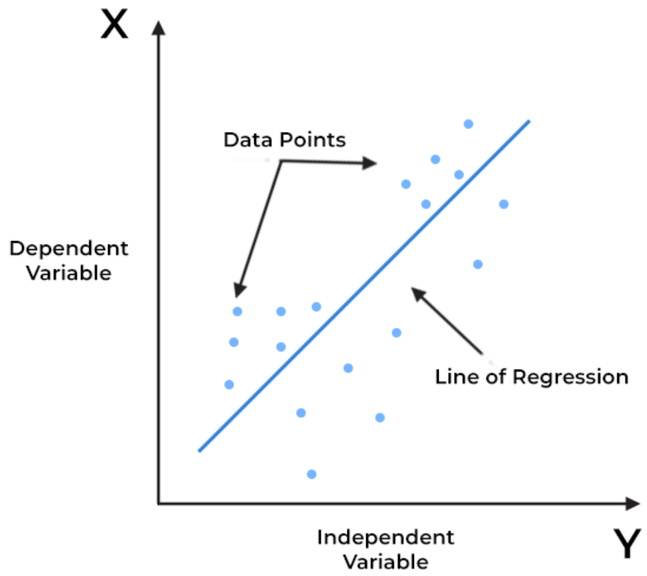
NBA Combine

Multi-day showcase where players go through various drills and tests for an audience of NBA coaches and scouts.

Machine Learning Regression

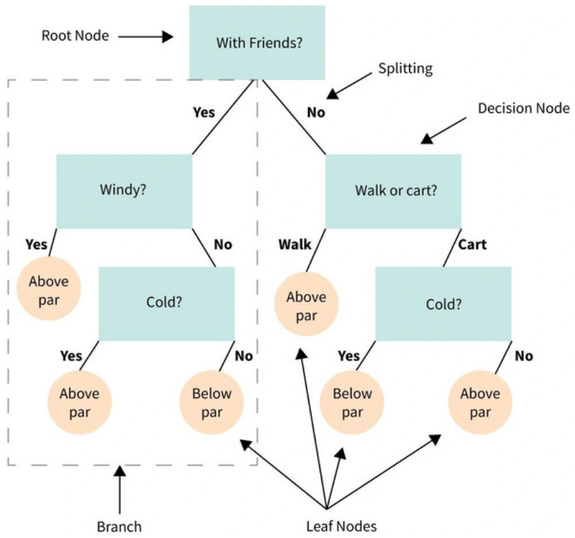
a statistical technique that relates a dependent variable to one or more independent (explanatory) variables.

Models



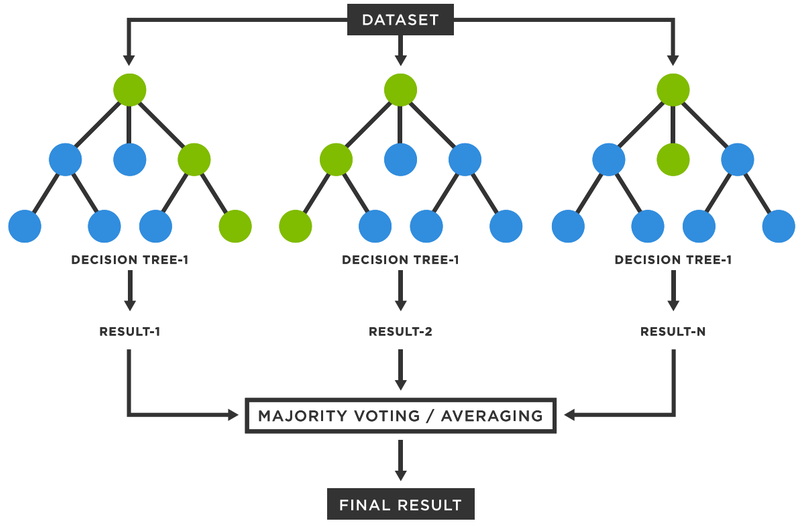
Linear Regression

Establishing a relationship between the dependent and independent variable



Decision Trees

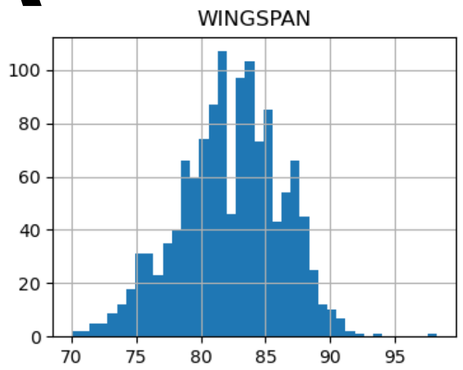
Takes a dataset, finds rules based on the X data and splits data into smaller datasets

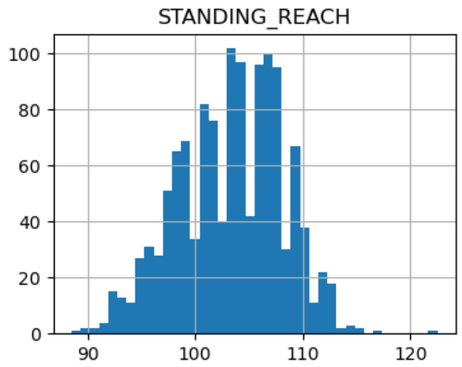


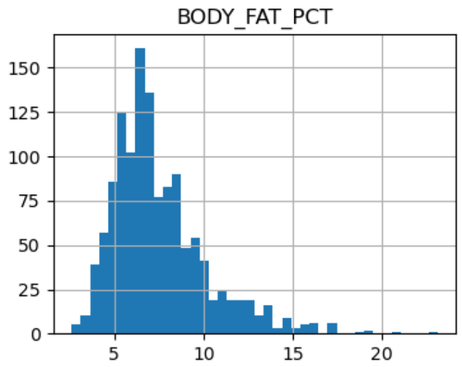
Random Forest

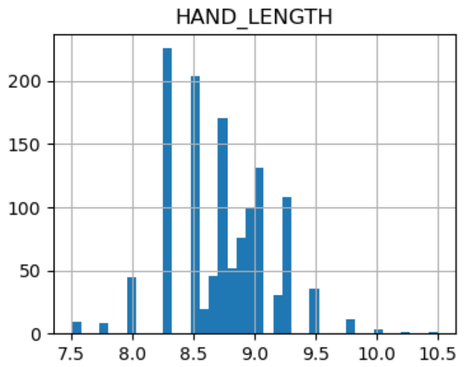
A number of decision trees on various subsamples of datasets

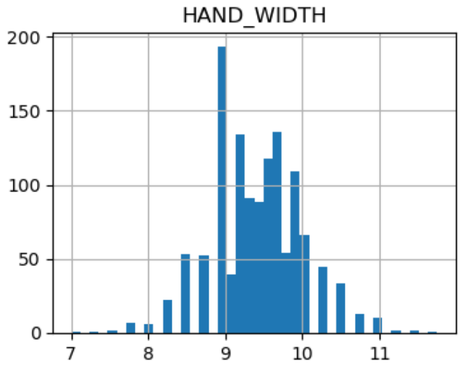
EDA

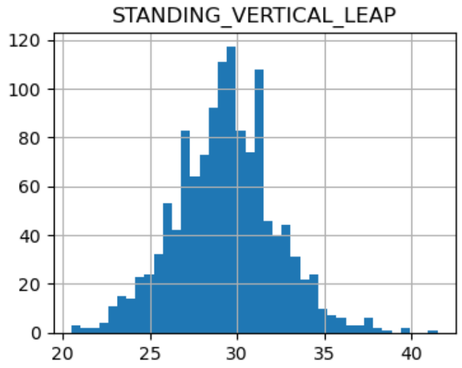


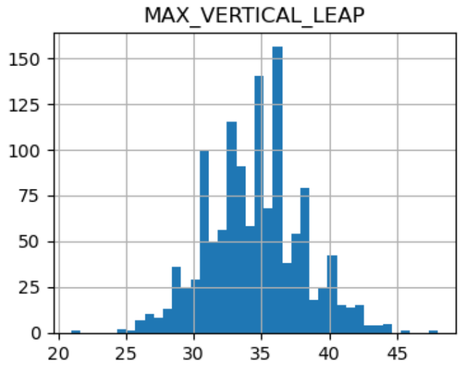


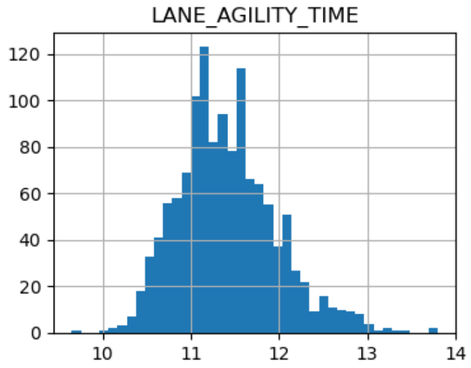


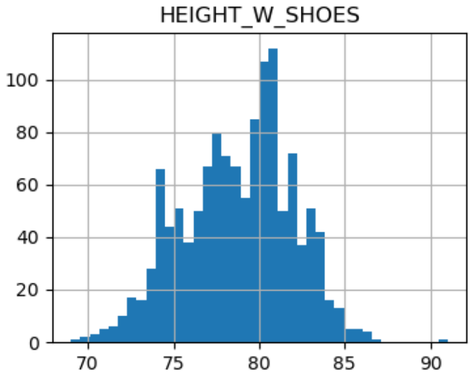


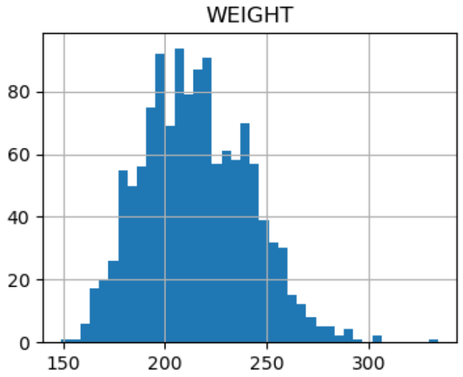










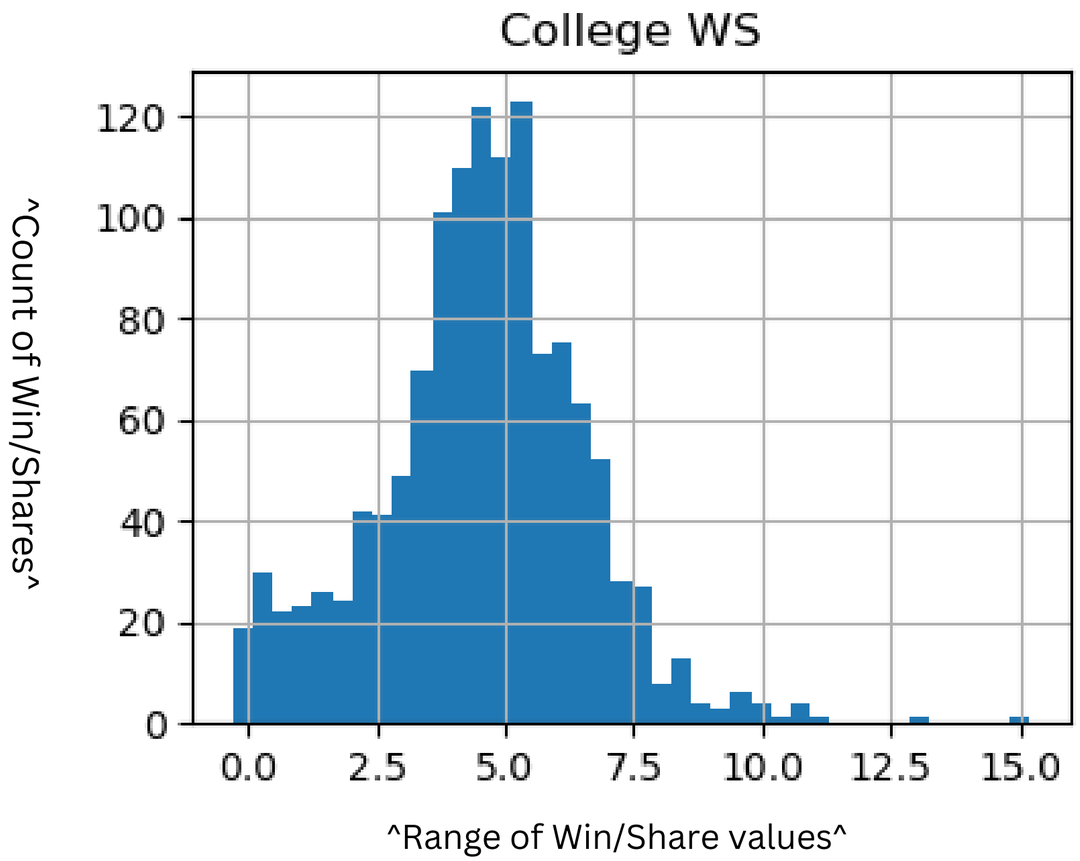


y-axis: count

x-axis: range of feature

-normal distribution

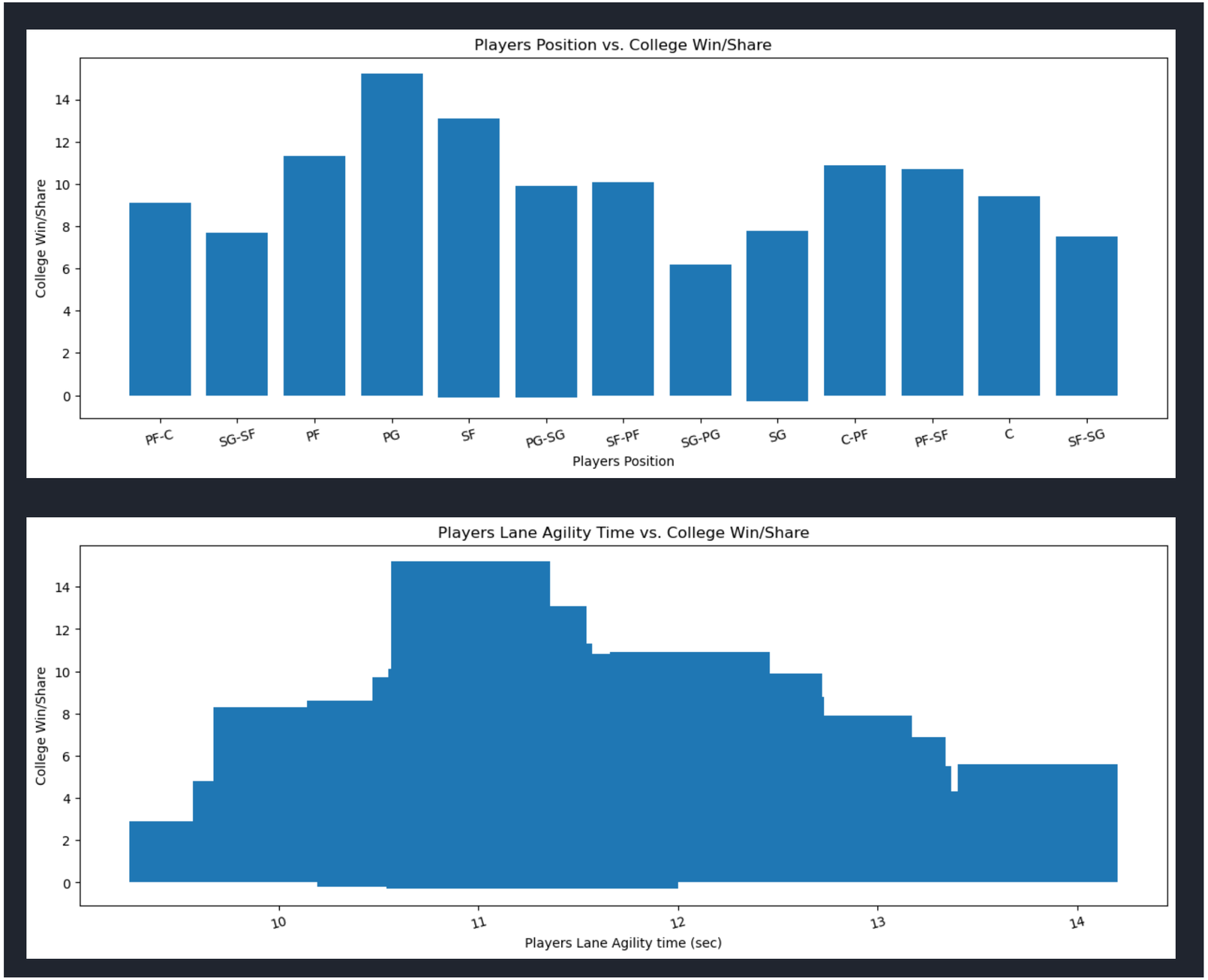
TARGET VARIABLE DISTRIBUTION

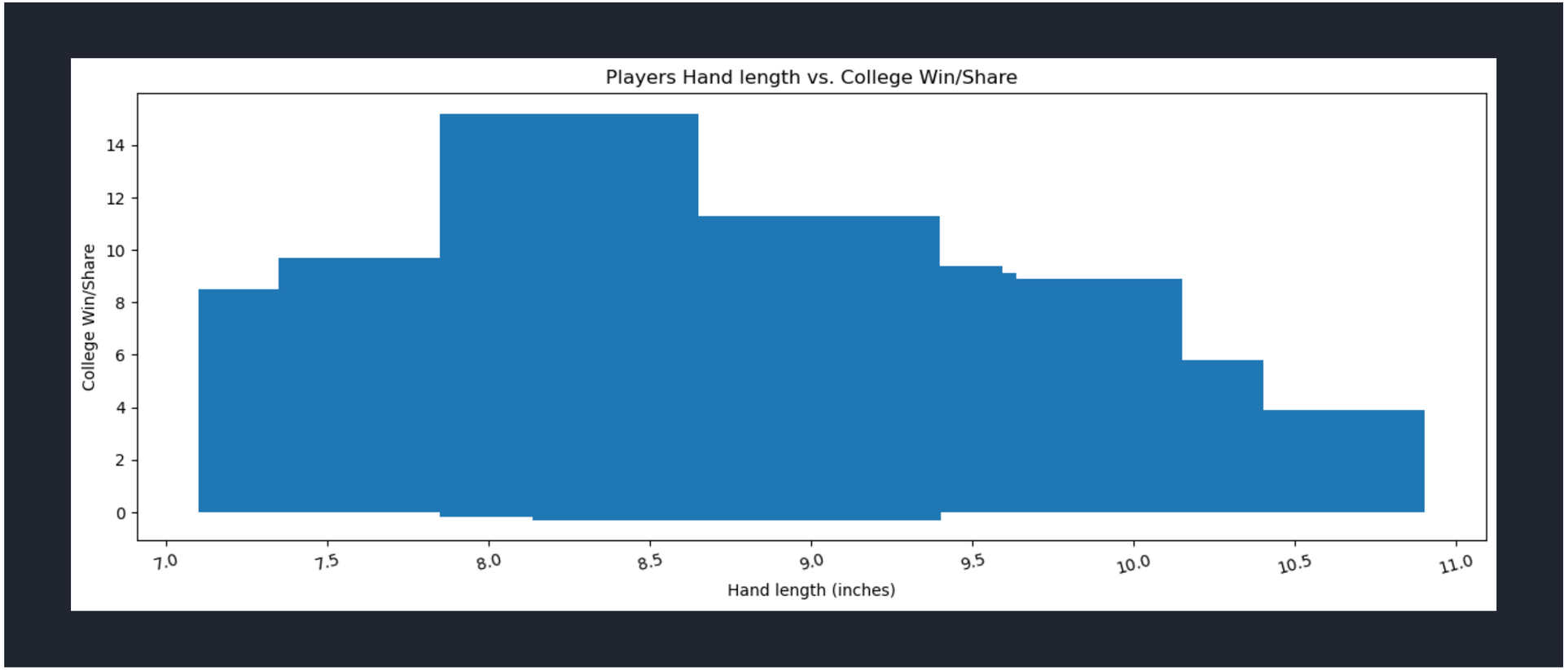


What factors increase Win Share?

Features:

position and
lane agility were two features
that highly affected
Win Share





Initial Hypothesis:

The larger the hand size, the better the player.

Not True

MODELING

6 Models Used

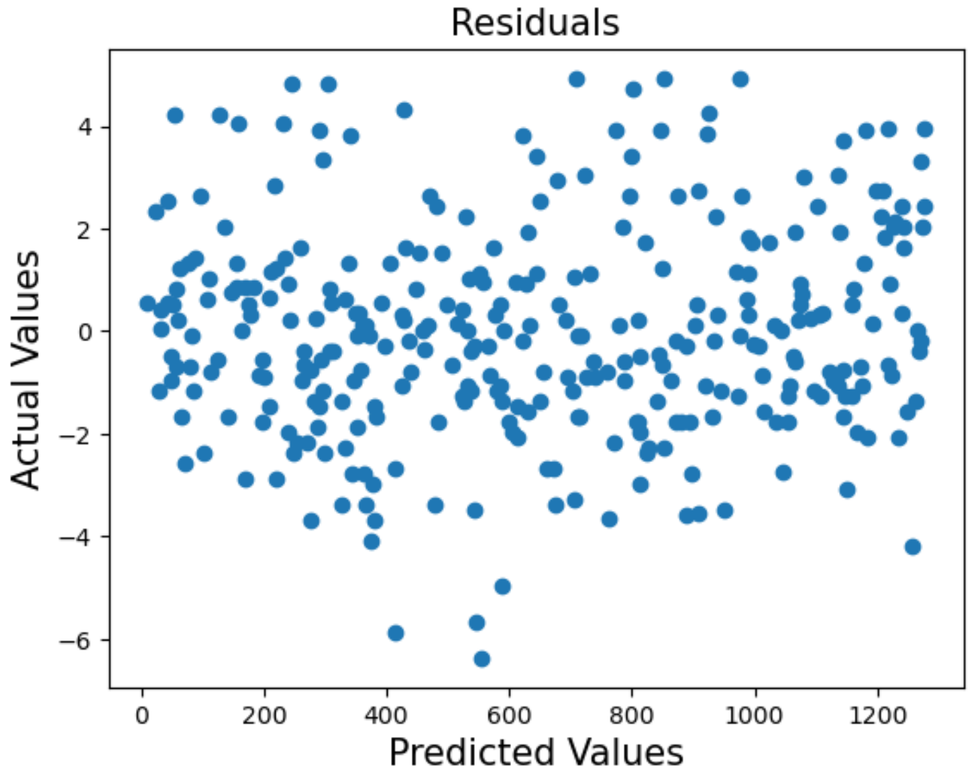
- Linear Regression
- Ridge Regression
- Lasso Regression
- Decision Tree
- Decision Tree: Gridsearch
- Random Forest: Gridsearch

Top 3

1. Linear Regression
2. Random Forest w/ Gridsearch
3. Decision Tree w/ Gridsearch

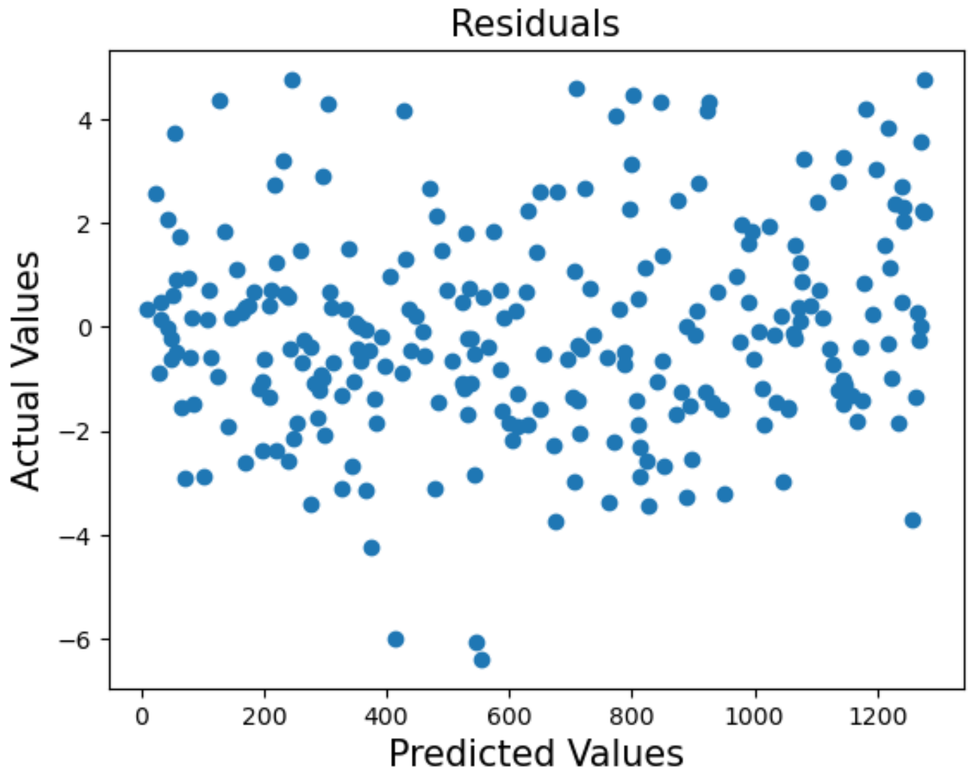
DECISION TREE W/ GRIDSEARCH

TRAINING RMSE	TESTING RMSE
1.870	2.032

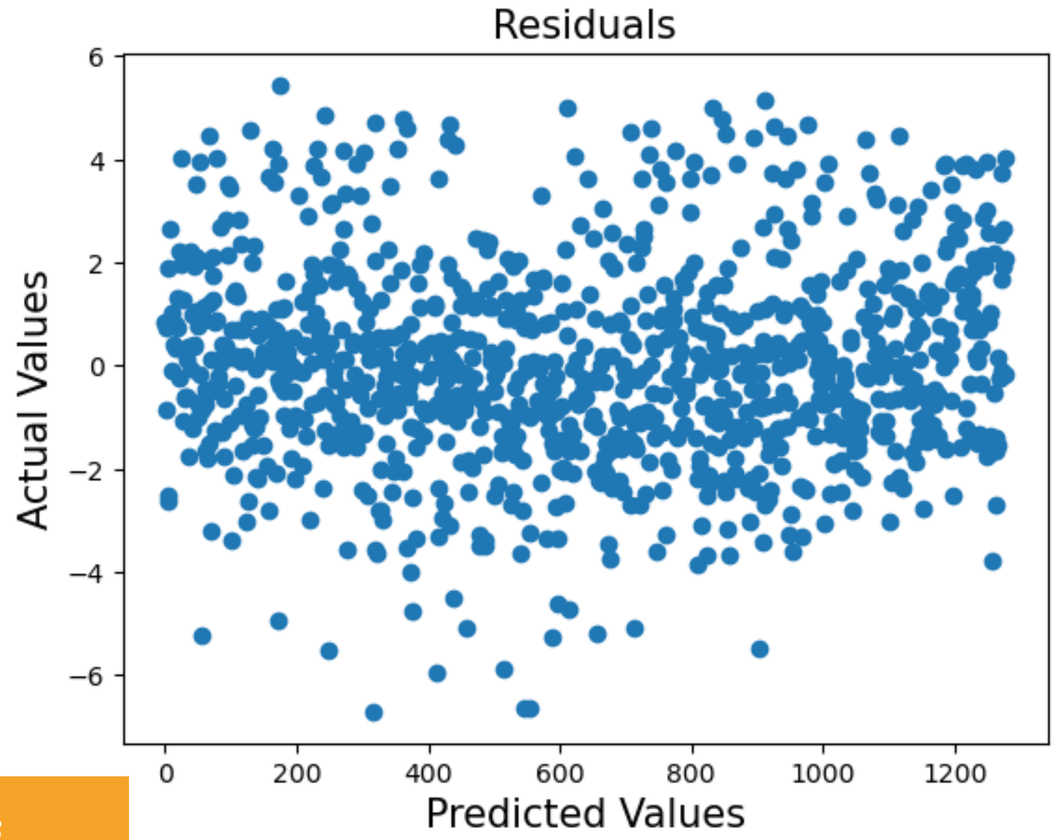


RANDOM FOREST W/ GRIDSEARCH

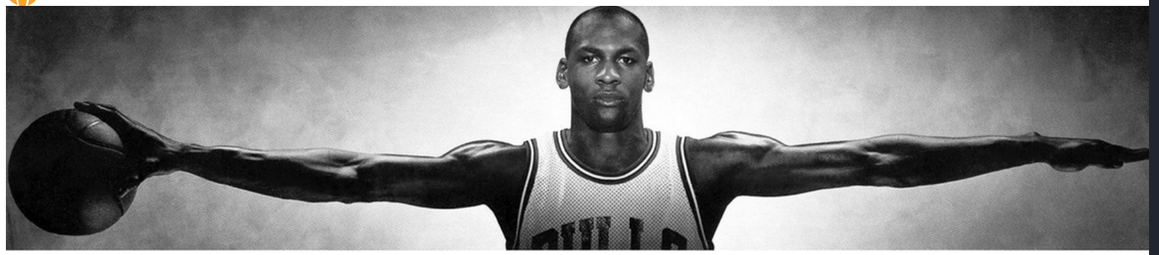
TRAINING RMSE	TESTING RMSE
1.870	1.998



LINEAR REGRESSION



TRAINING RMSE	TESTING RMSE	BASELINE
2.099	1.944	1.974



FEATURE IMPORTANCE

	feature importances
HEIGHT_W_SHOES	0.000000
WEIGHT	0.386307
WINGSPAN	0.376680
STANDING_REACH	0.000000
BODY_FAT_PCT	0.000000
HAND_LENGTH	0.000000
HAND_WIDTH	0.000000
STANDING_VERTICAL_LEAP	0.000000
MAX_VERTICAL_LEAP	0.000000
LANE_AGILITY_TIME	0.237013

TEST DATA OUTPUT

Three Created Players

1. **Player10 10:**
   a. position: PG
   b. Dec. Wt., Inc. Agility, Inc. Vertical
   c. A light, quick, point guard

1. **Player11 11:**
   a. position: C
   b. Inc. Hand size, Inc. Body Fat Pc.t, Inc. Vertical
   c. A shorter bigger center

1. **Player12 12:**
   a. position: SG
   b. Inc. Weight, Inc. Wingspan
   c. A larger & longer shooting guard

Predicted Win Share	POSITIONS
4.532662	PG
4.229131	PG
4.577730	SG
4.087634	SG
4.559291	SF
4.676197	SF
4.577238	PF
4.010440	PF
4.535112	C
4.[illegible]	C
4.408451	PG
5.847581	C
4.651049	SG

CONCLUSION

01

Point Guard and Center are valuable/important positions in basketball

02

NBA Combine events:

- Wingspan
- Weight
- Lane Agility Time

Are best predictors

01

High School

With more elite prep basketball school's each year, the talent and expectations grow.

These schools can evaluate players based off physical traits to help project their potential

02

College

With the introduction of NIL (Name, Image, and Likeness) deals, college basketball is much has become a larger business

This model can help analyze high school players and if they can help them win.

03

Professional

The NBA is a lucrative business with players accumulating large amounts of money.

This model can help predict potential of players while also adjusting with a team's budget

APPLICATIONS

SOURCES

- https://www.si.com/basketball
- https://www.gettyimages.com/photos/nba
- https://www.javatpoint.com/linear-regression-in-machine-learning
- https://www.mastersindatascience.org/learning/machine-learning-algorithms/decision-tree/
- https://www.tibco.com/reference-center/what-is-a-random-forest


NBATickets.com
THANK YOU