BrainStation

Prepared by Anastassia Troitskaia

September 27 2024

College 2 Contract

Predicting NBA Salaries for College Athletes

Deep diving into our data with clear direction

Content

Introduction

Purpose:

To predict NBA salaries for college basketball players based on their performance statistics and historical data, aiding NBA teams in draft and salary decisions.

Description:

This project tackles the challenge of translating college performance into professional success, using machine learning and comprehensive data from current and former NBA players to improve player evaluation and financial planning.

Back to Content Page

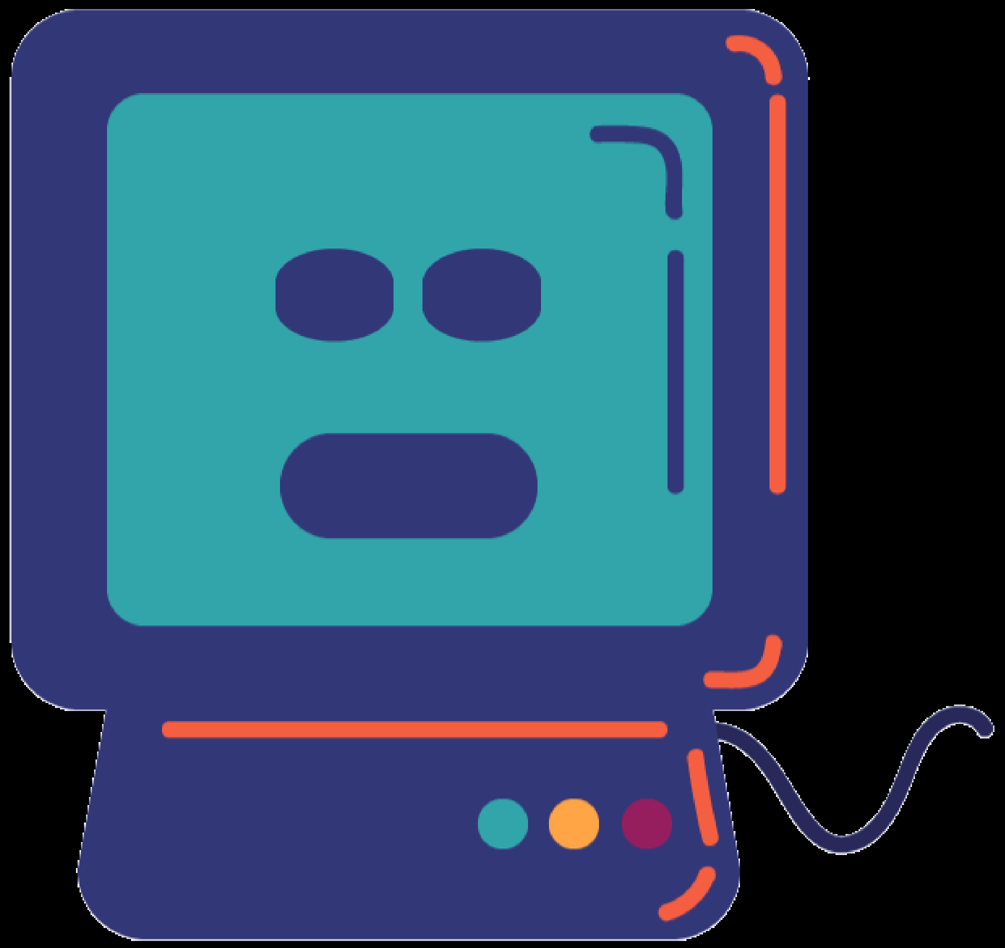
Methodology

- Data Collection
  - Web Scraping with Selenium
  - CSV Download
- Cleaning & Preprocessing
  - Impute missing values
  - Convert, encode values
- Model building
  - Model Selection
  - Feature Engeneering
  - Training and Testing

https://hoopshype.com
https://www.nba.com
https://www.sports-refere nce.com

Back to Content Page

Issues



Issues

- Missing Values in many different categories
- Data has a lot of issues with data types
- Many values were entered as string object and etc.

How was it handled?

- Lot’s of data cleaning had to be performed
- Missing values imputation
  - KNN imputation
- Convert all the data types from string object
- Data such as Draft Team, College Name had to be encoded

Back to Content Page

Features Overview

Game Statistics

- PTS (Points): Key offensive statistic that often correlates with a player's value and potential salary.
- G (Games Played): This indicates durability and consistency, which are valuable traits for NBA teams.
- MP (Minutes Played): Shows how much a player is trusted on the court, which can be indicative of their skill level and potential value.
- FG% (Field Goal Percentage), 3P% (Three-Point Percentage), FT% (Free Throw Percentage): These shooting percentages are crucial indicators of a player's offensive efficiency.
- TRB (Total Rebounds): Rebounding is a valuable skill that contributes to both offense and defense.
- AST (Assists): Important for guards and indicates playmaking ability.
- STL (Steals) and BLK (Blocks): These defensive stats can significantly impact a player's value.
- Draft Year: This could potentially capture trends in salary over time or how recent draft picks are valued.

Player Physical Statistics

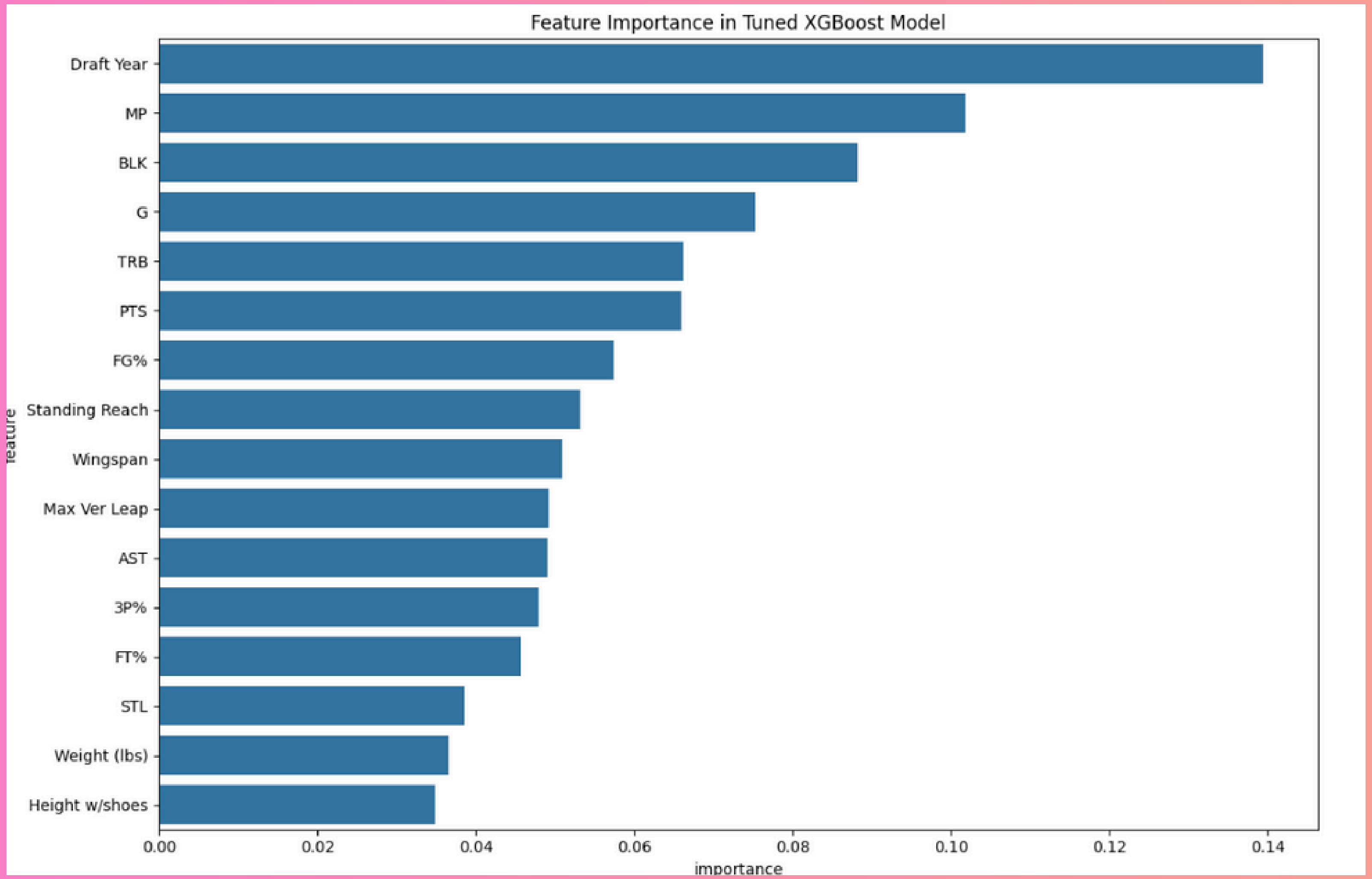
- Height: Important for positioning, especially for centers and forwards.
- Wingspan: Indicates defensive reach and ability to block shots or intercept passes.
- Weight: Shows a player's bulk, which can affect physicality in the game.
- Vertical Leap: Measures jumping ability, critical for rebounds and scoring near the basket.
- Agility and Speed: These metrics are tested in combine drills like the lane agility drill and three-quarter sprint.
- Standing Reach: Measures how high a player can reach without jumping, important for defense and rebounding.

Modeling: Iteration I

Model	Accuracy	Error Amount ($)
Linear Regression	~24.17%	*$1,577,872*
Random Forest Model	~24.17%	*$1,507,924*
XGBoost Optimized	~30.74%	*$1,501,051*

Modeling: Iteration I

XGBoost Optimized



Feature Importance

- **Draft Year**
- **MP (Minutes played): Career college**
- **BLK (Blocks): These defensive stats can significantly impact a player's value.**
- **G Games**
- **TRB (Total Rebounds): Rebounding is a valuable skill that contributes to both offense and defense.**

Modeling: Iteration II

Feature Engeneering

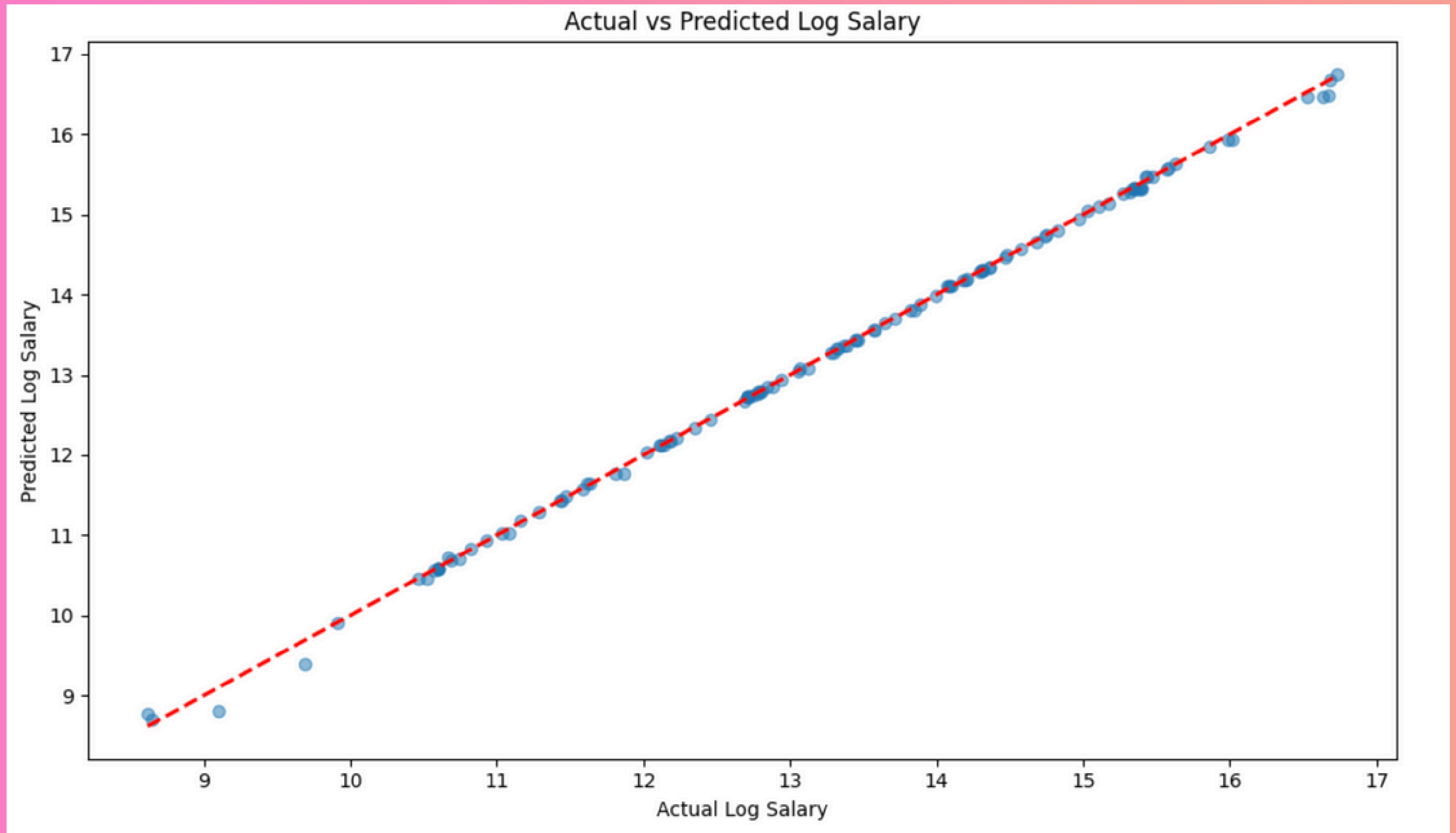
- Efficiency metrics
- Shooting efficiency
- College Experience
- College Experience Binning
- Physical Attributes
- Draft Position Binning
- Performance Binning
- Combine Shooting Percentages

- Per-minute statistics for points, assists, and rebounds
- Calculated true shooting percentage.
- Derived years played in the NBA.
- Physical Attributes: Calculated BMI and wingspan-to-height ratio.
- Draft position binned into categories.
- Performance metrics (points, assists, rebounds) binned into quintiles.
- Advanced stats (PER, Win Shares) binned into quintiles.
- Overall Shooting Percentage: Combined FG%, 3P%, and FT%. transformation to salary data

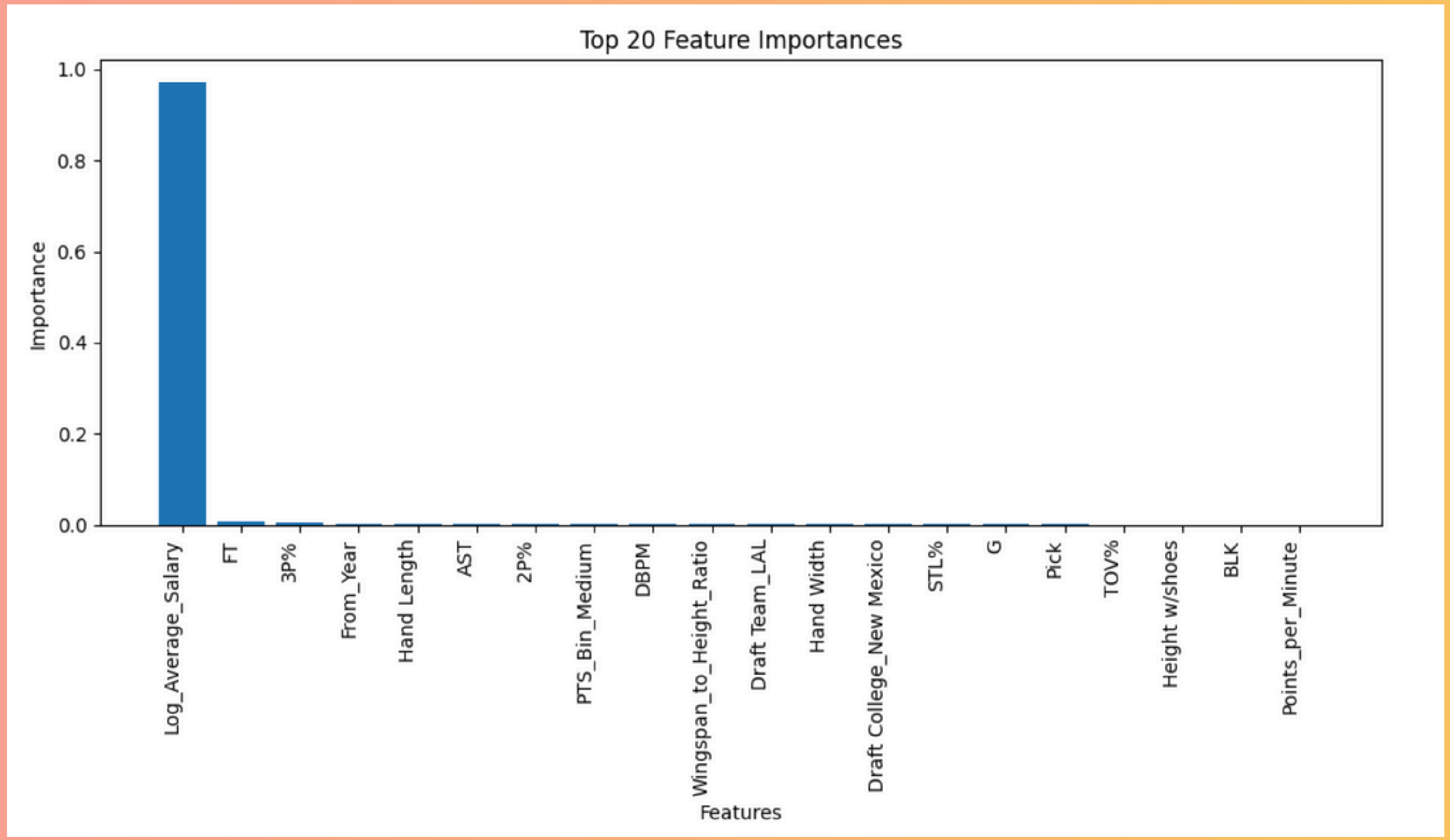
Modeling: Iteration II

Model	Accuracy	Error Amount ($)
XGBoost (Average Salary)	~99.91%	$55,320
XGBoost (Year 1)	~86.12%	$448,100
a. XGBoost b. Random Forest c. Neural Network d.Ensemble	a. ~98.00% b. ~98.00% c. - 100.13% d. ~ 74.00%	a. $253,718.96 b. $264,719.50 c. $2,644,677.01 d. 927,035.58

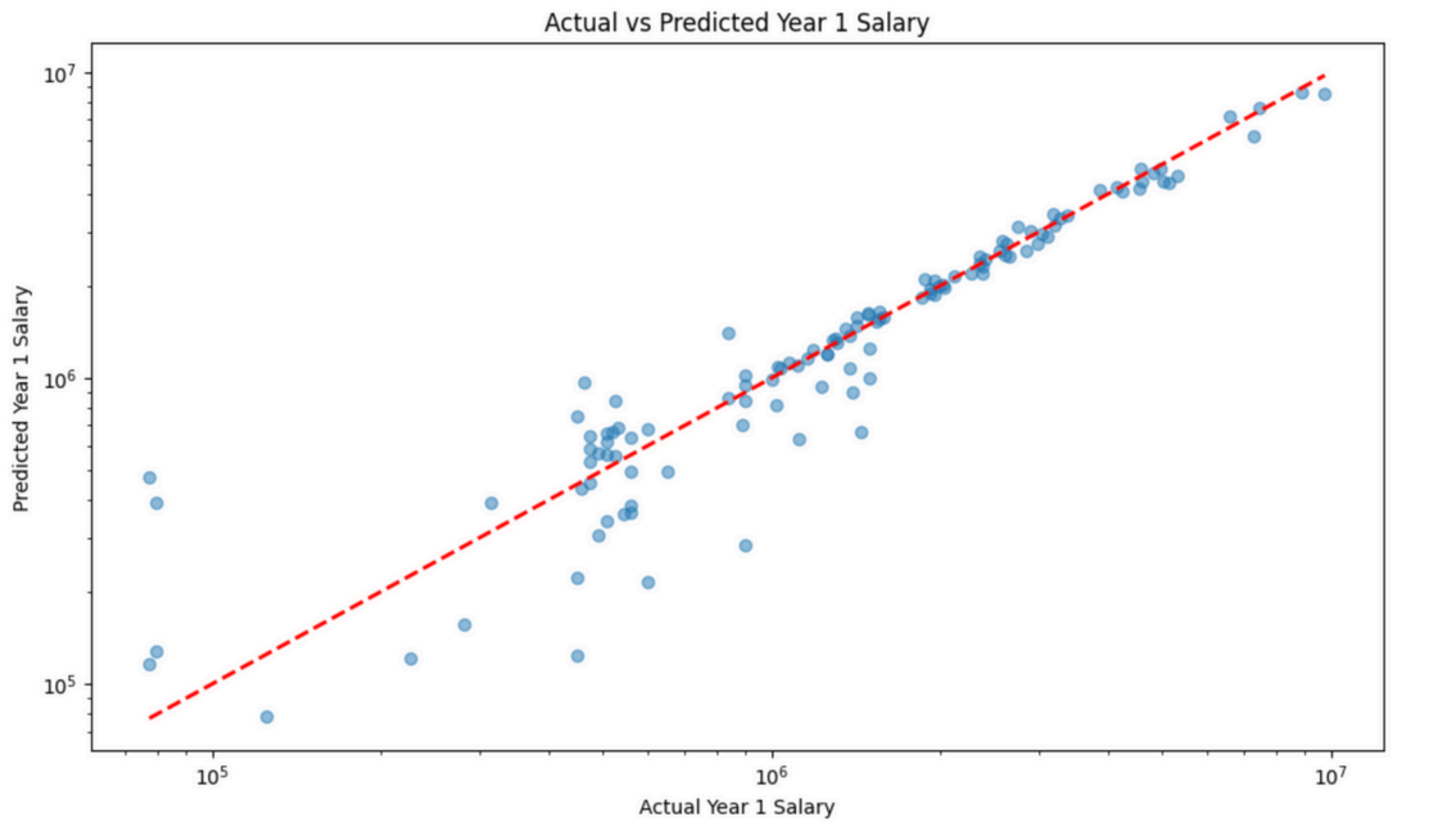
XG Boost (Average Salary)



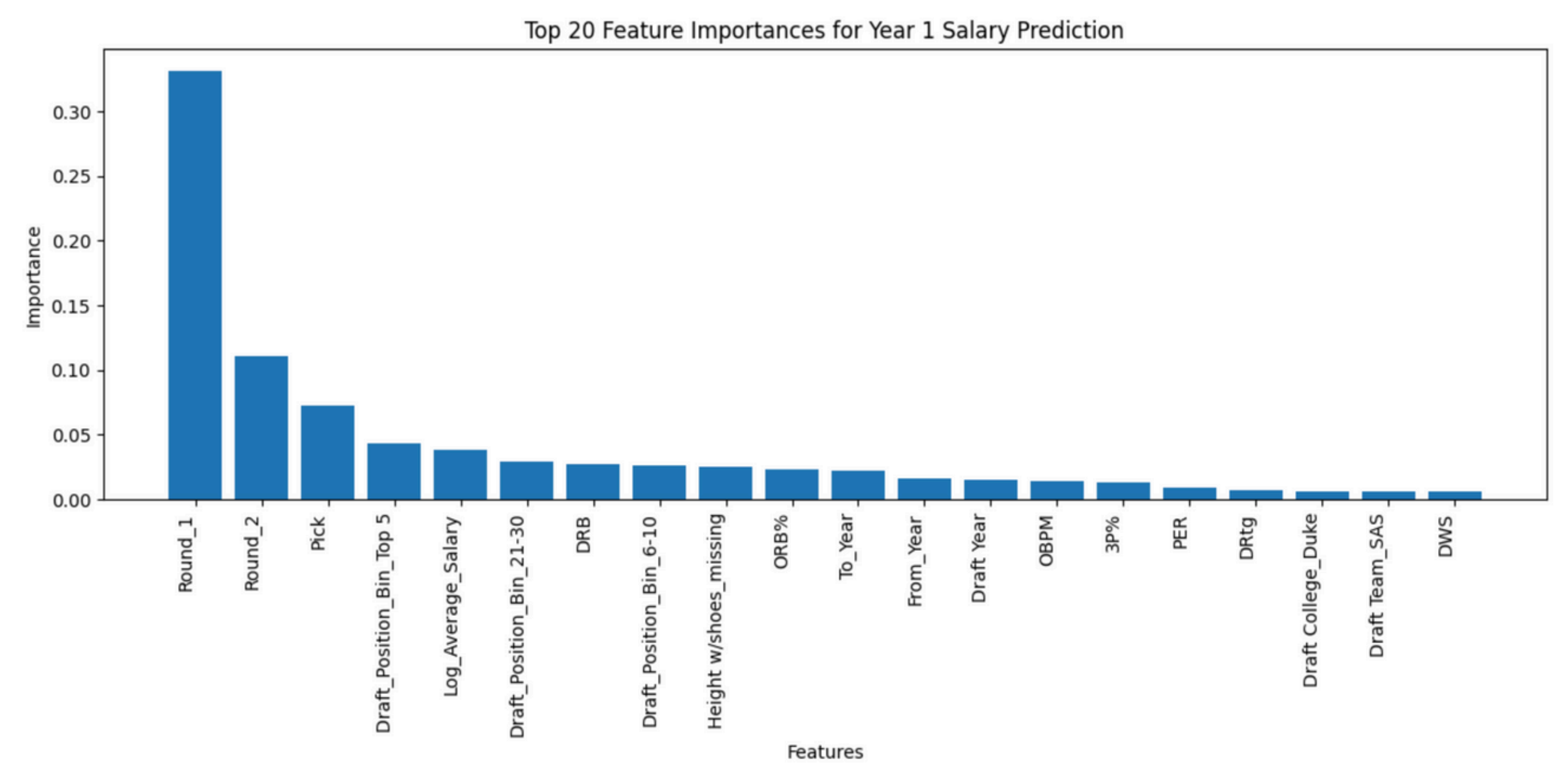
Feature Importance



XG Boost (Salary Year 1)



Feature Importance



Prediction Test of XGBoost

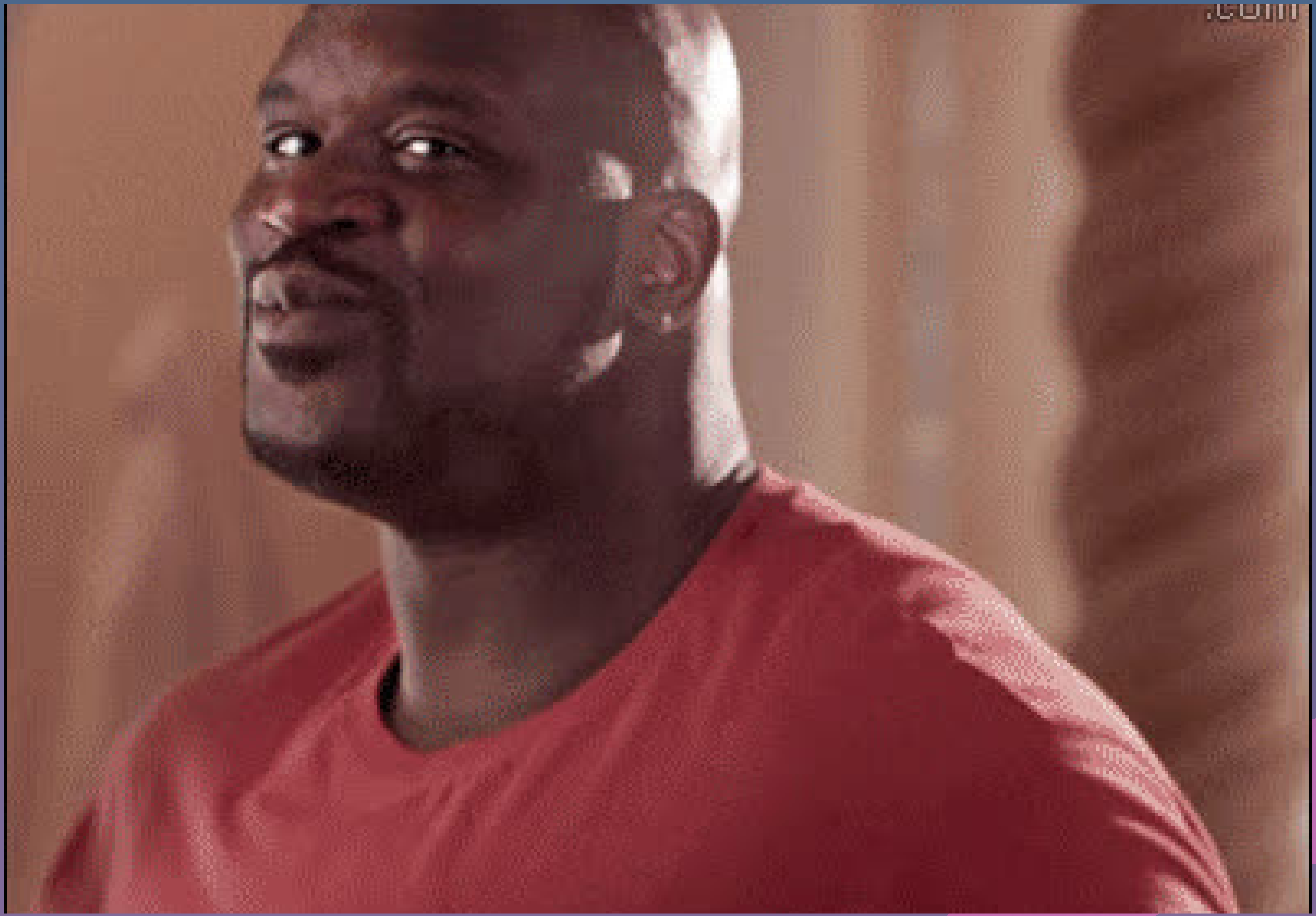
	Player	Actual Salary	Predicted Salary	Error (%)	PTS	Draft Position	College Experience
0	Gordon Hayward	2,356,320	2265243.00	4	572.0	6–10	Freshman
1	Andrew Nicholson	1,418,160	1497211.00	6	1727.0	11–20	Junior
2	Thomas Robinson	3,374,640	3639228.75	8	1026.0	Top 5	Junior
3	Tyler Lydon	1,874,640	1803886.50	4	823.0	21–30	Sophomore
4	Scottie Barnes	7,280,400	6419172.50	12	248.0	Top 5	Freshman
5	Nigel Williams–Goss	1,500,000	1222843.50	18	1536.0	31+	Senior
6	Terrence Jones	1,485,000	1573754.38	6	1064.0	11–20	Sophomore
7	Dwayne Bacon	1,378,242	1215589.25	12	1139.0	31+	Sophomore
8	Filip Petrusev	559,782	566106.94	1	783.0	31+	Sophomore
9	Meyers Leonard	2,126,520	2130876.50	0	502.0	11–20	Sophomore

Conclusion & Findings

- **Project Goal:** Predict NBA player salaries using machine learning models
- XGBoost outperformed other models:
  - ~99.91% accuracy for average salary
  - ~86.12% accuracy for Year 1 salary
- Optimized XGBoost (30.74%) surpassed Linear Regression and Random Forest (24.17%)
- Lowest error: XGBoost for average salary ($55,320)
- Draft pick number most crucial for Year 1 salary prediction
- High accuracy in individual player predictions (most errors <12%)
- Model performance robust across various draft positions and college experience levels
- XGBoost model valuable for teams, agents, and analysts in estimating player value
- Draft position crucial in determining initial NBA contracts
- **Future work:** Optimize model further, explore additional features, analyze across eras/positions

[Back to Content Page](#)

Thank You!



Back to Content Page