Generating Betting Insight on NBA Player Performance using Machine Learning

Group 42: Sahil Bishnoi, Josh Garretson, Avery Girsky, Oliver Hewett, Hardik Patel, Atticus Rex

Overview

We present an interactive dashboard to inform statistically optimal basketball betting strategy and describe trends in player-level data. We hope the use of interactive visualization around ML prediction should help to distill advanced data science concepts for everyone to receive data-driven advice to assist their intuition.

- **Modeling** To identify profitable bets given historical professional basketball data, our team trained a model outputting a probability distribution of the five main player stats: Points (PTS), Assists (AST), Rebounds (REB), Blocks (BLK), Steals (STL).
- **Visualization** These insights are displayed in a digestible and interactive formats allowing for user inputs to observe probability distributions for model predictions.

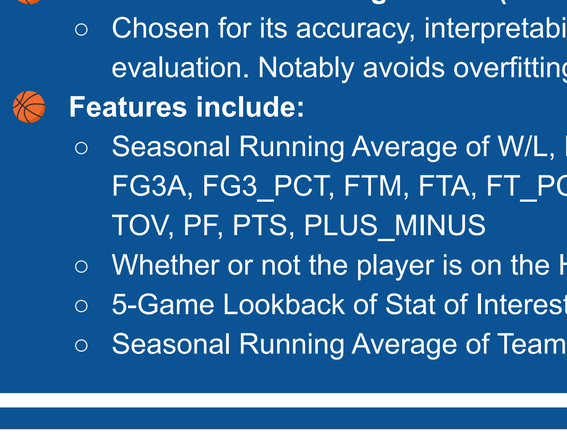
Modeling

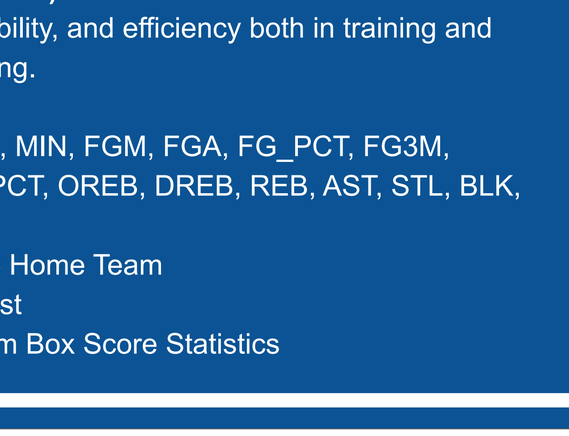
- **Data Summary:**
  - 142,320 Observations
  - All Active NBA Players and Regular Season Games since 2003
- **Models Evaluated:**
  - Ridge/Lasso Regression
  - Neural Networks
  - K-Nearest Neighbors Regression
  - AdaBoost and Random Forest Regression

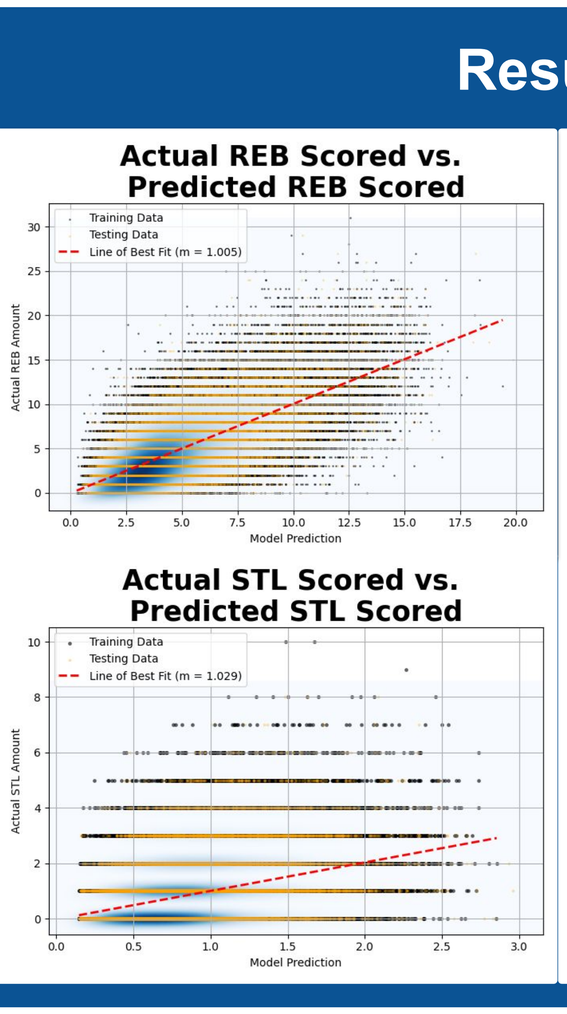
Test MAE By Model, Statistic	PTS	REB	AST	STL	BLK
Ridge Regression	4.866	2.031	1.402	0.570	0.744
Lasso Regression	4.866	2.001	1.401	0.572	0.745
Neural Network	**4.589**	**1.998**	**1.399**	**0.561**	**0.731**
KNN Regression	4.943	2.101	1.538	0.610	0.754
AdaBoost	4.712	2.062	1.477	0.590	0.746
Random Forest	4.640	2.048	1.492	0.577	0.737

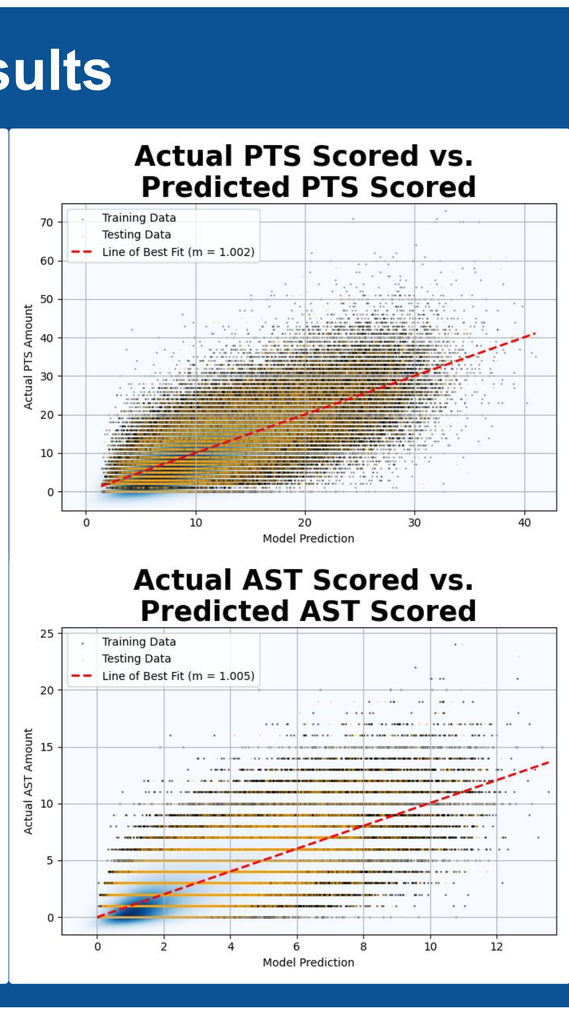
- **Final Model: Lasso Regression (α=9e-3)**
  - Chosen for its accuracy, interpretability, and efficiency both in training and evaluation. Notably avoids overfitting.
- **Features include:**
  - Seasonal Running Average of W/L, MIN, FGM, FGA, FG_PCT, FG3M, FG3A, FG3_PCT, FTM, FTA, FT_PCT, OREB, DREB, REB, AST, STL, BLK, TOV, PF, PTS, PLUS_MINUS
  - Whether or not the player is on the Home Team
  - 5-Game Lookback of Stat of Interest
  - Seasonal Running Average of Team Box Score Statistics

Results







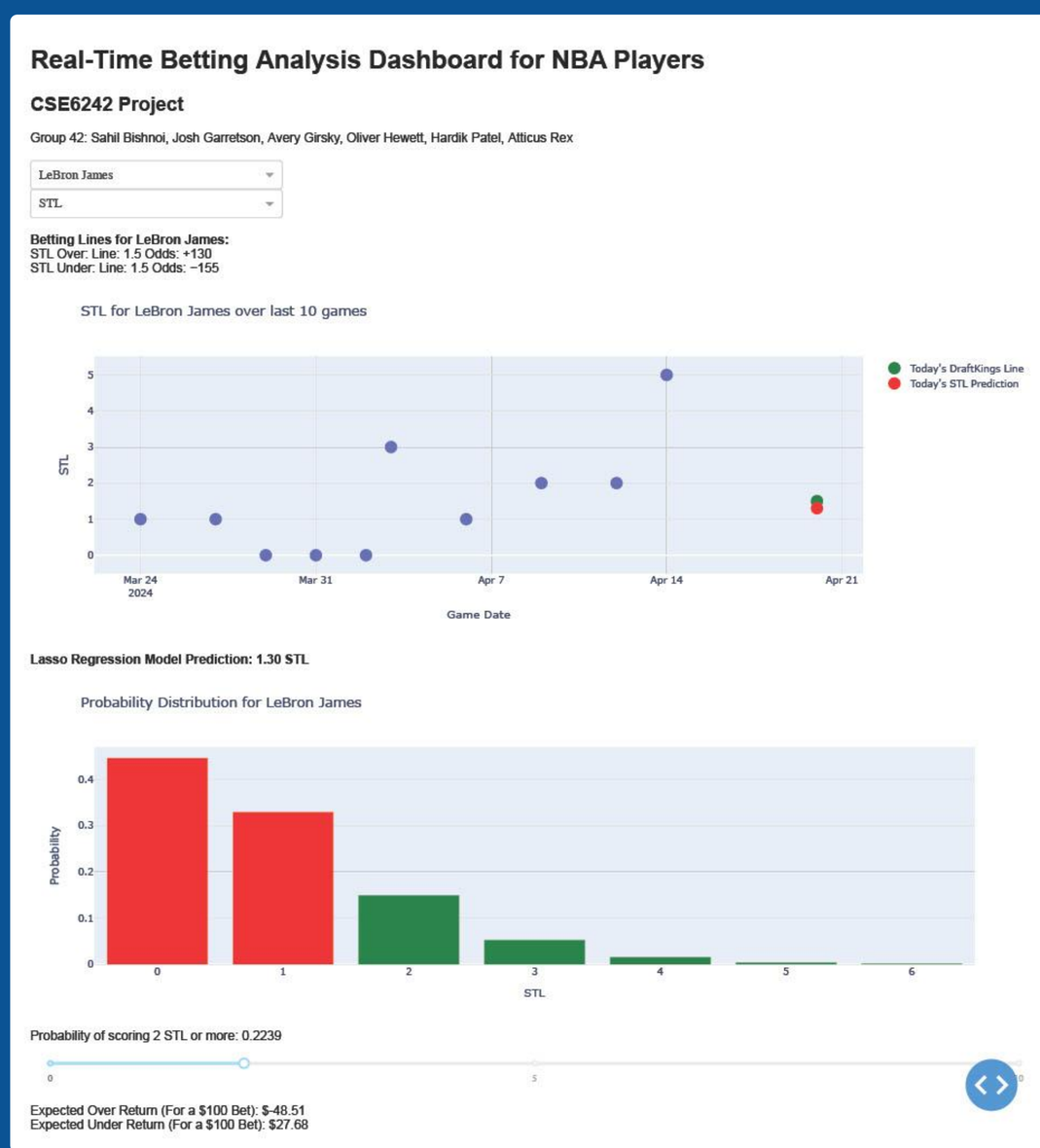


Visualization

Given user-input for a **player** and **statistic**, display two visuals:

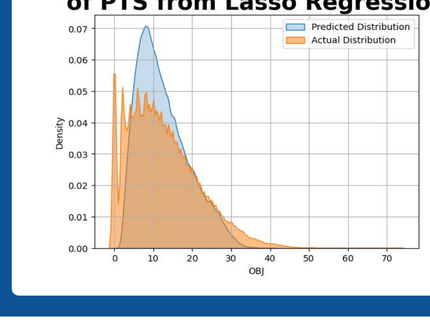
- Time-series plot of the player's statistical output over their last ten games combined with the projection and sportsbook line for their upcoming game
- Distribution plot of historical performances amongst all players that had been previously projected the same statistical outcome by our model
  - Vertical lines included for the model projection and the sportsbook line
  - The proportion of the distribution falling above/below the sportsbook line represents the percent "confidence" of the model's recommendation for an over/under bet

Snapshot



Features

- Scraped lines from DraftKings Sportsbook for each statistic for upcoming game
- Comparison between model prediction, sportsbook line, and current game statistic
- Model prediction to factor in player, team, matchup, and recent performance
- Probability distributions for prediction
- User input slider to observe over/under probabilities for any quantity of statistic
- Expected returns on over/under bets for informed decision-making



Conclusion

The final version of our product was developed using a LASSO regression model due to its balance of accuracy, interpretability, and computational efficiency. By integrating predictive analytics with sports betting, our model enhances users' experiences with online betting platforms through detailed statistical analysis and dynamic visualizations presented via an interactive dashboard. This project highlights the potential of applying machine learning in sports analytics, setting a promising stage for further research and the development of more sophisticated tools at the intersection of sports analytics and sports betting.