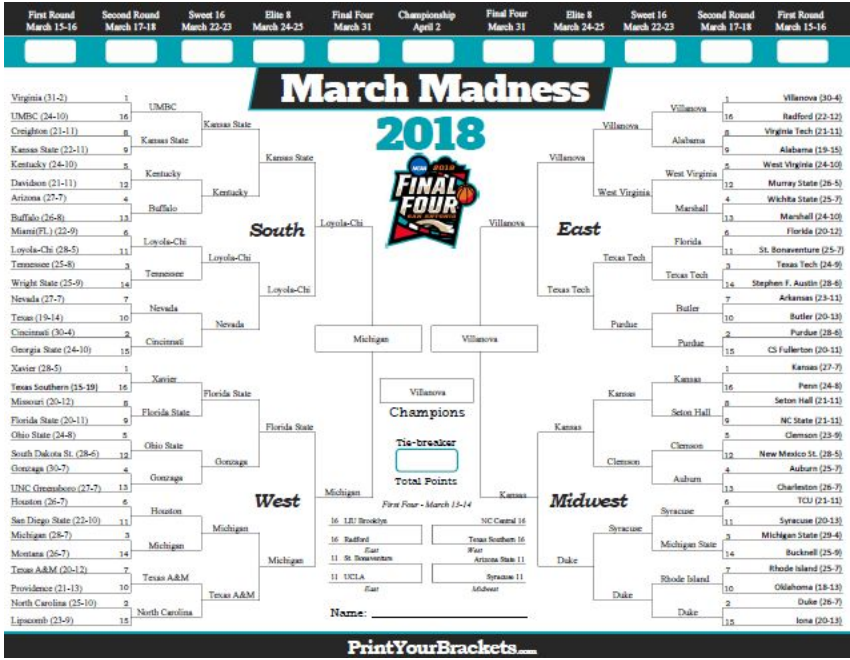
Making Sense of March Madness

Victor Ramirez
Springboard
Capstone Project 2
May 2018

Problem

- The NCAA Tournament is unpredictable.
- The casinos are very good at placing borderline betting lines
- Profiting off of sports betting is extremely difficult.



Objective: Create a Model and Profitable Betting Strategy

The goal of this project is to create a model that can:

1. Accurately predict the outcome of games, then
2. Using the predicted outcome, develop a profitable betting strategy

Clients

- Sports bettors looking to make good bets
- Sports analytics companies looking to gain an edge over the competition
- The casino looking to keep their advantage

Data

- Basketball Data:
  - Contains box score results and basic statistics since 2003
  - From 2018 March Madness competition hosted by Kaggle
- Betting Data:
  - Contains score results for each game and historic betting lines
  - From The Prediction Tracker

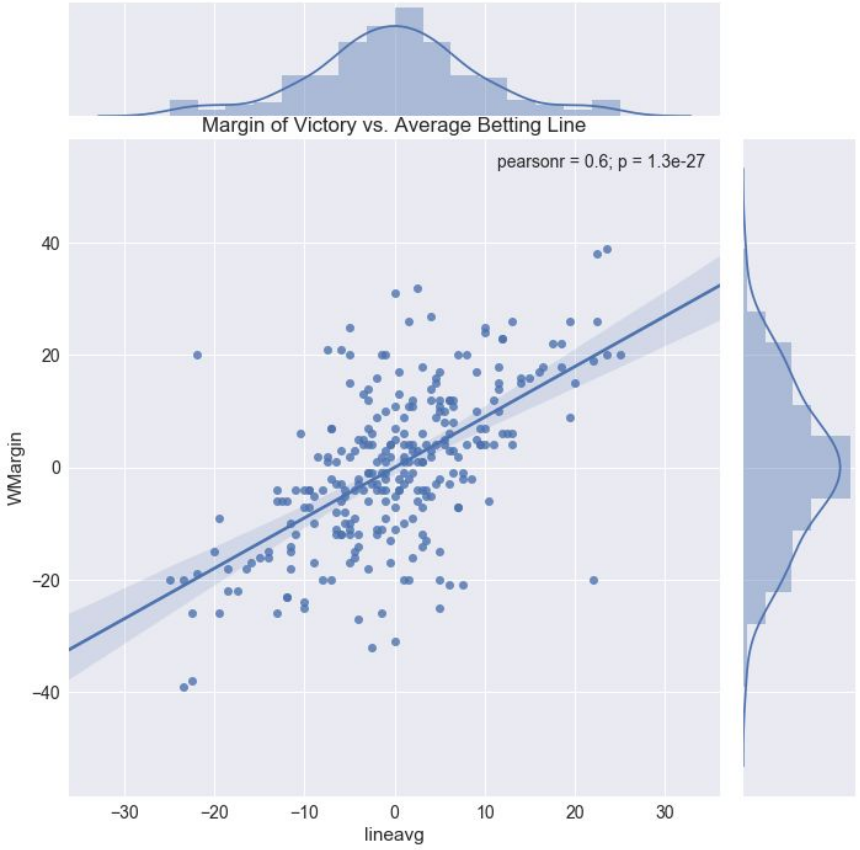
Data Cleaning

- Basketball Data:
  - Calculate advanced stats
- Betting Data:
  - Assigning Team ID’s to teams with inconsistent spellings
  - Dropping unreliable betting lines from various platforms

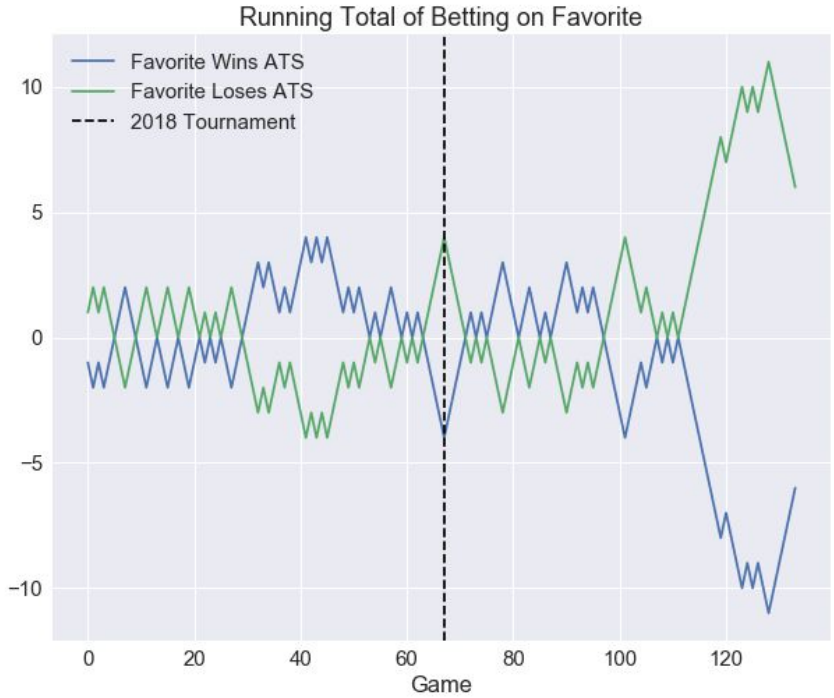
Approach: Predict Win Margins With Regression

- Betting lines are defined by predicted win margins
- Apply regression models to predict win margins
- Compare predictions with actual results and betting lines
- Bets are profitable if model predictions are closer to actual results than betting line

Betting Lines Are Closely Related to Actual Results



A Naive Betting Strategy is Profitable… If you choose the right one



Machine Learning Methods

- Linear Regression
  - 1st model (baseline) uses just Elo Ratings.
  - 2nd model extended to include advanced stats
- Linear SVR
  - 1st model uses just Elo Ratings
  - 2nd model extended to include advanced stats
- Decision Tree
  - Default parameters
  - Prediction mechanism similar to how analysts make prediction

Features

Elo Ratings:

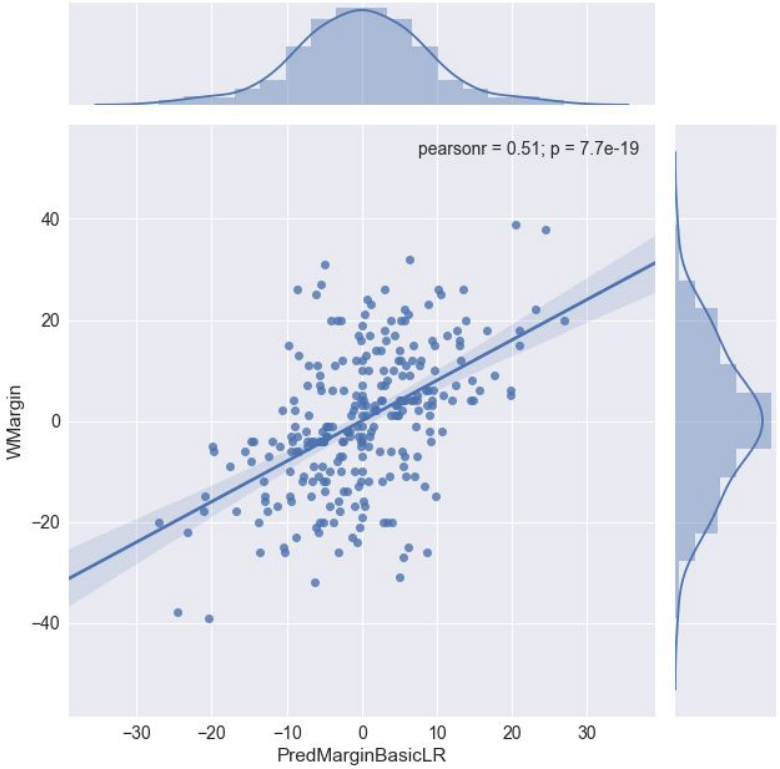
- Singular measure of strength of team
- Accounts for decisiveness of games and strength of teams faced

Advanced Stats:

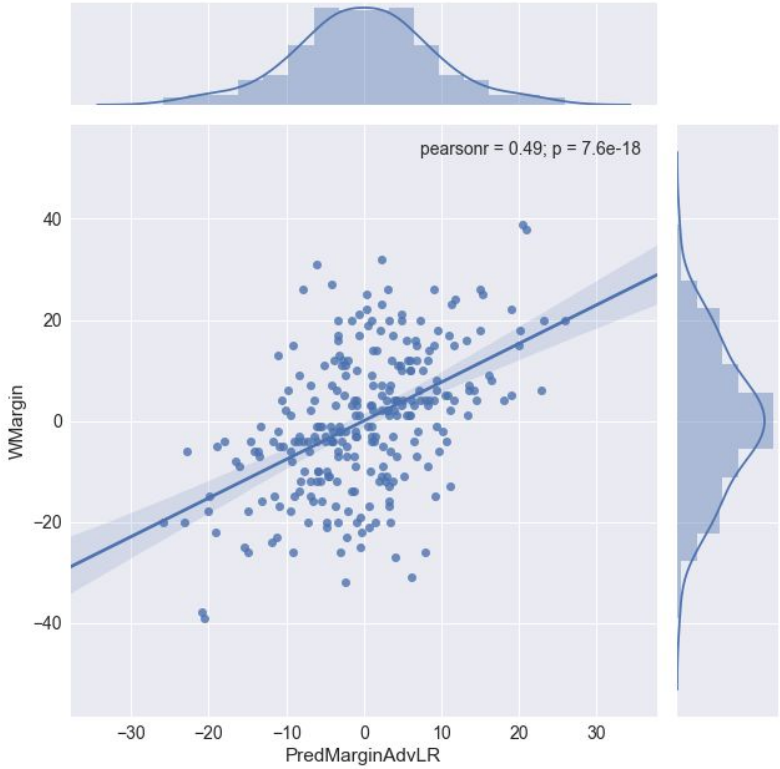
- In general, measures of how well a team is at one aspect of the game
- Example: Shooting - eFG%, Passing - Assist Rate, Rebounding - Rebound%, etc.

Results: Linear Regression

Elo Ratings

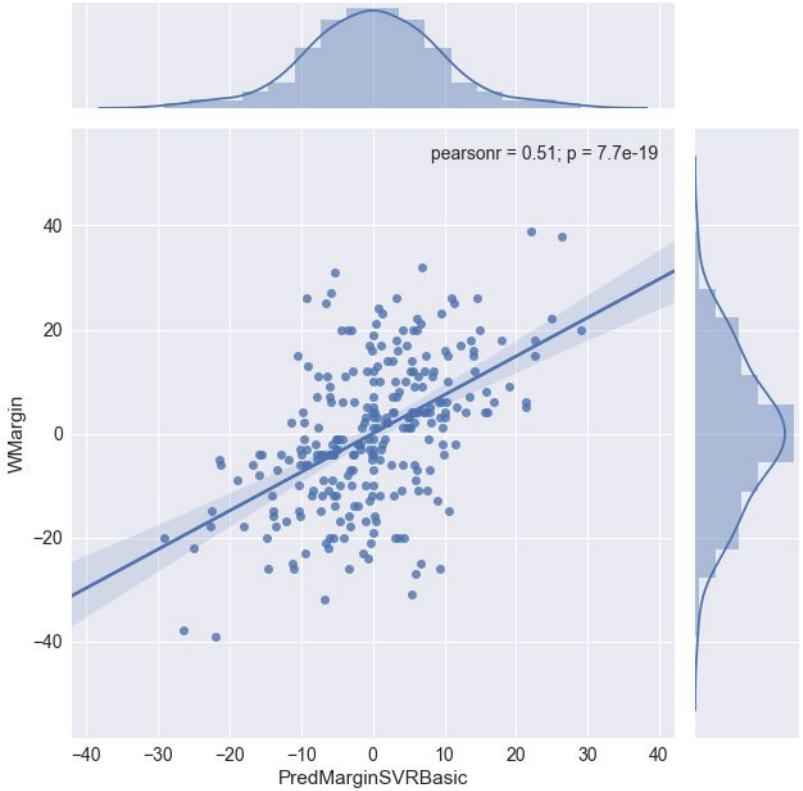


Elo Ratings + Advanced Stats

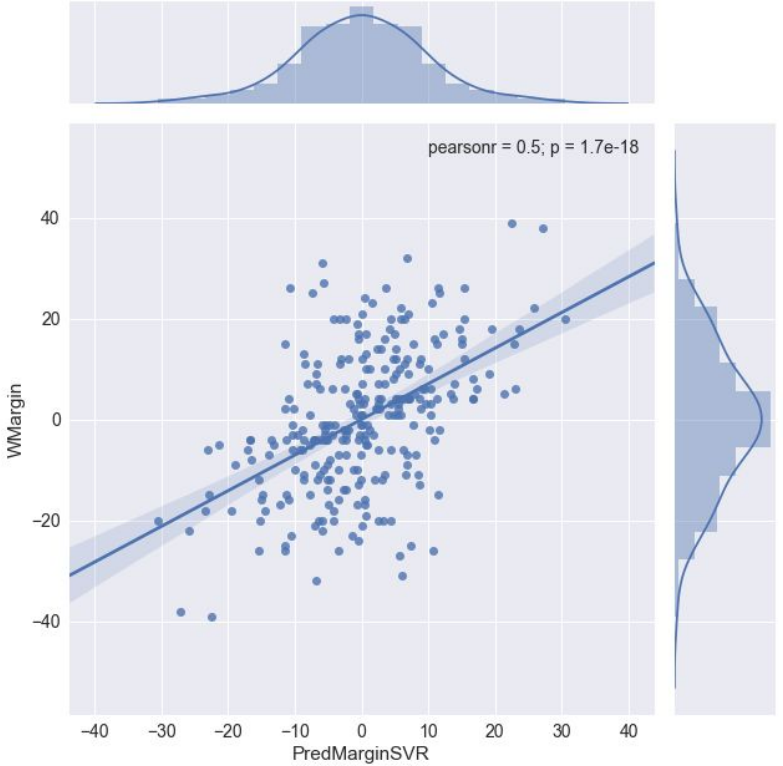


Results: Linear SVR

Elo Ratings

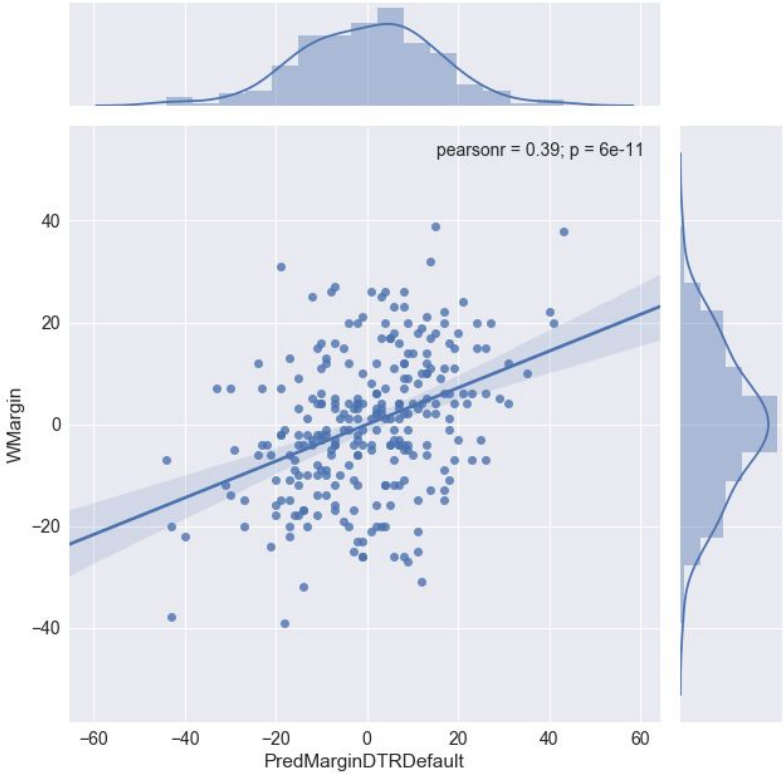


Elo Ratings + Advanced Stats

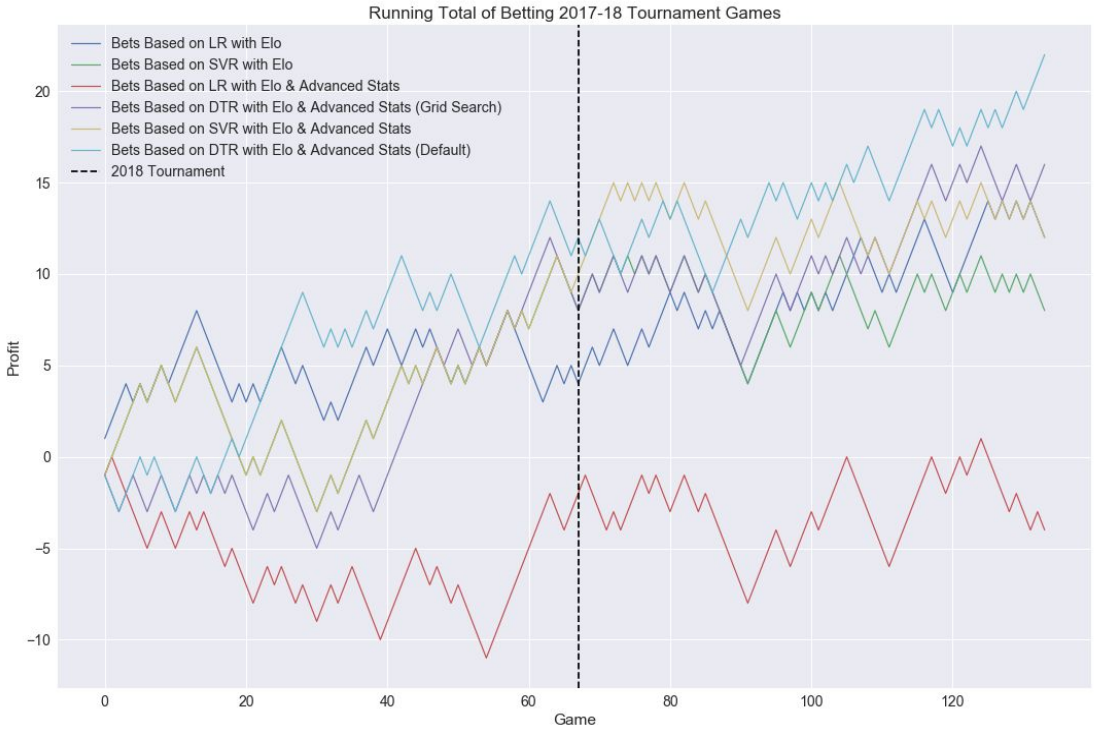


Results: Decision Tree Regressor

Elo Ratings + Advanced Stats



Results: Betting Simulation



Takeaways:

- DTR is most profitable
- Advanced stats improved SVR, did not improve LR
- Difficult to improve upon baseline model, LR with Elo Ratings

Conclusions / Recommendations

Regression Models are Generally Profitable

- Every model except LR with advanced stats were profitable

Decision Tree Regressors Are The Most Profitable

- DTR predict outcomes similarly to how games are actually predicted

Identify Key Bets

- For clients with less capital, focus on what makes a profitable bet.

Future Work

- Confidence Based Betting
  - Current model assumes all bets are made equal
  - Adjust bet sizes based on confidence by model
- Betting on Different Platforms
  - Some platforms are more accurate than others
- Add Non-Statistical Features
  - Game location, injuries, coach’s ability