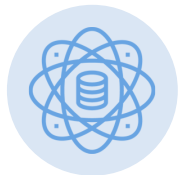

Beating the Book: Using ML to Identify an NBA Betting Edge

Thesis Overview



PART 1

- ✓ Convert betting lines into win probabilities



PART 2

- ✓ Build model to predict win likelihood for each team in a matchup



PART 3

- ✓ Test performance of my model versus betting odds

Data Sets

GAMBLING DATA

- Betting lines [archive](#) for each NBA game from 2007-08 to 2019-20

NBA TEAM GAME LOGS (REGULAR SEASON & PLAYOFFS)

- [nbastatR](#) package in R with vast API of NBA statistics
- Available team [data](#):
 - Basic statistics (points, rebounds, assists, etc.)
 - Matchup data (home/away, days of rest, game number of the season)
 - Advanced metrics (efficiency ratings, pace of play)



Part 1 – Betting Odds

COMPLETE

MATHEMATICAL DERIVATION FROM BETTING ODDS TO WIN PROBABILITY (using example of risking \$100 wager)

FAVORITE

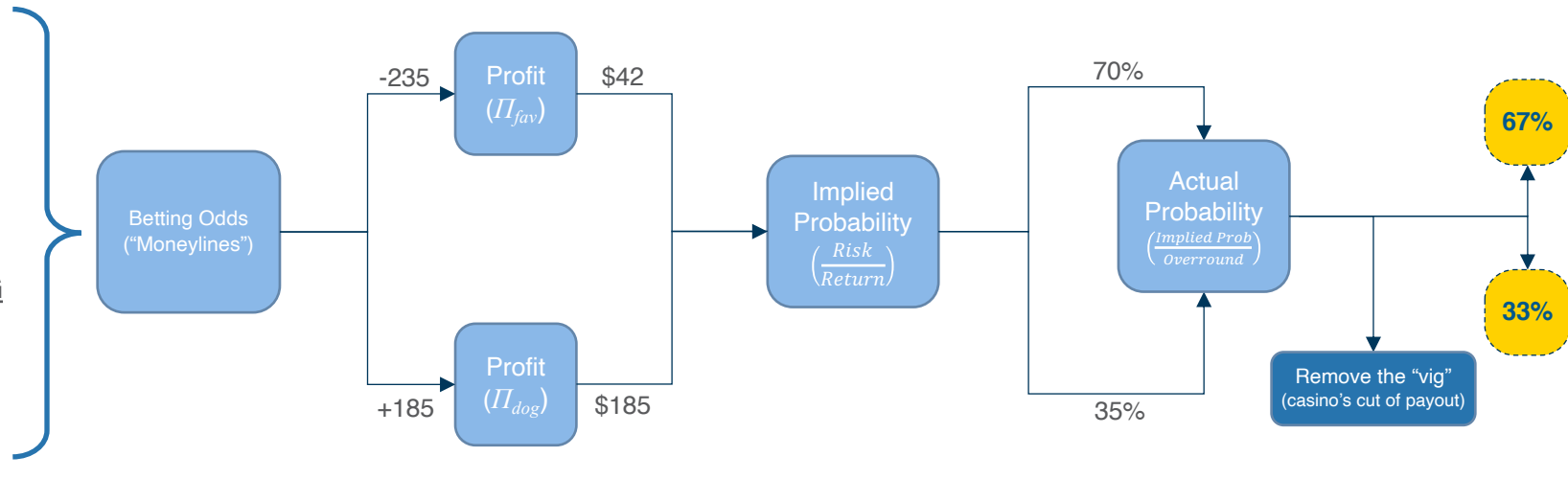


Warriors (-235)

UNDERDOG



Clippers (+185)



Notes

$$\Pi_{fav} = \frac{100}{-1 * ML} * Risk$$

$$\Pi_{dog} = \frac{ML}{100} * Risk$$

$$Return = Risk + Profit$$

$$Overround = IP_{fav} + IP_{dog}$$

Part 2 – Model Building

IN PROGRESS

GOAL: For each team in a matchup, predict the binary response variable (Win or Loss) along with a confidence level (0-100%) to use in comparison to the betting odds to determine if there's an edge.

AGGREGATION METHODS

- Use aggregation of full season to date to predict current game
- Use a rolling aggregation of last X games to predict current game

POTENTIAL PITFALLS

- Handle postseason and regular season differently?
- Lack of data for early season games. *Solution: use games from end of previous season?*

POTENTIAL MODEL TYPES

- Random Forest (with Boosting?)
- Neural Network
- Ridge or Lasso Regression
- PCA Regression

Part 3 – Model Application

COMPLETE

FIXED WAGER

1. Only wager on teams in which model predicts higher win probability than the betting odds.
2. Place fixed bet amount (e.g. \$10) on every wager.
3. Accumulate profits and losses from all bets to see the extent that the final amount is above or below starting bankroll of \$0.

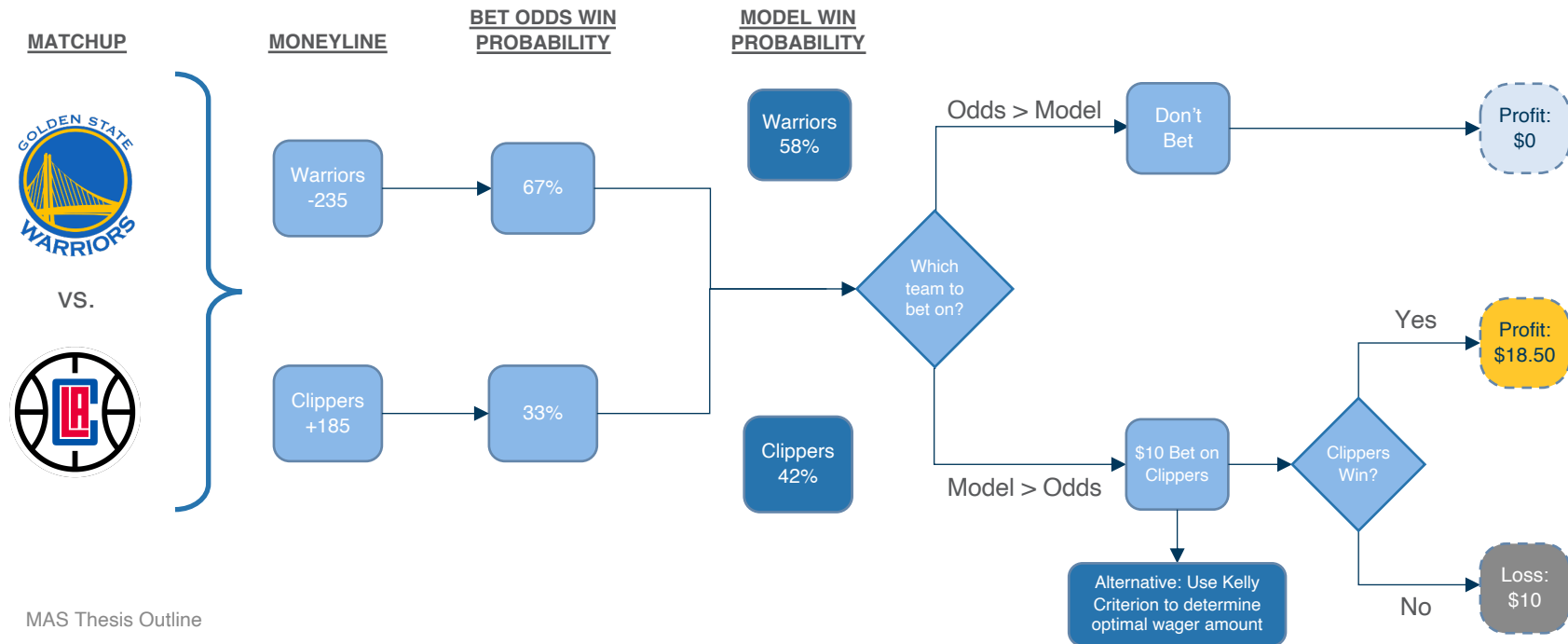


KELLY CRITERION

1. Only wager on teams in which model predicts higher win probability than the betting odds
2. Implement [Kelly Criteria](#) to determine optimal percentage of bankroll to wager based on edge identified from model confidence versus betting odds.
 - Compare results of fractional Kelly: $\frac{1}{4}$ Kelly, $\frac{1}{3}$ Kelly, $\frac{1}{2}$ Kelly, and Full Kelly
3. Accumulate profits and losses from all bets to see if bankroll grows from initial \$1000 or trends toward \$0.

Logic Flow for Fixed Bet Payout

Example application of fixed \$10 wager placed for a sample matchup and random model result

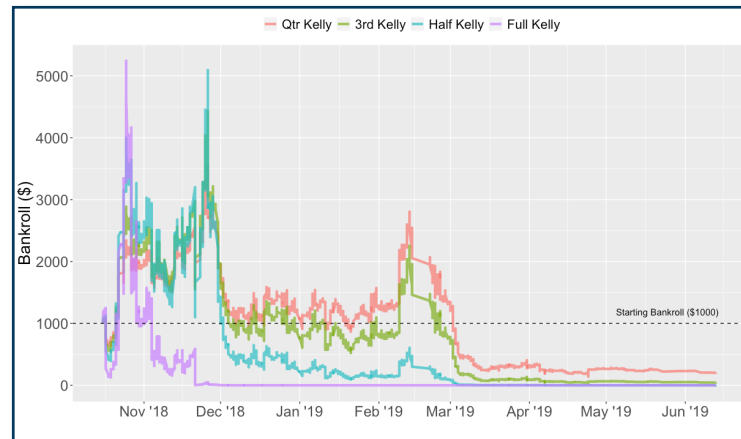


Comparison of Betting Strategies

Proof of concept using predictions from a randomized model on 2018-19 NBA season



BANKROLL FROM FIXED \$10 WAGER



BANKROLL FROM IMPLEMENTING KELLY CRITERIA

(beginning with \$1,000 bankroll)

RANDOM MODEL RESULTS:

Win	Loss	No Bet
667 (25.4%)	674 (25.7%)	1283 (48.9%)

Thank You
