

# A Bayesian-Boosted Approach to Predicting NCAA Tournament Outcomes

Jaden Rothschild

# Introduction / Research Question

- March Madness is the most popular tournament in the country
- 26.6 Million brackets filled out on ESPN
- Machine Learning can help improve our bracket predictions

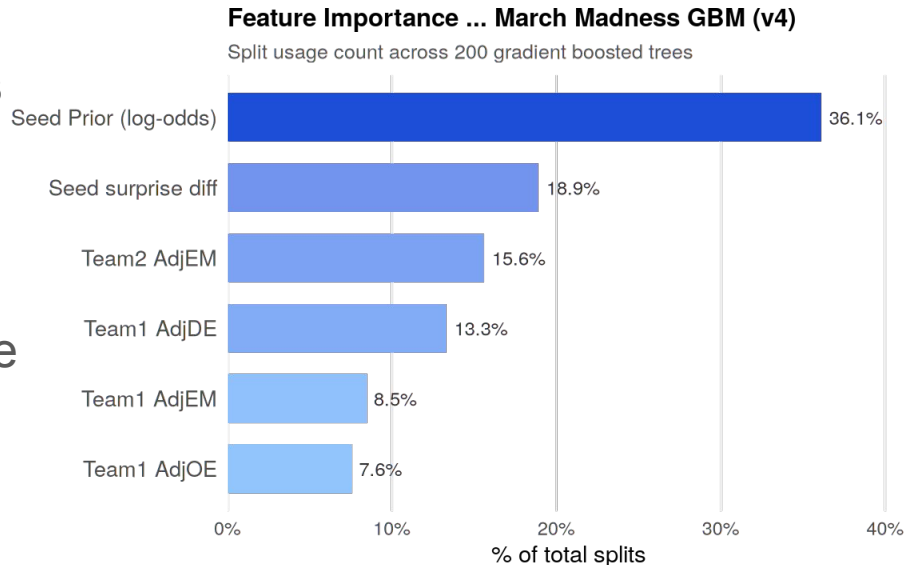
## Two Main Research Questions

1. What is the probability of each team winning their March Madness game?
2. What is the margin we expect each team to win or lose by?



# Data Source

- Data collected from KenPom (2003-2026 March Madness games)
- Focuses on the Seed Prior
- Adj OE, Adj DE and Adj EM metrics
- Seed Surprise difference
- Variables are limited to reduce noise
- Models including box score metrics were overfit
- Adjusted efficiency is used to account for SOS

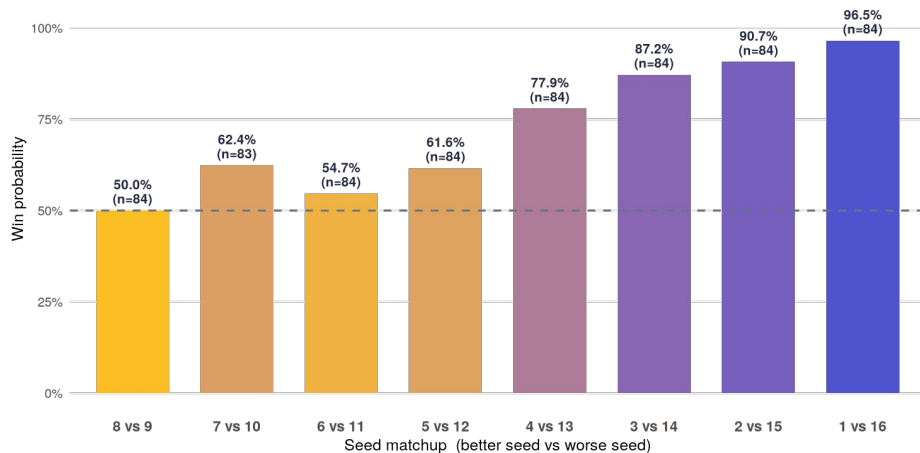


# Bayesian Prior

- Seed matchup win rates are stable enough across 40 years to serve as a reliable prior
- The 1v16 has held at 96%+ since 1985
- The prior alone gets you ~70% accuracy; everything else is refinement
- Win probabilities for each seed matchup used beyond first round
- This prior serves as a baseline before we see team stats

## P(Higher Seed Wins) ... Standard R64 Matchups

Bayesian prior from 2003...2024 NCAA tournament data (Laplace-smoothed)



# Gradient Boosted Tree

- Xgboost regression predicts final point margin (100 trees, learning rate) while using seed matchup as a prior
- Predicted margin is converted to win probability using a Normal distribution with  $\sigma = 11$  pts:  $P(\text{win}) = \Phi(\frac{\text{Predicted Margin}}{11})$
- No separate classifier needed; win probability flows directly from the margin model, keeping the model simple and interpretable
- Spreads from sportsbooks used for comparison

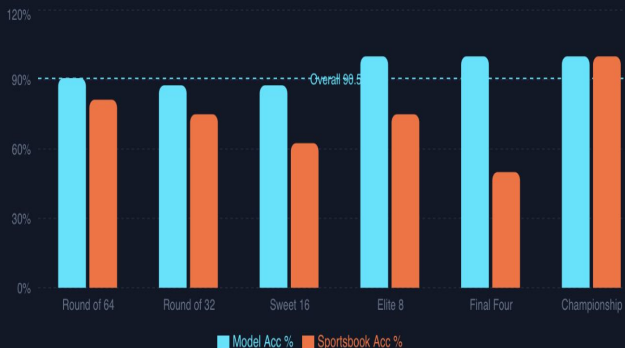
Predicted vs Actual Margin

Each dot = one game · cyan = correct, red = wrong



Accuracy by Round

Validation 2026 - Reg Margin → Prob



MAE by Round: Model vs Sportsbook

Mean absolute margin error (pts)



# 2026 March Madness Results

- **90.5%** accuracy (missed 5.5 games, 14% better than sportsbooks)
- Correctly picked Michigan to win championship (1.5 ME in championship)
- MAE only 0.12 better than books (performed better in close games)
- Test and Validation performed better due to less upsets in the tournament
- Log Loss was 0.35, Brier Score was 0.11



**Questions?**