Simple Enough to Explain, Strong Enough to Compete: The Accuracy-Interpretability Frontier in NCAA Selection Models

Nic Abbagnaro, Nate Kirschman, Andrew Miller, Andrew Sternberg, Dr. Matthew Lanham (Advisor)
Butler University Lacy School of Business
nabbagnaro@butler.edu; nkirschman@butler.edu; aamiller1@butler.edu; asternberg@butler.edu

BUTLER
LACY SCHOOL of BUSINESS

ABSTRACT

This study evaluates how much predictive performance must be sacrificed to keep NCAA tournament models transparent. The question is motivated by a productive tension in the literature: simple models such as Coleman et al.'s probit approach and Dutta and Jacobson's decision tree are easy to explain and closely tied to committee-style reasoning, while more flexible machine-learning models may capture nonlinearities that improve prediction but reduce interpretability. We compare depth-limited decision trees and other explainable specifications against boosted-tree benchmarks using a compact, committee-facing set of resume variables. Leave-one-season-out validation will assess AUC, missed bids, and calibration alongside structural measures such as depth, rule count, and feature simplicity. The expected contribution is an empirical map of the transparency-accuracy tradeoff in this setting. Rather than assuming that explainability and performance are always in tension, the project asks where interpretable models remain competitively strong and where added complexity begins to earn its keep.

BUSINESS PROBLEM FRAMING

Research Question: Can interpretable models (linear regression and depth-limited trees) perform as well as less interpretable models (Random Forest and Neural Networks) to support seed decisions?

- **The Trade-Off:** High accurate results are complex and hard to understand, and interpretable models are simpler but sacrifice accuracy.
- **Impact:** Organizations rely on models for high-stakes decisions such as financial decisions and even in the hiring process. Stakeholders want accuracy, but they also want explanations for these models so they can trust them.
- **Constraints & Challenges:** Human decisions also include the "eye test", which is subjective. Data limitations also occur as metrics don't capture everything. The tradeoff reality also infers that perfect accuracy plus interpretability is rare.
- **Key Takeaway:** Effective decision making requires balancing predictive accuracy with model transparency to ensure both performance and trust.

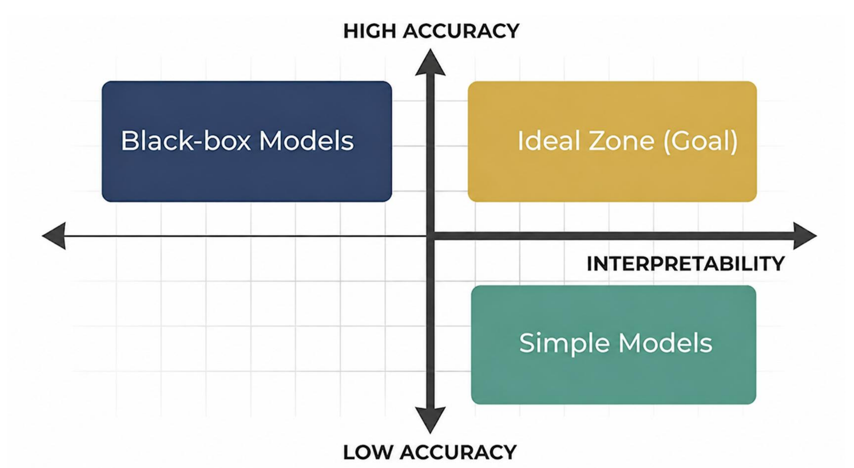


Fig 1. Accuracy Model

ANALYTICS PROBLEM FRAMING

- **Analytical Problem:** This problem is a regression-type problem as we are predicting the team's seed in the NCAA tournament.
- **Assumptions:** assume data is representative in nature of past committee decision seed placements and team results.
- **Success Metrics:** captured R2 and MSE (Kaggle evaluation) statistical evaluation metrics.
- **Justification of Approach:** The approach of the analytical problem is justified due to the predictive nature of the NCAA tournament, where games cannot have clear answers for every outcome.

DATA

Important Data Relationship

- We found that NET Rank was the highest predictive indicator of Overall Seed with our model

Analytics Problem Framing

- Our team translated the business problem of predicting March Madness bids into an analytical problem by using the previous four years of picks to train our model.

Column	Description
RecordID	Unique identifier for each record
Season	The academic year associated with the relevant basketball season. E.g. The 2024-2025
Team	The NCAA Member Institution sponsoring Men's Basketball for Division 1.
Conference	The Conference the Team associated with for Division 1 Men's Basketball for the
Overall Seed	The overall rank assigned for the 68 teams selected to compete in the NCAA
Bid Type	AQ = Automatic Qualifier. Assigned to team that won their Conference's tournament;
NET Rank	The NET Ranking (NCAA Evaluation Tool) is the primary metric used by the NCAA to
PrevNET	The NET Ranking (NCAA Evaluation Tool) is the primary metric used by the NCAA to
AvgOppNETRank	Average Opponent NET Rank. Ranks team's AvgOppNet compared to the rest of the
AvgOppNET	Average Opponent NET. Measures the average NET ranking of a team's opponents over
WL	Win-Loss; Number of Wins and Losses separated by a dash.
Conf.Record	Win-Loss against teams in the same conference

Table 1. Data Dictionary

METHODOLOGY

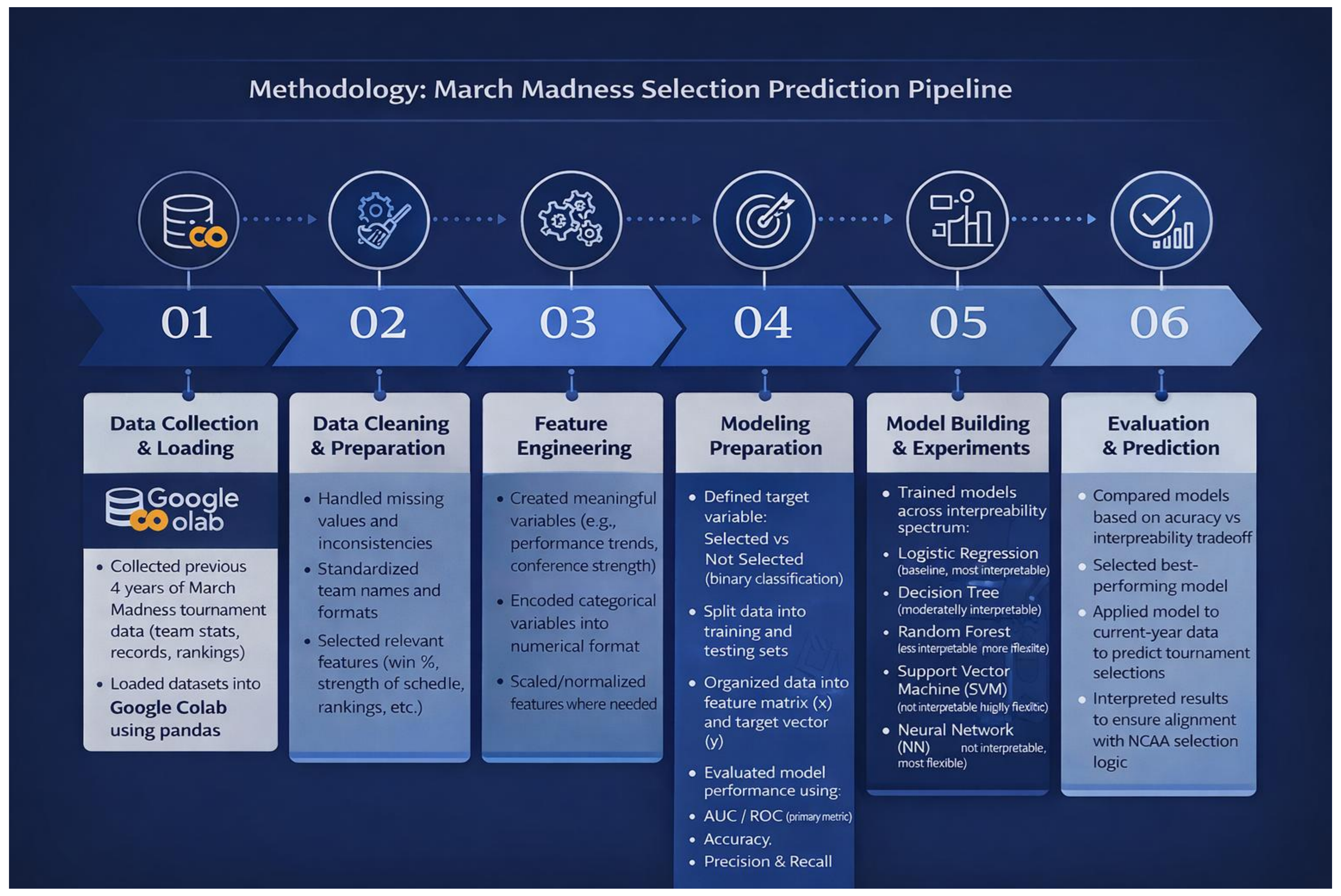


Fig 2. Methodological Workflow

MODEL BUILDING & EXPERIMENTAL RESULTS

Table 2 depicts the experimental accuracy of our predictive models. We found that the Random Forest model performed the best among all experiments with a 98.05% R-squared. The Decision Tree model was close with 96.66, and our more interpretable linear regression model obtained an 86.27 R2.

Training vs. Validation R-squared Table

Model	Set	R-squared
Linear Regression	Train	0.8645
Linear Regression	Validation	0.8627
Decision Tree	Train	1.0000
Decision Tree	Validation	0.9666
Random Forest	Train	0.9964
Random Forest	Validation	0.9805
Neural Network (MLP)	Train	0.9311
Neural Network (MLP)	Validation	0.9138

Table 2. Experimental Results

Figure 3 shows that all model experiments were candidate models (not overfit)

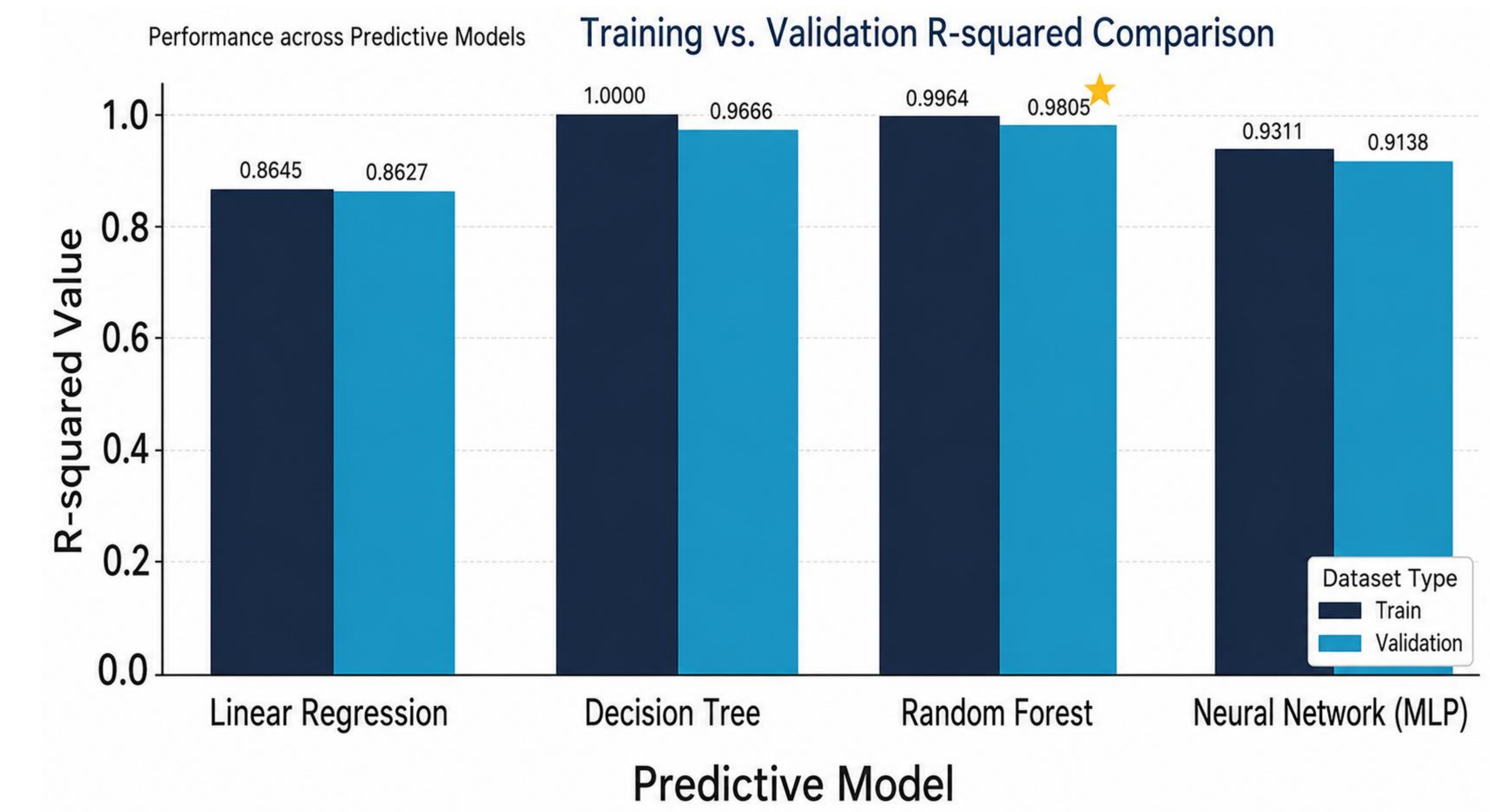
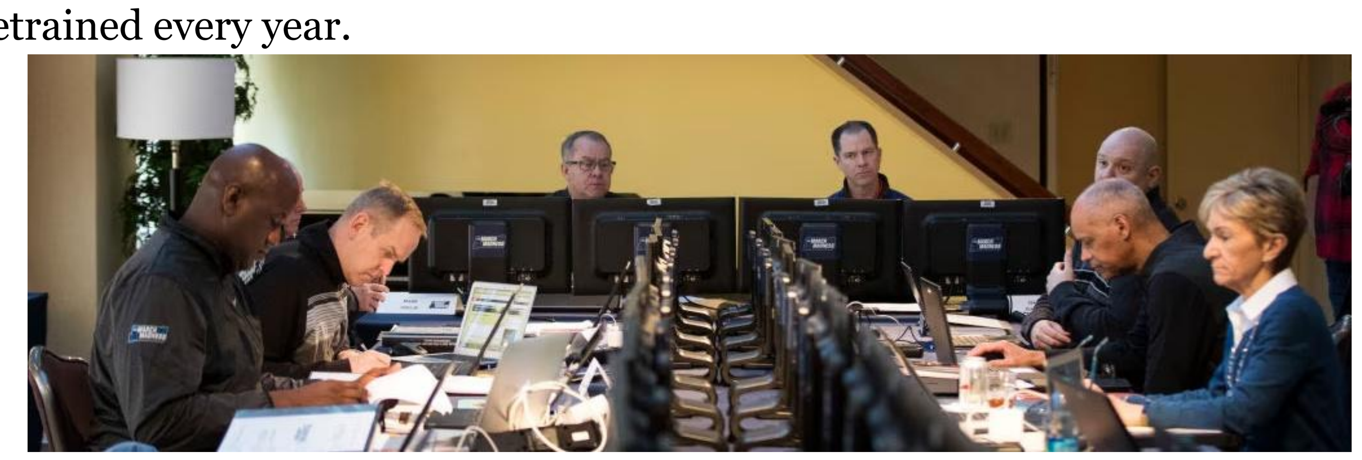


Fig 3. Experimental Results

DEPLOYMENT & LIFECYCLE MANAGEMENT

- Our Logistic Regression model achieved a Kaggle MAE score of 34.35 (public) / 34.52 (private), ranking 62nd out of 83 teams. While we were not one of the top performers, the model offers a transparent, interpretable baseline that others could use to examine seeding decisions with easy to explain logic.
- Using NET rankings, win/loss records, strength of schedule, and conference stats together produced better predictions than any single variable did by itself. This confirms nuance improves overall seed predictions.
- The model would be used each year in the weeks before Selection Sunday by the selection committee to help facilitate decisions. Since team rosters, coaching staff, and conference strength change every season, the model should be retrained every year.



KEY TAKE-AWAYS

Research Question:

- A decision tree provides good interpretation and is competitive in predictive accuracy (96.6% R2). It could even outperform a Neural Network (91.3% R2), but a mildly interpretable model like Random Forest appeared to perform the best (99.6%).

Future Ideas:

- Perform a deep dive into the Random Forest (RF) model predictions to see where it did well and where the shortcomings might be.
- Do the linear regression drivers align with the RF drivers?

Competitive Outcome:

- Kaggle MAE Score of 34.35, which ranked 62nd out of 83 teams. We were not able to test our Random Forest predictions in time of the FFAC deadline

Personal Development & Outcomes

- Completed 2 courses on DataCamp about Python Skills
- Received badge on SAS learning for their advanced analytics course on SAS Viya Model Studio
- Developed a stronger understanding of Python and Gemini AI through Google Copilot

datacamp

ADVANCED ANALYTICS LEARN sas