(Money)ball is Life – Point Spread Prediction

Joseph Glasson, Daniel Murong

Overview

Our goal was to predict NCAA men's basketball point spreads (home score − away score) using team efficiency metrics. We used historical ACC game results and opponent-adjusted team statistics to train a regression model and generate predictions for upcoming ACC games.

Data and Preprocessing

We used three datasets: Historical ACC 2025-26 game scores; Barttovik's 2025-26 team efficiency metrics; the submission schedule. We defined the response variable as: spread = home points − away points

Model

We used opponent-adjusted efficiency metrics as our predictor variables. Rather than using raw values, we constructed *home–away differences* to model relative strength: **ADJOE** – Adjusted Offensive Efficiency → adjoe_diff; **ADJDE** – Adjusted Defensive Efficiency → adjde_diff; **ADJT** – Adjusted Tempo → tempo_diff.

We fit a linear regression model: spread ∼ adjoe_diff + adjde_diff + tempo_diff

We initially tested additional predictors (e.g., BARTHAG) but removed them when they added little predictive value and risked overfitting. We chose linear regression because of the small size of the dataset (~85 games), for ease of interpretability, and simplification of the model.

Evaluation

We used an 80/20 train-test split. Performance metric mirrored the competition's evaluation methods i.e. **Mean Absolute Error (MAE).** We achieved a baseline test MAE 8 points, while large misses were concentrated in a few blowout games

We checked for systemic bias and observed that predictions slightly overestimated home performance. We tested small adjustments but avoided tuning heavily to the small test set to prevent overfitting.

Conclusion

Opponent-adjusted efficiency differences provide meaningful signal for predicting spreads. Even a simple linear model captures a substantial portion of outcome variability. Given the limited data, we prioritized robustness and interpretability over complex machine learning methods. Future improvements could include multi-season data, recent performance trends, and injury information.