Optimizing NCAA Tournament Chances: A Deep Learning Approach to Strategic Scheduling for Mid-Major Conferences

William Zhu

Abstract

Predicting NCAA basketball rankings is essential for mid-major conferences aiming to maximize their chances of securing at-large bids to the NCAA Tournament. Given the challenges these conferences face in attracting and retaining elite talent, strategic non-conference scheduling becomes critical for improving their NCAA Evaluation Tool (NET) rankings, which directly tie into whether or not a team will get an at-large bid. Motivated by a scheduling optimization challenge presented by the Atlantic 10 (A10) Conference, this study develops a deep learning model to classify NCAA Division I basketball teams into tournament-selection quadrants based on historical performance data. The neural network models achieved an overall F1 score of 0.6225 for men's teams and 0.6609 for women's teams. While the models demonstrated strong predictive capabilities in classifying teams in Quadrants 3 and 4, performance was comparatively lower for accurately predicting Quadrants 1 and 2. These results underscore the difficulty in differentiating elite and near-elite teams but affirm the practicality of using deep learning for broader scheduling and strategic planning purposes. Future research will aim to enhance predictive accuracy for higher-ranked teams by integrating additional metrics and exploring advanced model architectures.

1. Introduction

Securing at-large bids to the NCAA Basketball Tournament is a significant goal for collegiate athletic conferences, particularly those classified as mid-major. Unlike major conferences, mid-major conferences often face substantial hurdles in attracting and retaining elite-level talent due to two main reasons: lack of Name, Image, Likeness (NIL) deals, which allow players to profit from their personal brand by engaging in activities like endorsements, social media promotions, and selling merchandise and the introduction of the transfer portal which allows players transfer to better schools that can get them better NIL deals. Consequently, these conferences must strategically approach scheduling, particularly non-conference games, to enhance their NCAA Evaluation Tool (NET) rankings—a key metric used by selection committees to determine tournament eligibility.

Predictive modeling has emerged as a powerful tool in sports analytics, enabling teams and conferences to make data-driven decisions to optimize performance and strategic planning. By accurately forecasting team rankings and tournament selection potential, conferences can more effectively structure their schedules to improve their standings and maximize the probability of receiving at-large bids.

Motivated by this critical scheduling challenge, the Atlantic 10 (A10) Conference initiated a project aimed at leveraging advanced predictive analytics to optimize scheduling strategies across NCAA Division I basketball teams. Although inspired by the A10's objectives, the scope of this study encompasses all NCAA Division I teams, reflecting a broader applicability. This paper details the development and application of deep learning models designed to classify teams into NCAA tournament-selection quadrants based on historical performance data. The primary aim is to provide actionable insights that conferences like the A10 can use to make informed scheduling decisions, thereby enhancing their teams' competitive profiles and overall chances of securing at-large tournament bids.

2. Methods

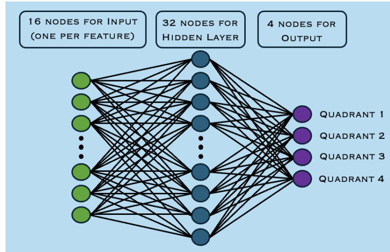
2.1. Data Collection and Preprocessing

Historical game data for NCAA Division I basketball was collected from Massey Ratings, comprising over 17,000 games spanning the 2021-22 through the 2023-24 seasons. Additionally, NCAA NET rankings were collected from the official NCAA website immediately prior to Selection Sunday for 2023, 2024, and 2025. Each dataset entry represents one NCAA Division I basketball team's seasonal performance. In total, 1,080 team-season examples were gathered, each described by 16 distinct features: total wins and losses, along with home, away, and neutral wins and losses for both the team and its opponents. Based on NCAA NET rankings, teams were classified into tournament quadrants:

- Quadrant 1: Ranked 1-50 (Men), Ranked 1-35 (Women)
- Quadrant 2: Ranked 51-100 (Men), Ranked 36-65 (Women)
- Quadrant 3: Ranked 101-200 (Men), Ranked 66-105 (Women)
- Quadrant 4: Ranked 201+ (Men), Ranked 106+ (Women)

Data was standardized using a StandardScaler prior to training to ensure equal feature weighting.

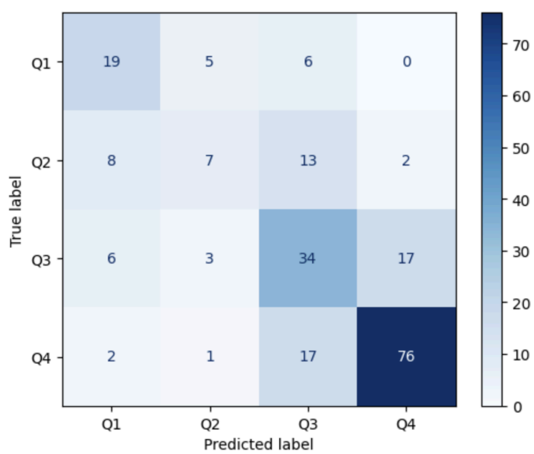
A multi-class classification neural network was constructed using PyTorch. The architecture included an input Layer with 16 nodes corresponding to each feature, one hidden Layer with 32 nodes with ReLU activation, and an output Layer with 4 nodes, each representing a quadrant, with a SoftMax activation

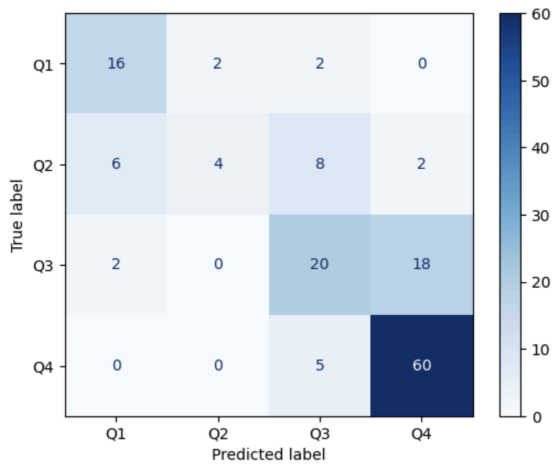


The dataset was partitioned into training (60%), validation (20%), and testing (20%) sets. The neural network was trained for 50 epochs with a learning rate of 0.001. The best-performing model, determined by the lowest validation loss, was selected for generating predictions. For evaluation, we looked at accuracy within each quadrant as well as the overall F1 score.

3. Results

The neural network achieved an overall F1 score of 0.6225 for men's teams and 0.6609 for women's teams. The confusion matrices are shown below, with men’s on the left and women’s on the right.





4. Discussion

Looking at the confusion matrix for men's and women's, we can see that the model predicted fairly well for Q3 and Q4 teams. On the men’s side, when a team is predicted as Q4, there is a 93% chance it is either Q3 or Q4, and when a team is predicted as Q3, there is a 73% chance that the team is either Q3 or Q4. For women's teams, when a team is predicted as Q4, there is a 98% chance it is either Q3 or Q4, and when a team is predicted as Q3, there is a 71% chance that the team is either Q3 or Q4. This is most likely due to the number of teams that are distributed between the quadrants. There were over double the number of teams in Q3 and Q4 compared to Q1 and Q2 in both men's and women's. Wins against Q3 and Q4 teams do not add much, if anything, to NET rankings, and losing to such a team drops a team's NET rankings by a lot. Scheduling games against such teams is not ideal in that case. When making decisions about scheduling, this model would work best at filtering out those Q3 and Q4 teams. That is not to say that this will completely replace statistics-driven decision making, but the model would assist in such decision making.

5. Future Work

There is always more data to be collected. As the years go on, more and more data will be readily available to scrape about NET rankings and team records. Many different team statistics were not explored in this research for simplicity, which might be another avenue to follow in future work. Finally, the model was also an extremely simple neural network built from scratch. With more data in the future, looking at more complicated models or even using pretrained models might result in better predictability and give us more insights about the quadrant rankings.

6. Conclusion

This study shows that deep learning methods can effectively predict NCAA basketball rankings, providing valuable assistance to mid-major conferences in improving their non-conference scheduling strategies. The developed neural network models accurately classified teams into the lower quadrants (Quadrants 3 and 4), making them practical tools for identifying matchups that could negatively impact NET rankings. However, distinguishing between top-tier teams (Quadrants 1 and 2) remains difficult, highlighting the complexity involved in accurately predicting outcomes for closely matched elite teams. Future research could benefit from incorporating more detailed team and player data, trying more advanced model structures, and expanding the dataset over multiple seasons. Ultimately, this research supports the use of deep learning predictions in strategic scheduling to enhance the competitiveness of mid-major NCAA basketball conferences.

7. References

1. Boettger, E. (2019). Mid-Major Scheduling: Where Can Teams Go for Quality Games? AthleticDirectorU. Retrieved April 30, 2025, from https://athleticdirectoru.com/articles/mid-major-scheduling-where-can-teams-go-for-quality-games/ Athletic Director U

2. Kuchera, M. (2025). Deep Learning [Davidson College course, PHYS / CS].

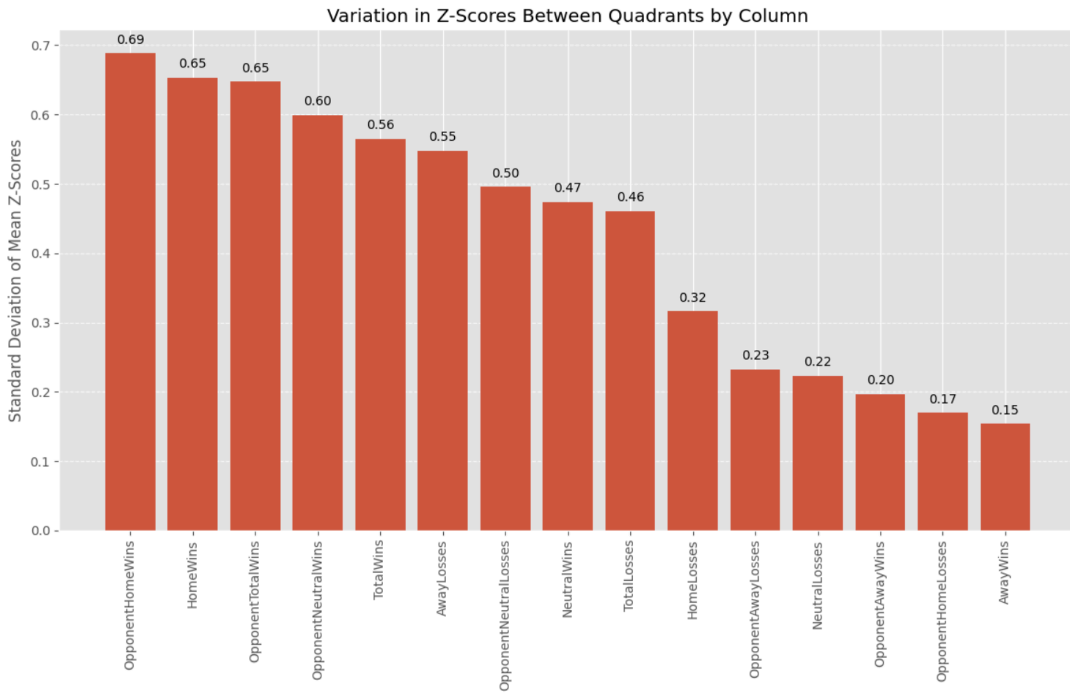
3. Massey, K. (n.d.). Massey Ratings. Retrieved April 30, 2025, from https://masseyratings.com

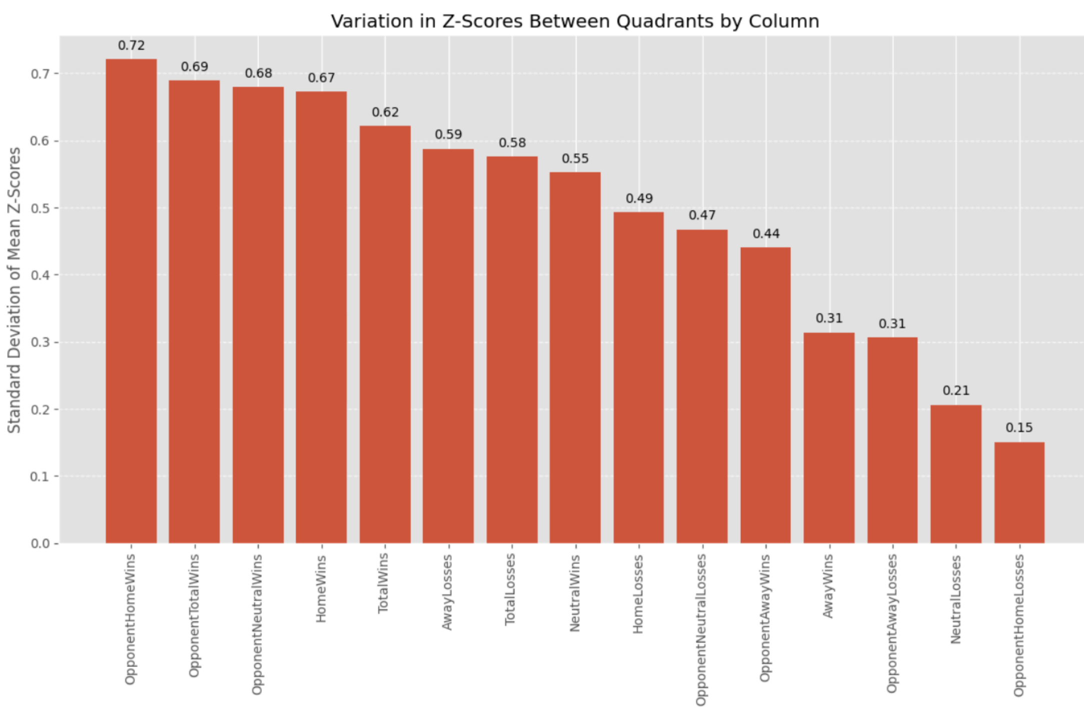
4. NCAA. (2025, March 16). DI Men's Basketball Rankings – NCAA Men's Basketball NET Rankings. Retrieved April 30, 2025, from https://www.ncaa.com/rankings/basketball-men/d1/ncaa-mens-basketball-net-rankings

5. NCAA. (2025, April 6). DI Women's Basketball Rankings – NCAA Women's Basketball NET Rankings. Retrieved April 30, 2025, from https://www.ncaa.com/rankings/basketball-women/d1/ncaa-womens-basketball-net-rankings

8. Appendix

Feature ranking of variation in Z-score based on quadrant (Top: Men's, Bottom: Women's)





Data exploration of Z-Score distributions of the features based on quadrants (Top: Men's, Bottom: Women's):