A SYSTEMS-BASED APPROACH FOR PREDICTING NCAA BASKETBALL TOURNAMENT OUTCOMES

Oscar Manuel Contreras Gacha
David Felipe Ariza Ariza
Julian Mendez

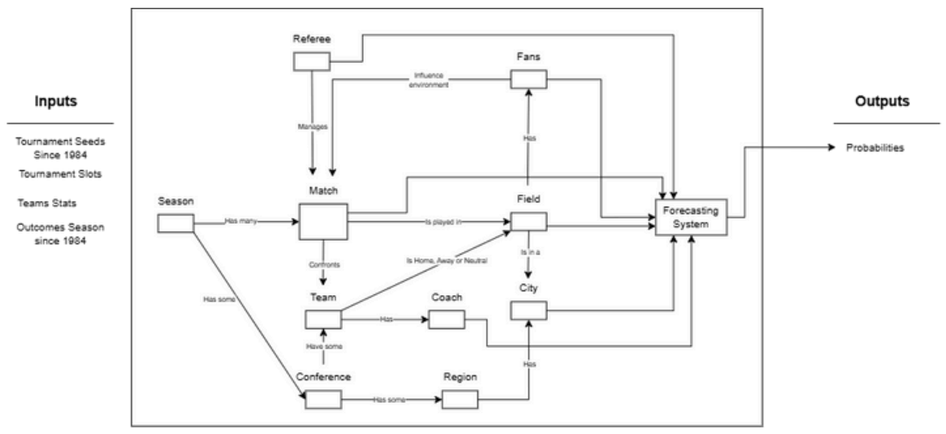
Universidad Distrital Francisco José de Caldas

The NCAA Men's Basketball Tournament ("March Madness") is notoriously hard to predict due to its unpredictable, single-elimination nature. Current prediction models often miss complex interactions. This project aims to use systems thinking to understand these dynamics and design a better predictive system for the 2025 tournament

GOALS

- To design and lay the groundwork for implementing a robust, modular, and adaptable machine learning system for predicting NCAA tournament match results
- To iteratively implement a simulation flexible, adaptable, and well-architected, of the system focusing on delivering clean, documented, and testable code.

SYSTEM DIAGRAM



METHODOLOGY

Our approach involves a modular pipeline designed to process historical data, train predictive models, and generate forecasts. The architecture emphasizes clear separation of concerns, from data acquisition to prediction output, facilitating adaptability and maintainability

INGESTION Loads and organizes all essential raw tournament data

FEATURE Transforms raw data by adjusting statistics, structuring the data, and crucially, incorporating and comparing team seeds

TRAINING The model is trained using a cross-validation strategy.

PREDICTION Model outputs are calibrated to convert predicted point spreads into more reliable win probabilities.

RESULTS

- Using all statistics as a baseline does not ensure optimal performance.
- Reducing the variable set slightly improved the results.
- Including seed rankings and their difference significantly improves performance.
- The best results come from combining all statistics with seed information, with a strong correlation observed between point difference and win probability.

Best result with seed information:

Score: 0.11790

Note: Visualizations comparing seed values with point differences showed clear trends, highlighting the practical importance of seeds.

CONCLUSION

The project demonstrates that combining systems engineering with machine learning enables a scalable, accurate forecasting system that captures the complexity of NCAA tournaments, emphasizing the importance of modularity, domain knowledge, and adaptability in handling real-world unpredictability.