

Summary: NBA Championship Win Prediction Using Machine Learning

By Faiz and Rafay

This presentation explores the use of machine learning models to predict NBA championship winners. We start with a literature review that highlights previous work using neural networks to analyze performance factors from both regular season and playoff statistics. Despite achieving high prediction accuracy, previous models struggled with data imbalance and overfitting, issues partly addressed through class weight adjustments. However, challenges remain in handling imbalanced datasets and accurately accounting for player trades and injuries. We used various methodology, which includes using logistic regression, ordinal logistic regression and random forest models alongside neural networks to gain insight into predicting the NBA Champions. WE identified key variables for logistic regression such as Points Per Game, Win Ratio, and Age, among others, emphasizing the significance of certain player and team statistics in predicting outcomes. We used R-squared from linear regression and the multicollinearity stats to pick our variables. The random forest section explains its ensemble learning approach, which combines multiple decision trees to improve classification accuracy and reduce overfitting. This model benefits from minimal data preprocessing and can handle non-linear data effectively, with an emphasis on feature importance to identify predictive variables. In the neural networks segment, the presentation describes the structure of these models, including layers, weights, biases, and activation functions. It addresses the challenge of overfitting by utilizing regularization techniques like dropout and L2 regularization to improve model generalizability. The results section presents the prediction accuracy for the 2024 NBA championship, reporting a mean absolute error (MAE) of 4.066667 and a root mean square error (RMSE) of 5.434458. Given the figures, we acknowledge limitations such as the potential bias from our data, the need for refined neural network models, and the lower accuracy of random forests in this context. We noted the significant predictive power of variables like Age, Win Ratio, and Points Per Game, while finding the Elo rating method not to be significant in the model. We conclude with a prediction for the 2024 NBA championship winner to be **Boston Celtics** and stress the importance for need to do further research to refine these predictive models, particularly to better account for dynamic factors like player trades and injuries.

