The Nexus of NBA Performance and NFT Value:
Machine Learning Project about NBA NFT Pricing

Feiting Li, Shuting Lin

Dec 9th, 2024

Introduction

In recent years, the intersection of sports and digital finance has witnessed a remarkable evolution, particularly through the emergence of Non-Fungible Tokens (NFTs). NFTs represent unique digital assets verified using blockchain technology, allowing for the ownership and trade of digital collectibles with proven scarcity and authenticity. The sports industry, especially professional basketball, has seen significant engagement in this digital transformation. Prominent examples include NBA star Stephen Curry, whose NFTs hold substantial financial value, and the Golden State Warriors, the first NBA team to issue their own NFTs in April 2021.

The launch of "NBA Top Shot," an officially licensed platform, has further exemplified this trend by offering digital highlights of NBA games as tradable NFTs. Given the burgeoning market and high transaction volumes on platforms like NBA Top Shot, a pertinent question arises: How does an NBA player's on-court performance impact the market value of their associated NFTs? This paper seeks to explore this question through a predictive modeling approach, employing machine learning algorithms to assess the influence of players' performances on their NFT pricing.

The burgeoning market of sports NFTs offers a unique opportunity to explore new economic models and asset valuation strategies that combine traditional sports metrics with modern financial technologies. This study replaces a purely econometric approach with machine learning algorithms—random forest, logistic regression, and XGBoost—to predict the impact of performance metrics such as points, rebounds, assists, steals, and blocks on the pricing of NBA Top Shot NFTs. By comparing the models' predictive capabilities, this paper identifies which method performs best and determines which variables most significantly affect NFT prices.

This study aims to test the hypothesis that NBA player performance metrics, along with external factors such as playoff appearances and social media influence, can effectively predict NFT valuations. By incorporating machine learning methodologies, the findings offer actionable insights for investors, fans, and industry stakeholders, potentially influencing future digital asset valuations and sports marketing strategies.

This paper is structured as follows: It begins with a review of existing literature on NFTs, including their operation, market dynamics, and prior analyses involving sports-related digital assets. The methodology section describes the machine learning algorithms employed, the process of data collection, variable construction, and model evaluation metrics. Finally, the results are discussed, emphasizing both the predictive power of the models and the impact of individual variables on NFT pricing. Through this exploration, this study engages with a cutting-edge topic at the nexus of digital technology, sports analytics, and economic research.

Literature Review

Definition of NFTs

According to Sestino (2022), Non-Fungible Tokens (NFTs) are cryptographic ownership certificates for digital objects such as artwork, music, or online features. These tokens are unique, not interchangeable, and their authenticity and ownership are verified through blockchain technology. This uniqueness allows NFTs to embody verifiable ownership and authenticity of digital items. Within the sports industry, NFTs serve as a new way for fans to engage, representing game-worn jerseys, limited-edition digital trading cards, or digital memorabilia (Glebova, 2023). These definitions highlight not only the theoretical underpinnings of NFTs but also their transformative potential in sports fan engagement.

Market Dynamics and Valuation Factors

The fervor surrounding the NFT secondary market mirrors the dynamism and volatility characteristic of the cryptocurrency sector, positioning it as a burgeoning arena for investment. Scholars such as Nadini (2021), Ho (2023), and Polat (2023) emphasize that market visibility, platform popularity, and social media engagement significantly affect NFT valuations. Nadini (2021) identifies Twitter and Discord follower counts as critical predictors of NFT demand and price. Similarly, studies like Barolli (2019) and Pelechrinis (2023) demonstrate that fan engagement metrics and social network activity play pivotal roles in shaping the pricing of "NBA Top Shot" NFTs.

Machine Learning in NFT Valuation

Recent research highlights the adoption of machine learning algorithms to predict NFT prices and uncover the factors influencing their valuations. For instance, Nadini(2021) and Wang (2023) employ machine learning to predict prices based on intrinsic (e.g., rarity and content) and extrinsic (e.g., social media influence) factors. Pelechrinis's (2023) use of random forest models for "NBA Top Shot" valuation underscores the effectiveness of these algorithms in identifying key variables such as platform popularity and fan interactions.

Building on these studies, this research employs machine learning algorithms—random forest, logistic regression, and XGBoost—to predict "NBA Top Shot" NFT prices. By evaluating model performance and identifying the most influential variables, this study contributes to understanding the dynamics of NFT valuation within the sports domain. Unlike prior work that often emphasizes prediction accuracy, this paper aims to illuminate the relative importance of performance metrics and external factors in determining NFT prices.

This body of literature underscores the potential of machine learning in advancing the analysis of NFTs. However, a gap remains in integrating sports performance metrics with these models to analyze their effect on NFT valuations. By bridging this gap, this research provides insights for investors, fans, and stakeholders into the economic implications of athletic performance on NFT pricing.

Data Description

The dataset employed in this study integrates data from two primary sources: the official "NBA Top Shot" platform and the National Basketball Association's (NBA) official website. This integration facilitates a comprehensive analysis of the relationship between players' on-court performance metrics and the market value of their corresponding NFTs. The data were gathered over a one-week period through web scraping, focusing on real-time information regarding player performance and NFT market activity. This approach allowed for the capture of dynamic fluctuations in NFT pricing and player statistics on specific match days, ensuring the relevance and timeliness of the dataset. Table 1 shows the description of the variables that been used in the model. The dependent variable is “Lowest Ask, which is a binary variable that indicates whether an NFT is actively traded and valued. A value of 1 signifies that the NFT is being actively traded, while a value of 0 indicates inactivity or lack of market interest. The independent variables are “Movement”, “MIN”, “PTS”, “Three”, “ORB”, “DRB”, “AST”, “STL”, “BLK”, “TOV”, “PF”. These variables provide a detailed view of both player performance metrics and contextual factors hypothesized to influence NFT valuations.

Methods

1. Data Preprocessing

Feature Selection

A subset of features was selected based on domain relevance and initial exploration, including both player performance metrics (e.g., ”Minutes Played (MIN),” ”Points Scored (PTS)”) and NFT-specific attributes (e.g., ”Movement”). The target variable, ”Lowest Ask”, was binarized, with 1 indicating active NFTs with market interest and 0 denoting inactive NFTs.

Encoding Categorical Variables

The LabelEncoder was applied to the ”Movement” feature to convert categorical data into numerical format for model compatibility.

Handling Missing Data

Rows containing missing values were dropped to ensure a clean and consistent dataset.

Feature Scaling

StandardScaler was used to normalize numerical features, ensuring they are on the same scale to enhance model performance during optimization.

2. Class Imbalance Adjustment

The dataset exhibited class imbalance, with significantly fewer active NFTs (Class 1) compared to inactive NFTs (Class 0). To mitigate this issue, the Synthetic Minority Oversampling Technique (SMOTE) was implemented to generate synthetic samples for the minority class, improving the model’s sensitivity to underrepresented data.

3. Feature Selection

Feature importance was assessed using SelectKBest with ANOVA F-statistics (`f_classif`) as the scoring metric. The top eight features were selected to streamline the model and reduce overfitting, including ”Movement,” ”PTS,” ”MIN,” and ”Three-Point Shots (Three).”

4. Model Training and Evaluation

0.0.1 Algorithms

Logistic Regression: A baseline linear model with L1 regularization was chosen for its interpretability and ability to identify key predictive features. Class weights were set to "balanced" to address class imbalance.

Random Forest: A non-linear ensemble method was selected for its ability to handle feature interactions and provide insights through feature importance scores. Class weights were also applied to improve performance on the minority class.

XGBoost: A gradient-boosting algorithm was used due to its strong predictive capability and flexibility in handling complex feature relationships. Hyperparameters such as learning rate, number of estimators, and tree depth were optimized using GridSearchCV.

0.0.2 Evaluation Metrics

Cross-Validated ROC AUC: To ensure robustness and generalizability, cross-validation with five folds was performed, and the mean ROC AUC score was calculated.

Confusion Matrix: Used to quantify classification accuracy, highlighting true positives, false positives, true negatives, and false negatives.

Classification Report: Precision, recall, and F1-score were computed to provide a balanced view of the model's performance.

Feature Importance Analysis: Feature importances from the Random Forest model were extracted to identify the key drivers influencing market activity.

Discussion and Conclusion

Interpretation of Results

The findings of this study offer valuable insights into the factors driving market activity for NBA Top Shot NFTs. Among the models evaluated, the Random Forest algorithm achieved the highest ROC AUC score of 0.823, demonstrating its superior capability to differentiate between active and inactive NFTs. The analysis identified key player performance metrics, including Points Scored (PTS), Minutes Played (MIN), and dynamic gameplay actions such as Blocks (BLK) and Steals (STL), as significant predictors of NFT market interest. Additionally, NFT-specific attributes, such as Movement and Rarity, were validated as critical factors influencing market value. These results align with the hypothesis that high-impact plays and unique attributes enhance the desirability of NFTs.

An intriguing finding was the role of Turnovers (TOV), traditionally considered a negative basketball performance metric. In this study, Turnovers were associated with higher NFT valuations, likely reflecting the market's focus on high-visibility players whose prominence outweighs minor performance flaws. Conversely, Minutes Played (MIN) exhibited a negative correlation with value when excessive, underscoring the market's preference for scarcity and exclusivity. Despite these successes, the study faced certain limitations. The dataset's inherent class imbalance presented challenges, although the application of SMOTE mitigated some of these issues. Furthermore, while the selected features provided strong predictive power, the absence of other potentially influential factors, such as player popularity or macroeconomic trends, constrained the analysis. These gaps highlight opportunities for future research to expand on the current findings.

The study also provided several key visual insights. The feature importance analysis (Figure 1) revealed that Movement was the most influential predictor, followed by Points Scored (PTS) and Minutes Played (MIN). This confirms that dynamic gameplay moments and player contributions are critical in determining NFT value. The cross-validation results (Figure 2) indicate consistent model performance across folds, with the ROC AUC score stabilizing at a high level. This suggests the model generalizes well to unseen data, reinforcing its reliability. The confusion matrix (Figure 3) highlighted the model's ability to classify active NFTs (Class 1) with moderate precision but strong performance in identifying inactive NFTs (Class 0). Although there were some misclassifications, the results validate the model's overall effectiveness.

Limitation

The study underscored the importance of careful feature engineering in achieving strong predictive performance. The combination of player performance metrics and NFT-specific attributes emerged as a critical factor in driving model accuracy. Preprocessing techniques, including StandardScaler for normalization, LabelEncoder for categorical data conversion, and SMOTE for addressing class imbalance, proved essential in improving model performance and ensuring a clean, balanced dataset. The Random Forest algorithm stood out not only for its high predictive accuracy but also for its interpretability, with feature importance analysis providing valuable insights into the key drivers of NFT market activity.

Additionally, the project illuminated nuanced market dynamics, such as the valuation of high-visibility players despite their performance flaws. This observation reflects the market's preference for narrative-driven assets, where player prominence and memorable gameplay moments often outweigh traditional performance metrics. The findings demonstrate how domain-specific knowledge, coupled with robust machine learning methodologies, can provide actionable insights into complex markets.

Future Work

To build on the current findings, future research should consider expanding the feature set to incorporate external factors such as player popularity, social media sentiment, and broader economic conditions. These additions could provide a more comprehensive understanding of the drivers of NFT market activity. Advanced algorithmic approaches, such as deep learning models, could be explored to capture more complex non-linear relationships and interactions between features. Neural networks, for instance, may uncover subtle patterns that are not apparent with traditional machine learning methods.

Furthermore, the analytical framework developed in this study could be applied to other domains, such as art, music, or gaming NFTs, to generalize the findings and explore cross-domain applications. This extension would provide a broader perspective on NFT market dynamics and reveal whether similar predictors and trends are consistent across different industries. Such work could contribute to a deeper understanding of the evolving NFT ecosystem and its intersection with diverse fields.

Conclusion

This study sheds light on the complex interplay between NBA player performance metrics and the market value of their associated NFTs, particularly within the context of "NBA Top Shot." By leveraging machine learning algorithms such as Random Forest, Logistic Regression, and XGBoost, the analysis identified key predictors, including Points Scored (PTS), Minutes Played (MIN), and unique NFT attributes like Movement and Rarity. The findings underscore the importance of dynamic gameplay moments and high-impact player actions in driving NFT desirability. Despite limitations, such as class imbalance and the exclusion of external factors like player popularity, the study demonstrated robust model performance, with Random Forest achieving the highest predictive accuracy. These results not only validate the hypothesis that player performance influences NFT valuations but also provide actionable insights for stakeholders in the NFT ecosystem. Future research should expand on these findings by incorporating additional variables and exploring advanced algorithms to further unravel the dynamics of NFT pricing. This work highlights the transformative potential of combining sports analytics with digital finance, paving the way for innovative valuation strategies and investment opportunities in emerging markets.

References

1. NBA Topshot Marketplace: https://nbatopshot.com/search

2. NBA Official Website: https://www.nba.com/stats/leaders

3. Basketball Reference: https://www.basketball-reference.com/leaders/tov_season.html

4. ESPN NBA players career stats: https://www.espn.com/nba/stats

5. Sestino, Andrea, Gianluigi Guido, and Alessandro M. Peluso. *Non-Fungible Tokens (NFTs): Examining the Impact on Consumers and Marketing Strategies.* Springer Nature Switzerland AG, 2022.

6. Glebova, Ekaterina, and Paulína Mihaľová. "New Currencies and New Values in Professional Sports: Blockchain NFT and Fintech Through the Stakeholder Approach." *Journal of Physical Education and Sport*, vol. 23, no. 5, Art 153, May 2023, pp. 1244-1252, DOI: https://doi.org/10.7752/jpes.2023.05153.

7. Ho, Kin-Hon, et al. "Spillover Analysis on NFTs, NFT-Affiliated Tokens, and NFT Submarkets." *Finance Research Letters*, vol. 60, 3 Dec 2023, p. 104598.

8. Klein, Niklas Konstantin, Fritz Lattermann, and Dirk Schiereck. "Investment in Non-Fungible Tokens (NFTs): The Return of Ethereum Secondary Market NFT Sales." *Journal of Asset Management*, vol. 24, no. 4, Palgrave Macmillan UK, 2023, pp. 241–254, DOI: https://doi.org/10.1057/s41260-023-00316-1.

9. Huang, Jintao, Ningyu He, Kai Ma, Jiang Xiao, and Haoyu Wang. "Miracle or Mirage? A Measurement Study of NFT Rug Pulls." *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 7, no. 3, ACM, 2023, pp. 1–25, DOI: https://doi.org/10.1145/3626782.

10. Polat, Onur. "Dynamic Interlinkages between Cryptocurrencies, NFTs, and DeFis and Optimal Portfolio Investment Strategies." *China Finance Review International*, 2023, DOI: https://doi.org/10.1108/CFRI-03-2023-0061.

11. Nadini, Matthieu, Laura Alessandretti, Flavio Di Giacinto, Mauro Martino, Luca Maria Aiello, and Andrea Baronchelli. "Mapping the NFT Revolution: Market Trends, Trade Networks, and Visual Features." *Scientific Reports*, vol. 11, no. 1, Cornell University Library, 2021, pp. 20902–11, DOI: https://doi.org/10.48550/arxiv.2106.00647.

12. An, Jaehyung, Alexey Mikhaylov, and Tsangyao Chang. "Relationship between the Popularity of a Platform and the Price of NFT Assets." *Finance Research Letters*, vol. 61, Elsevier Inc, 2024, pp. 105057-, DOI: https://doi.org/10.1016/j.frl.2024.105057.

13. Ho, Kin-Hon, Yun Hou, Tse-Tin Chan, and Haoyuan Pan. *Analysis of Non-Fungible Token Pricing Factors with Machine Learning.* IEEE, 2022, pp. 1161–66, DOI: https://doi.org/10.1109/SMC53654.2022.9945566.

14. Alon, Ilan, et al. "Predictors of NFT Prices: An Automated Machine Learning Approach." *JGIM*, vol. 31, no. 1, 2023, pp. 1-18, DOI: http://doi.org/10.4018/JGIM.317097.

15. Fitria, Vivi Aida, Arif Nur Afandi, Aripriharta, Lilis Widayanti, and Danang Arbian Sulistyo. "NFT Investments Analysis: A Strategic Approach with Ranking Insights and Sales Forecasting System for Informed Decision-Making." *The Asian Journal of Technology Management*, vol. 16, no. 2, Institut Teknologi Bandung, School of Business and Management, 2023, pp. 95–108, DOI: https://doi.org/10.12695/ajtm.2023.16.2.2.

16. Barolli, Leonard, Hiroaki Nishino, Tomoya Enokido, and Makoto Takizawa. *Card Price Prediction of Trading Cards Using Machine Learning Methods*, vol. 1036, Springer International Publishing AG, 2019, pp. 705–14, DOI: https://doi.org/10.1007/978-3-030-29029-0_70.

17. Pelechrinis, Konstantinos, Xin Liu, Prashant Krishnamurthy, and Amy Babay. "Spotting Anomalous Trades in NFT Markets: The Case of NBA Topshot." *PloS One*, vol. 18, no. 6, Public Library of Science, 2023, pp. e0287262–e0287262, DOI: https://doi.org/10.1371/journal.pone.0287262.

Appendix

Name	Type	Meaning
Lowest_Ask	Binary	Indicates whether an NFT is actively traded and valued. 1 means the NFT is actively traded and cared about; 0 means the NFT has no active trade or interest.
Movement	Categorical	Type of player movement captured in the NFT, e.g., assist, dunk, block, etc.
MIN	Continuous	Total minutes played by the player in the game. Represents game time contribution.
PTS	Continuous	Points scored by the player in the game. Indicates scoring ability.
Three	Continuous	Number of successful three-point shots made by the player. Highlights offensive efficiency.
ORB	Continuous	Offensive rebounds collected by the player. Reflects offensive hustle.
DERB	Continuous	Defensive rebounds collected by the player. Highlights defensive ability.
AST	Continuous	Assists made by the player. Represents playmaking skills.
STL	Continuous	Steals made by the player. Reflects defensive contributions.
BLK	Continuous	Blocks made by the player. Indicates defensive ability and physical dominance.
TOV	Continuous	Turnovers committed by the player. Reflects errors and risks taken during the game.
PF	Continuous	Personal fouls committed by the player. Indicates discipline issues or defensive aggressiveness.

Table 1: Variable Description

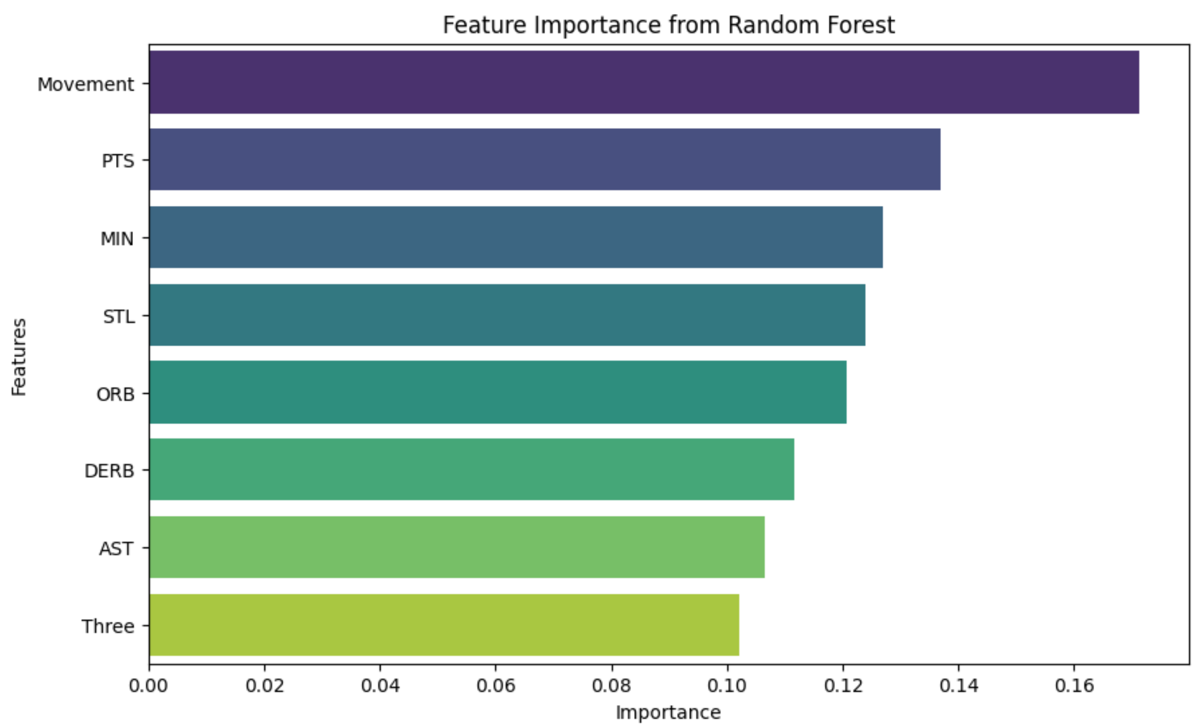


Figure 1: Feature Importance from Random Forest

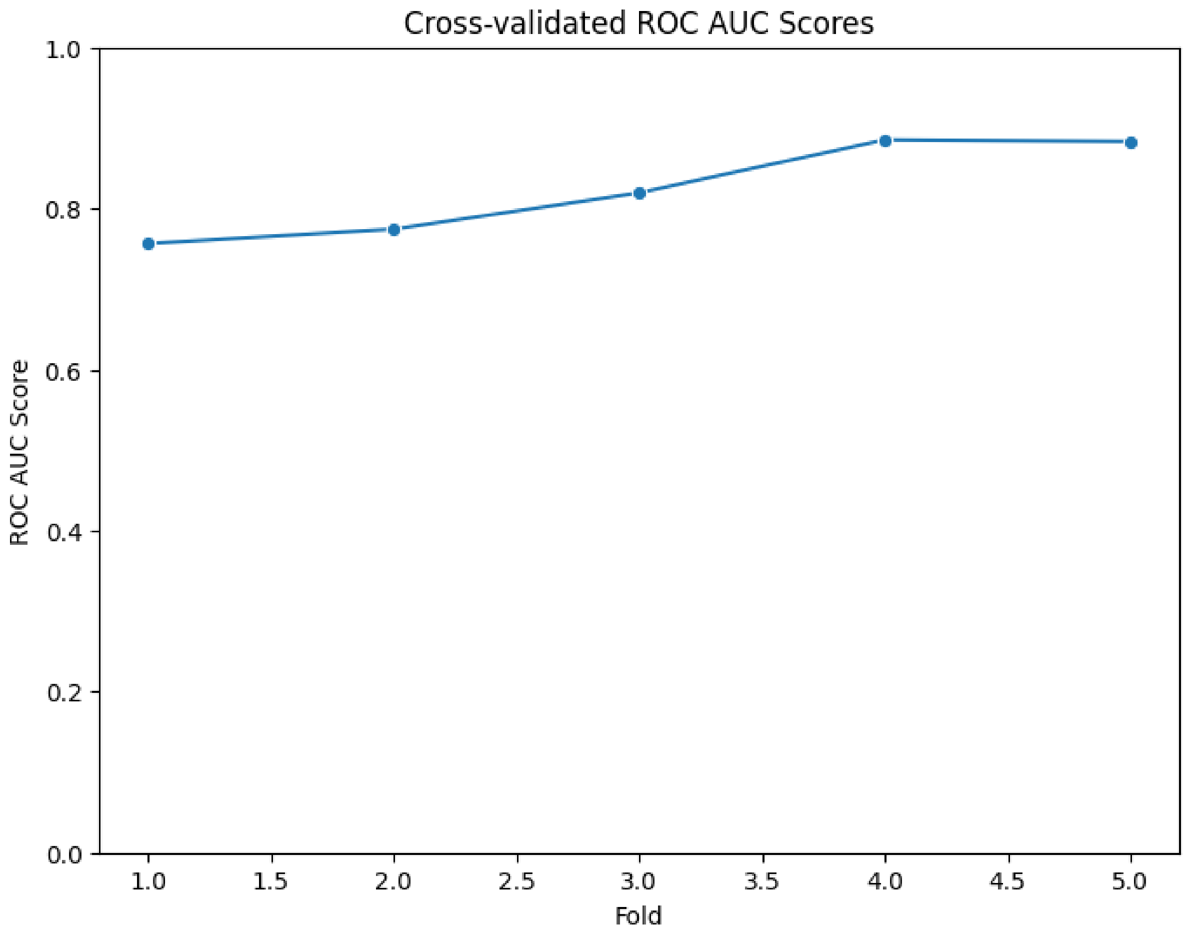
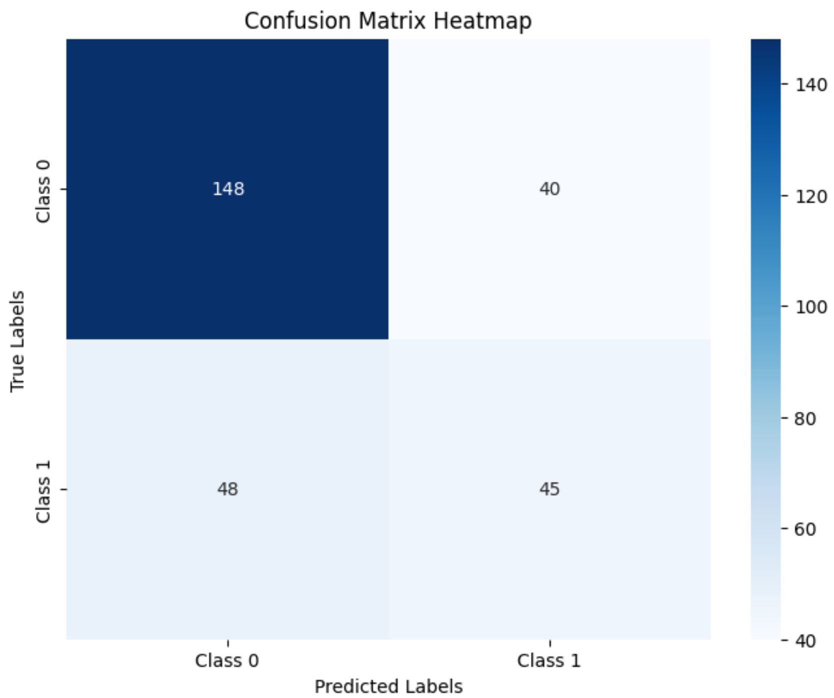


Figure 2: Cross-Validated ROC AUC Scores



[H] Figure 3: Confusion Matrix Heatmap

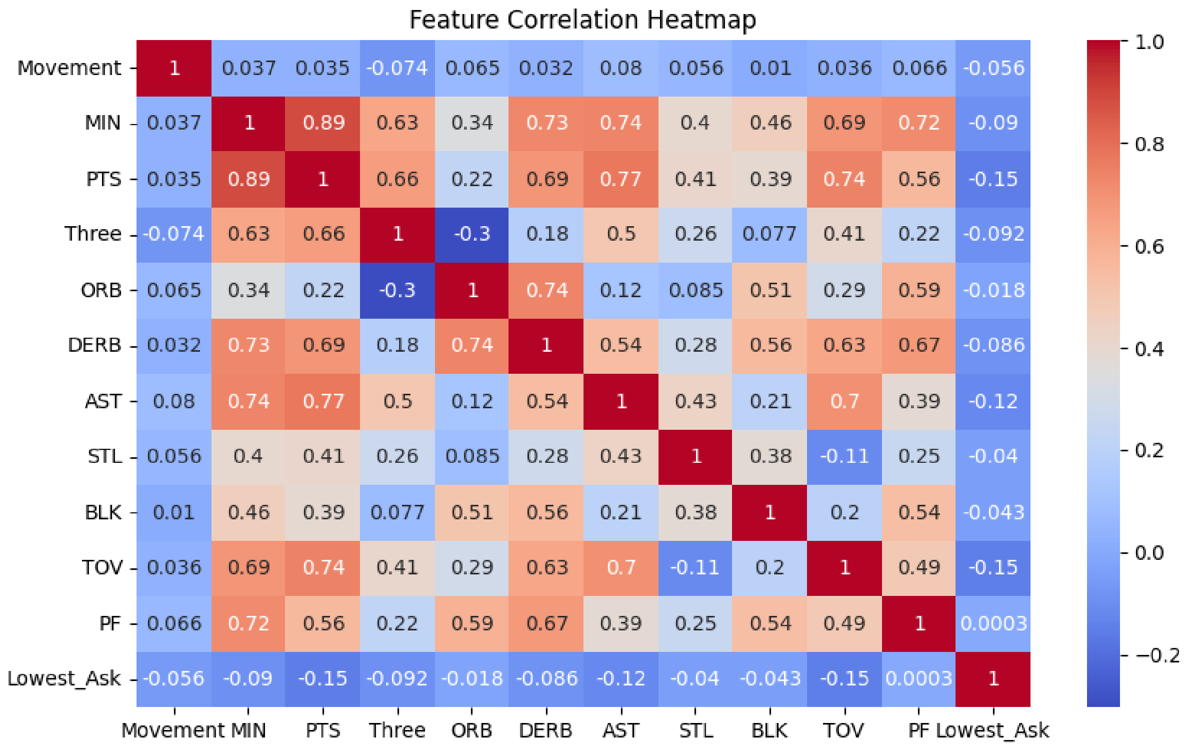


Figure 4: Feature Correlation Heatmap