Predicting NBA Player Career Outcomes and Performance Using Machine Learning and Statistical Modeling

Cameron Moore#1, Nathan Kizlinski#2

#1 *Eberly College of Science,*
The Pennsylvania State University,
United States of America

#2 *College of Information Sciences/Technology,*
The Pennsylvania State University,
United States of America

[1] cjm7736@psu.edu
[2] njk5644@psu.edu

Abstract— This document will explore research completed by multiple studies regarding recent developments in NBA player analytics and performance prediction using machine learning and data driven modeling techniques. The research papers discussed in this document focus on different topics such as long-term career prediction using early career statistics, performance forecasting through regression based models, and the identification of trends relating to player age, position, injury history, and economic outcomes in the NBA. Each study uses different datasets and machine learning methodologies, but they all aim to better understand the different factors that influence success, longevity, and value in the NBA. Through this document, readers will be able to see examples of three (3) research papers that cover long-term career prediction, machine learning performance modeling, and the impact of aging and injury patterns on athletic and economic outcomes. In addition to reviewing these research papers, this document will also take a deeper look at the methods used in the parent paper which focuses on predicting whether an NBA player will reach a full ten year career based on early career performance. We will also explore the motivations behind these studies and how their findings can assist NBA organizations, coaches, analysts, and front offices in making more informed decisions that improve roster building, player development, and long-term planning.

Keywords— Machine Learning, NBA Analytics, Player Prediction, Career Longevity, Injury Impact, Regression Models, Sports Economics, Data Mining, Player Evaluation

I. INTRODUCTION

The purpose of this document is to review, discuss, and assess the results of a set of three (3) research papers that are all related to the rapidly growing field of NBA analytics and sports data science. These research papers each deal with different aspects of player performance and career outcomes, using machine learning and data mining techniques to analyze large player datasets and uncover useful patterns. The papers examined in this document will be focused on the goals of predicting long-term NBA career attainment, forecasting player performance using machine learning regression models, and using data mining techniques to uncover the impact of age, position, and injury history on performance and economic value in the NBA. Once we have discussed the contents of the set of three (3) related research papers, we will take a more detailed look at the work conducted in the first research paper, which will be referred to as our "parent paper" for this document. The parent paper explains how early career statistics, even those limited to just their first two seasons, can be powerful predictors for whether that player will reach full retirement attainment, or ten or more seasons played. This ten year benchmark is meaningful, as it usually represents players who have shown timeless durability, performance, opportunity, and consistency that is rewarded with monetary retirement benefits. Now, we will briefly introduce each of the three (3) research papers to be fully discussed in the "Literature Review" section of this document.

The first research paper [1], our parent paper, is titled "Predicting Full Retirement Attainment of NBA Players." The research team in this paper aimed to determine whether a player's early career performance could be used to classify whether they would reach a ten-year NBA career. By collecting rookie and sophomore season data from players who entered the NBA from 1999-2006, the researchers developed several machine learning models to identify factors that yielded long term success. This will serve as a foundation for understanding how machine learning can predict long-term career success in the league.

The second research paper [2] that we will be further discussing is titled "Enhancing Basketball Team Strategies Through Predictive Analytics of Player Performance", which focuses on predicting game statistics, such as points, rebounds, and blocks. The purpose of this research paper was to highlight how performance metrics can be forecasted using advanced modeling techniques. This paper served as an opportunity for us to evaluate the researchers' methodologies, parameters, and variables to create predictions. We can also see how they compare to our parent paper.

The third research paper [3] that we will be further discussing is titled "Sports Analytics: Data Mining to Uncover NBA Player Position, Age, and Injury Impact on Performance and Economics." The research in this paper takes a data mining and clustering approach to examine how position, aging curves, and injury history shape both on-court performance and long term earning potential. Through these methodologies, the researchers were able to identify key trends that explain career length, salaries, and overall player decline as a player progresses in their career.

Our interest in this topic came from the increasing use of analytics in the NBA. Over the past decade, NBA organizations have placed a much greater emphasis on gathering and interpreting large amounts of data. The modern NBA is driven by a lot of information such as advanced tracking systems, detailed injury reporting, and expanded statistics such as player efficiency rating, defensive impact metrics, and lineup data. Data alone does not help teams make better decisions, but instead it matters how they interpret the data and which techniques they use to reveal findings that support roster building, draft evaluation, contract negotiations, and game strategy. Machine learning is now used frequently across the league and its usefulness only continues to grow. Predictive models can help teams evaluate player potential earlier and more accurately, identify undervalued free agents, predict declines in performance due to aging or injuries, and determine which players are most likely to remain productive over long careers. For these reasons, the topic of NBA career prediction and performance analytics is not only relevant to sports science but also to broader discussions about machine learning applications in real world settings.

Our purpose for writing this document is to combine the key findings from each of the three papers to create a comprehensive analysis of how machine learning models can be used to improve NBA player evaluation. Each of the papers add something different to our understanding of the league: the parent paper focuses on long-term career survival, the performance prediction paper focuses on predicting a specific performance metric, and the data mining paper focuses on aging, injuries, and economic trends. When put together, these studies provide a broad picture of how data science can assist NBA teams in making stronger, more informed decisions.

Another reason that this topic is worth covering is because of the evolution of the NBA over the past twenty years. As basketball strategies have shifted towards a faster paced game, greater emphasis on spacing, and more reliance on three point shooting, evaluation methods have also changed accordingly. The league now features more positionless basketball, more international players, and more advanced training and injury prevention systems. These changes mean that traditional scouting alone may not be enough when evaluating player potential. Machine learning methods can help fill this gap by identifying trends that reflect how a player might develop over time, what their strengths are, and which weaknesses may limit their long-term success.

In addition, the growing availability of detailed NBA data such as player tracking information and shot location charts, has made machine learning techniques even more valuable. These technologies allow analysts to quantify aspects of the game that were previously difficult to measure. This can be seen with factors like off ball movement, defensive rotations, and play efficiency. This means that studies (like the three discussed in this document) are not only valuable but also increasingly necessary as teams look for new ways to gain a competitive advantage.

Collectively, the three research papers provide a complete overview of how machine learning and analytics are being used in modern basketball. They explore topics of prediction, performance modeling, and data mining, which give insights into why certain players succeed, why some players have long careers, and how injuries and aging shape long-term performance. By combining the ideas from these papers, NBA teams can better identify which players are likely to have long-term value and which players may pose greater risks.

II. LITERATURE REVIEW

In this "Literature Review" section, we will further discuss each of the three (3) research papers that we introduced in the previous section. For each paper, we will highlight the importance and relevance to the topic of NBA analytics, the goals of each study, the datasets and methodologies used, and the findings from the researchers. Additionally, we will examine the limitations of each study.

[1] *Predicting Full Retirement Attainment of NBA Players*

As discussed in the introduction, the first research paper [1] we are reviewing will also be our parent paper. We will replicate the results of this paper later in this document. This research paper was published in June of 2025, and written by a team consisting of an independent researcher from Greece and two students from the University of Crete in Greece and Dental School/Medical Faculty in Switzerland. This paper [1] predicts whether NBA players will achieve a ten-year career using early career performance statistics.

The data used by the researchers includes NBA players who entered the league from the years 1999 to 2006. This dataset contained tracked metrics such as player position, minutes played, usage rate, points per game, field goal percentage, three point percentage, rebounds, assists, blocks, steals, turnovers, and more. The researchers utilized this dataset to test multiple machine learning models, such as:

- Ridge Logistic Regression
- Elastic Net Logistic Regression
- Random Forest Classifier
- Gradient Boosted Machine (GBM) Classifier
- Support Vector Machine (SVM)

Each model approaches the dataset and problem differently. Ridge regression uses regularization to control for correlated features. Elastic Net combines Ridge and Lasso regression, while Random Forest and GBM Classifier capture nonlinear patterns. SVM finds optimal separation boundaries in the data. Attempting different models is a great way to see how many different models perform, which can be compared and validated to create the strongest prediction model.

The researchers of this paper found that the Ridge Logistic Regression model produced the strongest performance. This technique achieved the highest accuracy and also demonstrated stable predictions, despite multiple testing conditions. The researchers explain that the regularization used by Ridge Regression helps prevent overfitting and effectively handles the multicollinearity present in player statistics.

Several limitations of the paper are noted. This study does not include injury data, changes in coaching, team strategy, off court behavior, or other external factors that could impact a player's career. Although some are hard to quantify, these can lead to stronger prediction accuracy and could/should be included. Binary classification gives us some insights, but it can lead to a few pitfalls. For example, if two players had significantly different career lengths (2 year and 8 year careers) the model would correctly predict they both do not last 10+ years in the NBA. However, their performance, salary, and injury statistics could be vastly different, hiding other information or otherwise unknown patterns that would result in stronger predictions or insights.

Through this research, we were able to establish a baseline that early career performance is a strong predictor of career longevity, and specifically variables like playing time and scoring efficiency strongly predicts long term career outcomes. We were also able to determine that regression models can be used to predict career length.

[2] *Enhancing Basketball Team Strategies Through Predictive Analytics of Player Performance*

The second research paper [2] we are reviewing was published in May of 2025, written by a team of research students from the Dundalk Institute of Technology and Lero, the Research Ireland

Centre for Software. This paper applies machine learning regression models to predict performance statistics, specifically rebounds.

The researchers used data from over 1,300 NBA rookie seasons, consisting of 21 variables such as games played, total minutes played, and points. The dataset includes detailed rebounds statistics such as offensive and defensive rebounds. The overall goal of this research was to see if there is a link between player statistics and team performance, while also determining which model yields the most accurate predictions.

They examined many different regression models, including:

- Linear Regression
- K-Nearest Neighbors (KNN)
- Multi-Layer Perceptron (Neural Network)
- Random Forest
- Gradient Boosted

To improve model performance, the researchers applied Recursive Feature Elimination (RFE) to determine which variables have the greatest impact on rebounding performance. They also used hyperparameter tuning to fine tune the data and find optimal results for each model.

The results of the study showed that Gradient Boosting had the highest predictive accuracy, with other models such as traditional linear regression performing slightly worse (however still strong and very accurate overall). Researchers were able to use these models to identify and remove the less important features, which reduced model noise and improved accuracy.

There are a couple limitations with this study. Like paper [1], injury data is not included and the models created by the researchers do not account for era adjustments within the NBA. Additionally, factors like team chemistry or fatigue levels aren't included in the data. As mentioned before, these can be crucial variables that allow for enhanced predictions. Through this research, we can improve on our ideas from the first paper [1] by implementing stronger parameter tuning and adjustments to strengthen our models. Additionally, we can adopt their methods of creating new efficiency metrics to strengthen our work.

[3] *Sports Analytics: Data Mining to Uncover NBA Player Position, Age, and Injury Impact on Performance and Economics*

The third research paper [3] we are reviewing was published in April of 2024, and written by a team of research students from the International Hellenic University. This paper [3] uses data mining techniques to determine how player position, age, and injury history affect players' performance and economic value. Unlike paper [1] and paper [2], this paper [3] also uses exploratory approaches to reach findings, as opposed to solely predictive modeling seen previously.

The researchers used data collected from many different sources, including ESPN and websites like NBA Basketball Reference from the years 1999 to 2006. Variables included more in-depth information about players beyond traditional statistics, such as:

- Injury history
- Position information
- Contracts (values, length, etc.)
- Demographic data

By applying algorithms such as K-Means Clustering, Cost Estimation Model, and association rule mining, the researchers were able to uncover meaningful patterns and correlations relating to player success and decline.

Through this research, they found that player age and injury history are two of the most important and critical predictors of performance and salary trends. Young players tend to improve rapidly early in their careers, especially if they receive constant playing time. As players age, especially after 30 years old, their performance quickly declines. Some positions such as Guards require constant speed and agility throughout a game, and have steeper declines after age 30. On the other hand, positions that require less from these skills like Centers are able to have longer careers.

There are a few limitations with this research study. Although the injury data is included, there are potential gaps with this information. Historically, NBA players can tend to play through injuries unknowingly and/or keep injury information private, which can skew data or represent inaccurate injury information. This would result in incorrect models and possibly the wrong interpretation. Additionally, this research focuses more on post career analysis as opposed to early career prediction, which isn't exactly aligned with the goals of our research.

Despite these limitations, the results of this study can help us improve on our ideas from previous papers [1-2] as it led to our realization that integrating player position, salary contract, and physical indicators like injury history return stronger models and more accurate predictions.

III. CONTRIBUTIONS

Even with all the extensive and informative research that we have covered thus far in the "Literature Review" section of this paper, we determined that more can be done to enhance the results of others' research. However, as with implementing any new ideas, unexpected challenges and obstacles arose along the way. This required us to adapt and take different approaches to contribute our knowledge. In this section, we outline our original research plan, what happened as we attempted to follow said plans, and how the challenges faced reshaped our final strategies and implementations.

A. ORIGINAL PLAN FOR IMPLEMENTATION

Our original plan was to closely replicate the methods, code, and results outlined in the parent paper, which utilized sophomore season performance metrics and statistics to predict which players would achieve a career of ten or more seasons. We then planned to use the methodologies and results from our parent paper to create robust models that more accurately predicted career longevity. We soon learned that there would be difficulties with this implementation strategy.

B. WHAT HAPPENED

The code and data set the parent paper used was not published or publicly available, so we needed to find a different data set and implement their code as best we could. We obtained two data sets [4 and 5] from Github that are both publicly available to view and use. As we progressed through the code implementation phase, several challenges emerged that complicated our approach and revealed new limitations with the previous research.

The most pressing issue was the substantial class imbalance present in the data sets, which was not resolved by the parent paper. Only about a quarter of the players in the data sets achieve the ten season benchmark, leaving an overwhelmingly large majority who had short careers. This distribution created bias within the models, which consistently demonstrated tendencies to predict shorter careers. A change was needed to accurately account for this imbalance.

C. NEW PLAN FOR IMPLEMENTATION

We created a new implementation strategy that kept the core research question while improving the thoroughness of the paper. Our new approach used several innovative techniques in order to address the limitations while building on the parent paper's original work.

The first improvement we made was using SMOTE (Synthetic Minority Over-Sampling Technique) which addresses the class distribution issue. We integrated SMOTE into our model to fix the problem of having more short career players compared to long term. SMOTE creates synthetic examples of players with long careers to balance out the data. We made sure to do this carefully so our results would be honest and accurate.

We also diversified our model approach by introducing XGBoost as an alternative to the standard Gradient Boosting implementation used in the original research. We used XGBoost because it is known for handling complex data patterns well and is computationally efficient. We also attempted ensemble stacking which combines predictions from three different models (Logistic Regression, Random Forest, and XGBoost) and uses a final model to make the ultimate decision. This model decides how to combine all the predictions in the best way in order to lead to a better overall performance.

We then created a systematic comparison framework to test whether these new techniques could improve upon the Logistic Regression baseline from the parent paper.

IV. DATASETS

We had to find a dataset that was similar to our parent paper, because it did not have an available data set. The data set we found and which is being used in both our original code and our additional novelties is a publicly available NBA player statistics data set from a GitHub repository, which contains historical records of player careers and seasonal performance metrics. The data spans from the early years of the NBA through the present day which provides a view of player development and career outcomes across different eras of basketball.

The most important attributes we will be using in our analyses are player identification, career length or number of seasons played, seasonal performance statistics, player position, and experience level. These attributes form the foundation of our predictive modeling approach.

The data set comes in two parts. The first file provides career level information for each player, including their total seasons played and Hall of Fame status. The second file contains detailed seasonal statistics, tracking performance metrics year by year for each player throughout their career. When merged together, these files create a complete picture of how early career performance relates to long term career outcomes.

Moving forward with our analysis, we will see all of these attributes used in different ways to build and test our predictive models. The complete nature of this data allows us to examine not only whether a player achieves a 10 year career, but also how their performance evolves over time and what statistical patterns emerge among players with different career trajectories.

V. IMPLEMENTATION

Once we picked our data set, the modeling process was established to be consistently accurate, easy to understand, and be computationally efficient. We picked logistic regression as the primary model because it lets us clearly see how each basketball statistic affects the chance of long term career survival. After preparing and cleaning the data, we split it into training and testing sets to make sure our results would hold consistent on new players, not just the ones used to train the model.

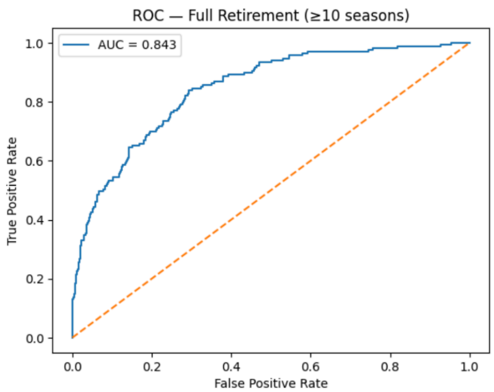
To prevent the model from overfitting, we added regularization. We then ran a grid search to find the best penalty level by testing a wide range of possible values for the tuning parameter C. The final selected value, C being approximately 0.0178, showed that moderate regularization gave us the best balance between flexibility and stability.

In addition to logistic regression, we tested several more advanced models to see whether nonlinear relationships could improve predictions. The random forest model was used to capture potential interactions and more complex splits in the data, while gradient boosting attempted to improve accuracy by learning patterns in a sequential way. These were the models we added to try and improve on our predictive accuracy, however they did not show major gains then the simpler and more transparent logistic model.

Since the number of players who lasted at least ten seasons was much smaller than those who did not, class imbalance became an important factor. We applied SMOTE to address this problem which created synthetic examples of the minority class for comparison runs. This helped us evaluate whether balancing techniques actually improved predictions. Meanwhile, the logistic regression model was kept unbalanced so that its coefficients could remain stable and interpretable.

We also created several visuals to present the model behavior in a straightforward way. These include ROC curves, coefficient rankings, permutation importance charts, and partial dependence plots. These graphs will be displayed alongside this section to offer a clear visual understanding of how the models performed and which variables contributed most to predicting long term NBA survival.

After we finished the model training and tuning process, we evaluated performance on the test set to determine how well the models translated to players not in the training set. The logistic regression model produced a holdout AUC of 0.843 which became our baseline to compare to. This confirmed that the simpler model was able to correctly separate long term NBA players from shorter career players at a very high rate. The ROC curve for this model illustrates a strong upward bend toward the top left corner indicating reliable true-positive capture with limited false-positive inflation.



When novelty models were introduced, performance differences were noticeable but not huge. Even though XGBoost is good at learning complex nonlinear relationships, it had a holdout AUC of 0.799 which was lower than the logistic benchmark. This suggests that while the model captured certain interactions, it also likely overfitted patterns tied to era or role-specific player types. Also, stacked ensembling produced an AUC of 0.838 which was only slightly below the logistic model. This indicates that most of the predictive information in the data set was already captured by linear relationships rather than nonlinear layering. To address the class imbalance, we applied SMOTE and ran the logistic model again using the balanced training data. This version reported a holdout AUC of 0.844 which was a very slight improvement compared to the original unbalanced model. The small change suggests that while balancing techniques help in certain classification settings, the original structure of NBA data was not skewed in a way that affected logistic regression's ability to learn separating patterns.

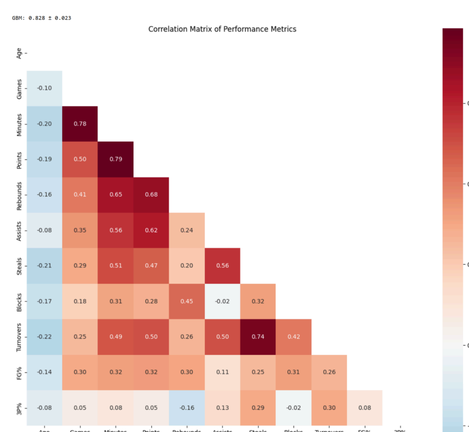
```
Holdout AUC: 0.843

NOVELTY 1: XGBOOST
XGBoost Holdout AUC: 0.799

NOVELTY 2: ENSEMBLE STACKING
Stacking Holdout AUC: 0.838

NOVELTY 3: SMOTE
Class distribution: 0.286
SMOTE Holdout AUC: 0.844
COMPARISON TABLE: NOVELTY VS ORIGINAL
                Model       AUC Improvement
0         Original LR  0.843468    Baseline
1             XGBoost  0.798761     +-0.045
2   Ensemble Stacking  0.838068     +-0.005
3           SMOTE + LR  0.843706     +0.000
```

Next, we looked at which features most influenced career survival. Both logistic coefficients and permutation importance highlighted age, points per game, rebounds per game, blocks, and total games played as the top drivers. Age showed the strongest negative effect while scoring and playing time showed the strongest positive effect. Following is the coefficient table and importance chart which shows the rankings.

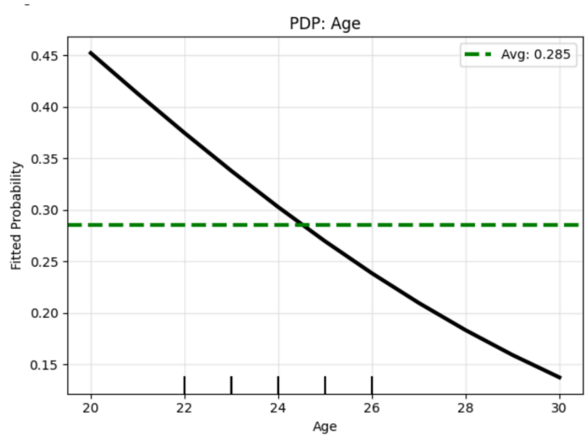
```
          feature      coef
0             age -0.417601
3             ppg  0.274175
4             rpg  0.265019
7             bpg  0.225331
1               g  0.222307
15         num_tm -0.215761
5             apg  0.175342
2              mp  0.158997
10     fg_percent  0.137535
12     ft_percent  0.133864
13         ast_to  0.114303
9            pfpg  0.090947
14  ast_plus_pts_to  0.060601
19         pos_PF -0.035402
18          pos_G -0.028849
20         pos_PG  0.025830
8             tpg  0.021925
6             spg  0.020333
17          pos_F  0.016156
22         pos_SG  0.010479
```

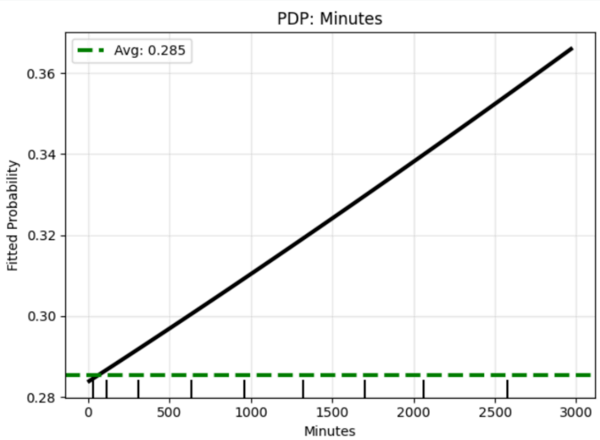
```
          feature  importance
0             age    0.021530
15         num_tm    0.009859
5             apg    0.008600
3             ppg    0.007976
1               g    0.007678
7             bpg    0.007271
4             rpg    0.004885
12     ft_percent    0.004352
13         ast_to    0.003831
10     fg_percent    0.003432
2              mp    0.003120
14 ast_plus_pts_to    0.001139
20         pos_PG    0.000775
6             spg    0.000378
19         pos_PF    0.000297
```

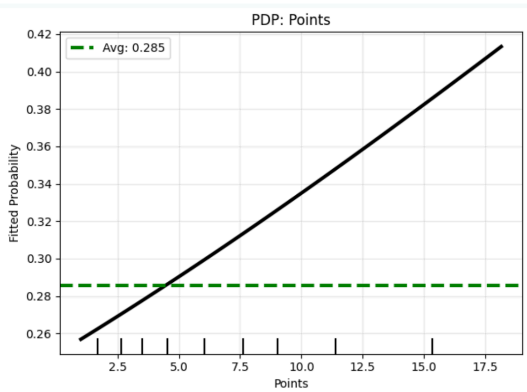
We also made a full correlation matrix of performance metrics to see how much overlap existed between player stats. For example, scoring, assists, rebounds, and efficiency metrics were highly related which helps explain why logistic regression worked so well since most of the predictive information was concentrated in a few core areas rather than spread out across independent variables.

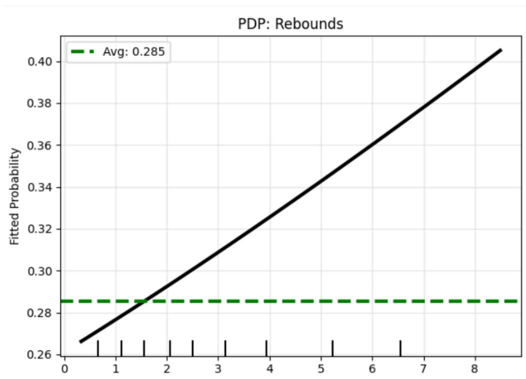


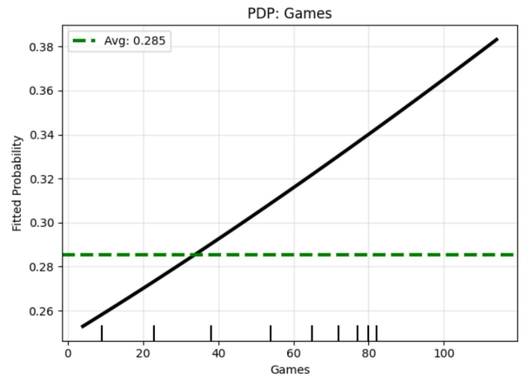
Finally, partial dependence plots give a more intuitive view of how predictions change as performance improves. For example, survival probability increases sharply with scoring and minutes per game but plateaus once certain thresholds are reached, while age shows a consistent downward survival trend as it rises. Below are some examples of the plots we made.











VI. NOVELTY

The true novelty factor that we have the privilege to add to this research is the integration of Synthetic Minority Over-Sampling Technique, or SMOTE. As previously mentioned, the parent paper did not provide us with full access to their data and code. This required us to independently reconstruct and follow all steps of the data science pipeline. We obtained different data sets, cleaned and visualized the data, modeled the data, interpreted the results, and revisited the model. Although we implemented different data sets and code, we were able to closely replicate the results obtained in the parent paper. This allowed us to accurately assess and compare our results with the original parent paper.

Our novelty becomes applicable in the modeling data step. When coding our models, we used SMOTE as a main strategy for our predictions. Previous research has shown no solution for the large class imbalance, which SMOTE handles efficiently. With this novelty implementation, we were able to improve the research and results obtained by the parent paper by an AUC score of 0.001. Creating synthetic data to balance the data sets and models to mitigate short-career prediction bias is now proven to improve machine learning accuracy. This confirms our original question and hypothesis that machine learning techniques can be implemented with second year season statistics to predict if an NBA player will last 10+ years in the league.

A second important novelty to the research we would like to note is our implementation of the XGBoost model. We attempted this machine learning strategy in hopes of achieving a stronger AUC score. Unfortunately, this wasn't successful, as this model returned an AUC score 0.044 lower than the parent paper. However, this was still an important step towards our final model and result. Through this step we then understood that complex models don't always return the best predictions. We found that through finer data preprocessing with simple models, higher AUC scores will follow.

Our third novelty contribution is our implementation of ensemble stacking. This approach produced an AUC score of 0.838, which is 0.005 less than the parent paper. Although counting for another failed attempt at outperforming the parent paper's AUC score, this gave us insights that lead to our successful model. Like the XGBoost model, ensemble stacking proved that complex models cannot consistently improve AUC scores or prediction models. Stacking also helps reduce variance and the risk of overfitting.

We believe that our novelty contributions to previous research is notable. Through our work, predictions for NBA player longevity are more accurate, and can be used to gain competitive advantages for NBA players, agents, coaches, and organizations. We emphasized the importance of balancing NBA player statistics when predicting career length and performance. New patterns were also discovered when uncovering information about positions and their different career trajectories. This research also provides a solid foundation for future research, which we will discuss in the next section.

VII. FUTURE WORK

This research has shown that it has created a stronger foundation for predicting NBA career longevity, using machine learning techniques applied to performance statistics. As we progressed through the different phases of our research project, there were additional ideas that came to mind. We found that there is a lot of opportunity for future work which can lead to increasingly stronger model accuracy. After discussing multiple topics for future research, we have narrowed down to our three most significant ideas, which are as follows.

First, our project would benefit from the switch of a binary classification to a multi-classification prediction framework. Our current model is effective at identifying players who are likely to achieve 10 year careers, however NBA organizations usually require more in depth and enhanced predictions. We could implement a tiered classification system, grouping career lengths as "short" (0-3 years), "medium" (4-7 years), and "long" (8+ years) to allow for deeper insights and more refined decision making. By moving beyond a yes/no outcome, this framework would give a clearer understanding of career trajectories, which enables for intricate scouting evaluations, developmental planning, and long term team strategies. This could also let teams differentiate between career types, such as short term role players or franchise players (those that stay with one team their whole career).

Our second idea for future work involves combining performance statistics with injury and salary data. Throughout this project, we noticed that many previous papers (including ones we have mentioned) struggled with the issue of missing injury data. Injuries are a large part of the game, and by adding this data into our models, it would return more realistic predictions. Accounting for injuries, a task that isn't as commonly practiced by other papers, improves overall model accuracy and would be useful for players and trainers. Additionally, implementing salary data gives players, coaches, and organizations a clearer understanding how financial investments align with projected career longevity and performance.

Finally, the third future research idea we discussed was the incorporation of advanced tracking data and physical metrics. We believe other metrics such as real-time data like player movement, defensive impact, speed, fatigue levels. and spacing efficiency would give the models a more complete view of different potential career outcomes. Although these variables might be hard to track and accurately measure, deep learning techniques might be able to handle the issue. This implementation would provide organizations, managers, coaches, and players with stronger justification for adopting these models.

VIII. REFERENCES

[1] R. Chandru, A. Kaushik, and P. Jaiswal, "Enhancing Basketball Team Strategies Through Predictive Analytics of Player Performance," *Electronics*, vol. 14, no. 11, p. 2177, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/14/11/2177

[2] G. Foutzopoulos, N. Pandis, and M. Tsagris, "Predicting Full Retirement Attainment of NBA Players," *Journal of Data Science and Analytics*, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s41060-025-00821-z

[3] V. Sarlis and C. Tjortjis, "Sports Analytics: Data Mining to Uncover NBA Player Position, Age, and Injury Impact on Performance and Economics," *Information*, vol. 15, no. 4, p. 242, 2024. [Online]. Available: https://www.mdpi.com/2078-2489/15/4/242

[4] HectorOik. "NBA-Career-Length-Prediction/Player Career Info.csv at Main - HectorOik/NBA-Career-Length-Prediction." *Github, https://github.com/HectorOik/NBA-Career-Length-Prediction/tree/main*

[5] HectorOik. "NBA-Career-Length-Prediction/Player Totals.csv at Main - HectorOik/NBA-Career-Length-Prediction." *Github, https://github.com/HectorOik/NBA-Career-Length-Prediction/tree/main*