PREDICTION OF BASKETBALL GAME RESULTS USING MACHINE LEARNING ALGORITHMS: ANALYSIS OF NBA & TBL

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

CANER KAHRAMAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

NOVEMBER 2022

Approval of the thesis:

PREDICTION OF BASKETBALL GAME RESULTS USING MACHINE LEARNING ALGORITHMS: ANALYSIS OF NBA & TBL

submitted by **CANER KAHRAMAN** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Statistics, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Özlem İlk Dağ
Head of the Department, **Statistics** _____

Prof. Dr. Barış Sürücü
Supervisor, **Statistics Dept., METU** _____

Examining Committee Members:
Prof. Dr. Ceylan Talu Yozgatlıgil
Statistics, METU _____

Prof. Dr. Barış Sürücü
Statistics, METU _____

Assoc. Prof. Hakan Savaş Sazak
Statistics, Ege University _____

Date: 30.11.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Caner, Kahraman

Signature:

ABSTRACT

PREDICTION OF BASKETBALL GAME RESULTS USING MACHINE LEARNING ALGORITHMS: NBA & TBL

Kahraman, Caner
Master of Science, Statistics
Supervisor: Prof. Dr. Barış Sürücü

November 2022, 81 Pages

There are several factors affecting the basketball game results including team strength, home court advantage, resting, and momentum. Using advanced metrics and data analysis, it has become much easier for teams to measure the impact of these factors on the game results over the past few years, especially in the NBA (National Basketball Association). Only a few studies are performed related to the prediction of the basketball game results, which consider advanced team statistics and player-specific factors together. This study analyzes the variables that affect basketball game results, including overall team strength via Four Factor metrics, home-court advantage, schedule, back-to-back games, momentum, and player-based variables such as maximum points per game for both NBA and TBL (Turkish Basketball League). Afterward, using the analysis findings of these factors, machine learning models are used to predict the game results in NBA & TBL for three seasons in 2016-2017, 2017-2018, and 2018-2019 regular seasons. In this study, ELO Rating Model, Logistic Regression, Support Vector Classifier, Decision Tree, Random Forest, Naïve Bayes, KNN, LGBM, XGBoost, and Neural Network models are used.

Analysis and results show that using superstar players to advance in the league is a valid option in NBA while it is not in TBL. Moreover, TBL is more predictable (up to 77.5%) than NBA (up to 67.5%) since there are power imbalances among teams in TBL and scheduling imbalances in NBA. Also using advanced variables has a better impact on the accuracies in NBA.

Keywords: Machine Learning in Basketball, Sports Analytics, Game Result Prediction, NBA, TBL

ÖZ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE BASKETBOL MAÇ SONUCU TAHMİNİ: NBA VE TBL ANALİZLERİ

Kahraman, Caner
Yüksek Lisans, İstatistik
Tez Yöneticisi: Prof. Dr. Barış Sürücü

Kasım 2022, 81 Sayfa

Bir basketbol maçının sonucunu etkileyen takım gücü, ev sahibi avantajı, dinlenme süresi ve momentum gibi çeşitli faktörler vardır. Başta NBA'de olmak üzere son birkaç yılda takımlar gelişmiş metrikler ve veri analizini kullanarak bu faktörlerin oyuna etkisini ölçebilir hale geldiler. Bu gelişmelere rağmen gelişmiş takım istatistiklerini ve oyuncu bazlı faktörleri bir arada ele alarak basketbol maç sonucu tahmini yapan çok az çalışma bulunmaktadır. Bu çalışmada, hem NBA hem TBL için Four Factor metrikleri ile takım gücü, ev sahibi avantajı, fikstür ile dinlenme süresi, momentum ve takımdaki en skorer oyuncu gibi oyuncu bazlı değişkenler dahil olmak üzere basketbol maçı sonucunu etkileyen değişkenler analiz edilmiştir. Bu faktörlerin analizinden elde edilen bulgular kullanılarak NBA ve TBL'nin 2016-2017, 2017-2018 ve 2018-2019 normal sezon maçlarının sonucunu tahmin etmek için makine öğrenmesi modelleri kullanılmıştır. Bu çalışmada ELO Rating Modeli,

Lojistik Regresyon, Destek Vektör Sınıflandırıcısı, Karar Ağacı, Random Forest, Naïve Bayes, KNN, LGBM, XGBoost ve Neural Network modelleri kullanılmıştır. NBA'de başarılı bir takım olmak için süper star oyuncu üzerinden oynamanın geçerli bir seçenek olduğu, TBL'de ise bunun tam tersi olduğu gözükmektedir. Ayrıca, TBL'de takımlar arasındaki güç dengesizliğinin fazla olması ve NBA'deki fikstür dengesizlikleri nedeniyle TBL'nin NBA'den daha tahmin edilebilir olduğu gözükmektedir (Model performansları TBL için %77,5'e kadar, NBA'de ise %67,5'e kadar çıkabiliyor). Ayrıca gelişmiş istatistiklerin kullanılması NBA'deki tahmin performansı üzerinde TBL'ye oranla daha iyileştirici bir etkiye sahiptir.

Anahtar Kelimeler: Basketbolda Makine Öğrenmesi, Spor Analitiği, Maç Sonu Tahmini, NBA, TBL

To my family

ACKNOWLEDGMENTS

I would like to thank my supervisor Prof. Dr. Barış Sürücü for accepting and encouraging me for this study. He guides me in all aspects during the thesis process, including modelling and basketball perspective.

Thesis during pandemic period was tough, and I may be able to perform my best, thanks to moral support of my family. Thus, I would like to thank my father İzzet Kahraman, my mother Aysun Kahraman, and my sister Cansu Kahraman for supporting and motivating me under any circumstances.

The Scientific and Technological Research Council of Turkey partially funded this work under grant number TUBİTAK 2210-A.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVATIONS

ACB: Spanish Basketball League

ANN: Artificial Neural Network

CAGR: Compound Annual Growth Rate

DRB%: Defensive Rebound Percentage

EFG%: Effective Field Goal Percentage

FG%: Field Goal Percentage

FT%: Free Throw Rate

KNN: K Nearest Neighbor

LGBM: Light Gradient Boosting Machine

NBA: National Basketball Association

NCAA: National Collegiate Athletic Association

ORB%: Offensive Rebound Percentage

PER: Player Efficiency Rating

SFS: Sequential Future Selection

SVM: Support Vector Machine

TBL: Turkish Basketball League

TOV: Turnover Rate

USD: United States Dollar

3PM: Three points made

CHAPTER 1

1 INTRODUCTION

Technology use in the sports industry to collect and process data increased significantly in the last decade. Thanks to value-added outputs of data extraction and analysis, team owners, coaches, analysts and even players can have a better understanding of the critical factors in sports competitions such as performance, mental and physical conditions of players, level of team play concerning role sharing between the players, and strengths and weakness of the team and opposite team. Information and inferences gained from data analysis can help teams build their roster and strategy to increase chances of winning more games and championships.

Recently many sports teams started to benefit from data science applications especially in football, basketball, and baseball (Grand View Research, 2021). The market size of the global sports analytics industry is estimated at around 1.9 billion USD in 2019 and it is expected to reach 5.2 billion USD in 2024 at a 22.0% CAGR (Markets and Markets, 2020). Software applications such as motion analysis, player tracking and video analysis make up approximately 60% of the market size while the remaining part is services such as data analysis, guidance and counselling services predictive analysis, and forecasting (Grand View Research, 2021).

Moreover, data analysis in sports has become so phenomenon that the importance of data science in sports can be seen in other platforms besides NBA, such as universities, data science related blogs and websites such as Kaggle, GitHub, Towards data science and Medium. Even some universities are offering sports analytics programs.

Basketball is one of the leading sports in terms of using data analysis effectively and NBA is the pioneer basketball league in the world in this respect. Team owners, coaches, players, academicians, business people and particularly executives in NBA

created a great culture of using data science. The commissioner of the NBA, Adam Silver, emphasizes the importance of data analytics.

"Analytics are part and parcel of virtually everything we do now" — NBA Commissioner Adam Silver

Thanks to high availability of data and convenient game dynamics, data science applications are performed very effectively in basketball. NBA environment also facilitates these applications since the power of data analytics is widely acknowledged. Some companies such as Genius Sports, Opta, Second Spectrum, Synergy Sports Technology, Stats Perform, and IBM specialize in technology applications and data analysis in basketball. Major data science applications in basketball are player and team analysis, video analysis and health assessment.

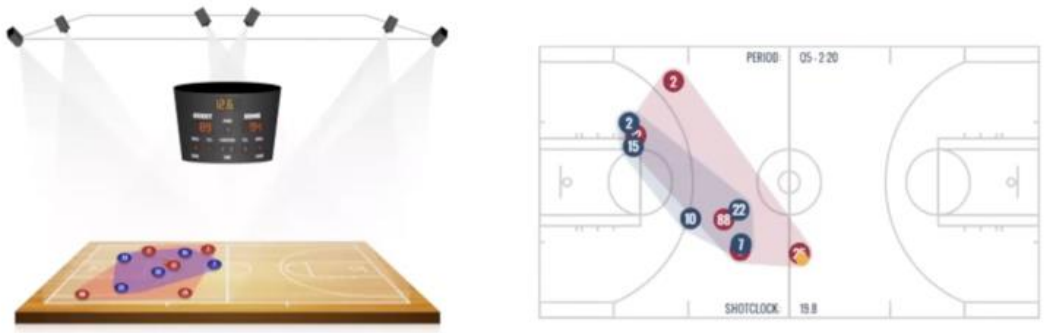


Figure 1: Player tracking technology to collect data from basketball game – Stats Perform

The employment of data analytics has been revolutionary for the NBA. At this point, almost each NBA team has an analytics department. In the last 20 years, thanks to statistical inferences from data analytics studies, basketball became more efficient with a more significant number of skilled players, and enhanced strategies practised by managers. Thanks to the data analytics revolution in the NBA, teams are now building their roster and optimizing the use of players on the court more effectively than in the past. The efficiency increase due to data analytics can be seen by examining the teams' offenses in the last two decades. Figure 2 shows the improvement in 3-points made per game and Effective Field Goal Percentage

(EFG%) in the last 20 seasons. EFG% is a metric for measuring shooting performance that considers the 2-points and 3-points separately. More details about EFG% are presented in the Chapter 3.

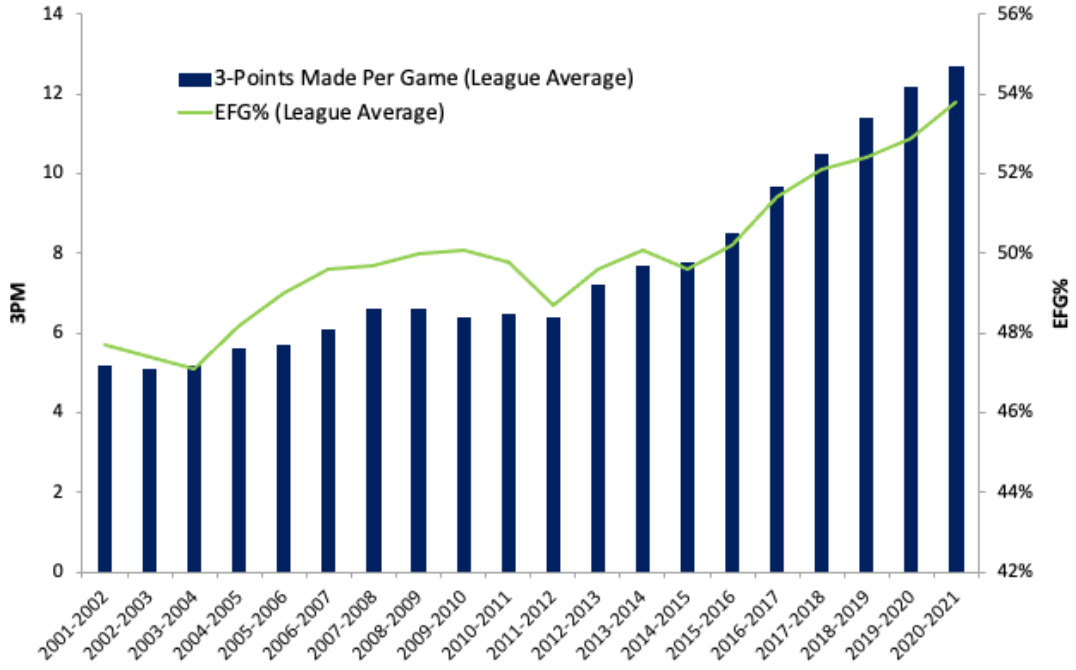


Figure 2: Improvement in 3 points made and effective field goal percentage in last 20 years in NBA

In the 2020-2021 season, the average 3-points made per game is 12.7, which is more than twice the average value in the 2001-2002 season which is 5.2. Moreover, along with the 3-points made per game, EFG% is also increased, which means the game become more efficient.

As well as optimal shooting preferences, teams have several questions that need to be answered, and data analysis can help answer all of them in different aspects. Some of them are presented below under related categories.

Performance/Quality of Players:

- Who is the best player in the league?
- Which players are good fits for our team?

- Which players deserve to play more/fewer minutes?
- Which players deserve to win more/less salary?
- Which players have an increase/decrease in their performance over the last few weeks

Performance / Strength of Teams:

- Which team is the strongest in the league?
- Which team has an increase/decrease in performance?

Game Result / Championship Prediction:

- What are the critical factors for winning the game?
- Which team will win the game? What are the probabilities?
- Which team will win the game and with how many points?
- Which team will win the championship? What are the odds?

As can be seen, several questions can be answered, and this study focuses on the game result prediction, which will be explained in, which will be explained in the section 1.1.

1.1 Aim of the Thesis and Motivation

The main aim of this study is modelling and predicting the game result in basketball, finding significant factors which affect the game results and revealing the difference of dynamics in NBA and TBL.

There are three major motivation arguments for this study. Firstly, there is a lack of game result prediction studies that cover important factors such as team strength, home-court advantage, resting, player-based metrics and use advanced data such as the Four Factor. There are several major basic statistics in basketball such as field goal made, points, rebound, assist, steal, block, and turnover. These statistics are basic because they do not need calculation, filtering, or adjustment. However, in the last few years, advanced metrics emerged such as Effective Field Goal Percentage (EFG%), True Shooting (TS%) and Player Efficiency Rating (PER) to evaluate performances. These metrics are developed to better understand the performance of

teams and players since considering only basic statistics might be misleading. For example, points per game only show the scoring potential of a team, however it does not show the efficiency and defense. For that reason, more adjusted metrics such as Effective Field Goal Percentage should be used to understand the team's real strength better.

Most of the studies focus on a specific topic such as the performance of players, home-court advantage or resting without analyzing all the factors together and without explaining the relation between each other. For example, to predict the game results, only basic team statistics are used without considering other factors such as resting or player-based statistics etc. Using original inputs rather than the classical ones shows the impact of expert opinion in basketball. Using advanced and numerous inputs can be helpful to increase the accuracy, which is a primary motivation of the thesis.

Secondly, most of the studies in the basketball field are related to NBA while only few of them are related to European Leagues, and almost zero for TBL. By many experts in basketball, the Turkish Basketball League is considered as a third-best basketball league in the world in terms of quality after the NBA and Spanish Basketball League (ACB). While the NBA teams take advantage of data analytics intensely, teams in other countries do not put emphasis on data analytics. Analysis and prediction of game results using data in different leagues allow us to understand and compare the dynamics of these leagues. Since the number of studies is extremely limited in TBL, the outputs of this thesis might be value-added for the TBL.

Thirdly, in terms of modelling perspective, a comparison of different machine learning algorithms and different input scenarios should be made to understand which models and input sets are best fit for basketball game result prediction. Most of the game result prediction studies only focus on the one specific model, thus this study can be beneficial to show which models and inputs are significant for basketball game result modelling.

The study includes the descriptive analysis of NBA and TBL game results in 2016-2017, 2017-2018 and 2018-2019 regular seasons. Playoffs are not included in this study since the dynamics of playoffs are extremely different from regular season and the number of playoff games is too small compared to regular seasons games. It should be noted that this work is totally based on match statistics which do not consider any extra information such as players played with minor injuries, heated debates between players, the mood of players.

Data preprocessing, numerical explorations, model development and data visualizations are conducted via Microsoft Excel and Python.

1.2 Structure of the Thesis

The structure of this thesis is as follows:

- Chapter 1, which is the introduction part, explains the importance of data analysis in solving some major sports and basketball problems. The current state of the sports analytics is briefly explained, including the technologies, business models and impact of the data analysis on the basketball. Lastly, motivation, aim and the scope of the thesis is presented.
- In Chapter 2, the literature review is presented regarding the related topics including evaluation of team strength, game result prediction, home court advantage, team form, momentum, and comparison of NBA with other leagues.
- In Chapter 3, a descriptive analysis of the significant elements of basketball game dynamics is presented. Significant factors such as team strength, home-court advantage, resting, momentum, individual players are briefly explained and the relation of these factors with the game result is shown.
- In Chapter 4, NBA's and TBL's game result prediction modelling in 2016-17, 2017-18 and 2018-19 regular seasons are presented. The methodology used in benchmark models and machine learning models is clearly explained. Details of

data preparation, variable selection and hyperparameter tuning are given along with the results of the models.

- In Chapter 5, the conclusion of the descriptive and predictive analysis, the model comparison and significant findings are presented. Possible improvements to be made for the future work are suggested.

CHAPTER 2

2 LITERATURE REVIEW

Most of the studies related to basketball analysis focus on the NBA, while only a few of them are associated with EuroLeague or national basketball leagues. This is an expected outcome when NBA is the most popular basketball league which provides easy access to the vast amount of data. Moreover, NBA teams are aware of the power of data analytic solutions, thus studying NBA is mostly favored.

Only few studies focus on the game result prediction. At the same time, most of them analyze specific parts of basketball such as home-court advantage, momentum (commonly known as hot hand phenomenon), evaluation of the value of players, schedule effect etc. Related literature review is given below.

2.1 Evaluation of Team Strength and Game Result Prediction

Elo (1978) developed a ranking system based on a mathematical formula to fairly evaluate and adjust the performances of chess players in a tournament. The fundamental problem in chess tournaments is the extremely high number of participants unlike football or basketball which makes it impossible for every single player to compete with all the participants one by one. Elo system gives the same initial points for each player and then updates their points at the end of each game according to the game result and current point of both players competing. The basic idea is to give more points if a player beats a powerful opponent.

Carlin (1996) used a regression method to calculate win probabilities for games and championships in the NCAA. He predicted the expected point spread to calculate win probability and assumed that point spread is normally distributed around the expected value. He used the difference of rankings of the teams as an input to represent the strength of the teams.

Paul Kvam (2006) combined logistic regression and Markov chain model to predict winning probabilities. They tried to predict winning chance by considering past results of the games between two paired teams. They thought the home-court advantage and point spread in calculating win probabilities. Markov chain model is used to evaluate team rankings and updated after the game results. Logistic regression is used to calculate transition probabilities in the Markov chain model. Their model accuracy was higher than the standard ranking systems such as Sagarin, Massey and Ratings Percentage Index.

Justin Kubatko (2007) published an article that is like a milestone in basketball literature and introduces many new metrics including offensive and defensive ratings, pace adjustment, true shooting percentage, plus-minus and four factors. Four factors are effective field goal percentage, offensive rebound percentage, free throw rate and turnover rate which are the most common metrics in measuring team strength.

Bernard Loeffelholz (2009) used neural network-based models to predict NBA game results. Feed-forward, radial basis, probabilistic and generalized regression neural networks models are used. Team statistics such as points, field goal percentage (FG%), offensive rebounds, blocks, steals are used as inputs. They used only the first 620 games of the regular season to reduce the effect of injuries and transfers. The average accuracy of the neural network-based models is 71.3% which is better than the 68.7% accuracy achieved by betting experts.

Dragan Miljkovic (2010) used various models to predict NBA game results. Team statistics such as field goal made per game, offensive rebounds per game, assist per game etc. are employed. Moreover, standings and momentum attributes such as percent of wins, number of games won at home, number of consecutive wins/losses and number of wins in last ten games are among included metrics. Decision trees, Naïve Bayes classifier, support vector machine and k nearest neighbors are used and it’s found that the best model is Naïve Bayes classifier with 67% accuracy.

Baghal (2012) tested if Four Factor metrics effectively predict winning percentages of teams in the NBA. He used linear regression and structural equation modelling to evaluate the importance of Four Factor metrics. He found that all four-factor metrics are significant for predicting winning percentage. Moreover, he discovered that replacing defensive free throw rate with steal per possession increase the model performance.

Manuela Cattelan (2012) developed a dynamic extension of the Bradley-Terry model for paired comparison to predict outcomes of sports games including time-varying strength. They assumed that team strength follows an exponentially weighted moving average process. They also assume that team strength at home and away are independent. Performance in home games is only related to home game strength and same logic applies to away games as well. They used the Brier score to evaluate a model and found that the dynamic model has a 0.421 Brier score which outperforms the static model with a 0.409 Brier score.

Manner (2016) used multiple linear regression to predict game results using team strength, home-court advantage, back-to-back games as inputs. Team strength is modelled in two different ways. In the first approach team strength is considered as constant during the whole season. In the second approach team strength is dynamic and updated after every game result based on the Gaussian autoregressive process. He tested the model for eight seasons of NBA and found that playing at home court gives [2.2, 2.7] point advantage and playing back-to-back games gives [-1.7, -1.9] point disadvantage.

Fadi Thabtah (2019) predicted the NBA games using feature analysis and machine learning algorithms. They used Naïve Bayes, artificial neural networks, and decision tree models. They found that defensive rebound is the most significant factor affecting NBA games. Other crucial factors are three-point percentage, free throws made and total rebounds. They used multiple regression, correlation feature set and RIPPER algorithm for variable selection. There are 21 variables and seven of them are selected using the Multiple Regression method: home-court (1 if home team, 0 if away team), three points made, three-point attempts, defensive rebound, steal,

turnovers, and personal fouls. Home court, field goal made, field goal percentage, three points percentage, free throws made, defensive rebound and total rebound is selected by using the correlation feature set model. Finally, using the RIPPER algorithm, they selected field goal attempt, field goal percentage, total rebound, three-point percentage, and free throw made. They used accuracy and F1 score metric to evaluate the performances of models. The accuracies of the models were in between %71 and %83. They tried to find the most significant factors for an NBA game, thus they assumed that the parameters were known beforehand.

Ping-Feng Pai (2016) used an ensemble model with SVM and Decision Tree to predict basketball games. They used correlation-based feature selection to find the best feature subsets. This approach considers the correlation between features and class and takes correlation between features in subset into account. Two-point field goal percentage, three-point field goal percentage, free throws, defensive rebounds, total rebounds, steal, and assists are selected as input variables. They used 400 games data which are in between 2008-2010 regular seasons. The model accuracies were in between 71% to 73.5%.

2.1 Home Court Advantage and Schedule

Barry Schwartz (1977) shows the significant effect of home advantage in his paper which delivers a leading approach to the field. It's stated that home advantage is much more effective in indoor sports such as basketball and ice hockey than football and baseball which are considered outdoor sports. It's also mentioned that home court primarily affects offense rather than defense and that the audience effect is more important than resting and being accustomed to home court.

David A. Harville (2015) analyzed college basketball NCAA (National Collegiate Athletic Association) and estimated that the advantage of playing at home is +4.58 ± 0.28 points per game. They stated that the home team has the advantage for three significant factors: audience support, court familiarity and away team is exhausted

by travelling. They also found that team-specific home advantage differences are minimal.

Jones (2007) studied home court advantage considering the in-game analysis. He analyzed the 2002-2003 and 2003-2004 NBA regular seasons and stated that the home team has a +3.59 to +3.89 points advantage. It's found that the effect of home-court advantage is mainly observed in the first quarter of the game, and it becomes more distinct if the home team is losing at the beginning of any quarter.

Oliver Entine (2008) studied home-court advantage in NBA considering the resting and schedule. They believed that home-court advantage had not been appropriately studied until this study since the schedule effect is ignored. There are road trips that contain several away games in a short period in NBA. Road trips may create an imbalanced schedule which makes the difference between the tiredness of teams. It's found that home teams play 12 back-to-back games (no resting day) while away teams play 27 back-to-back games in a single season on average. It's also found that resting one day between games instead of playing back-to-back games creates a significant difference in terms of performance and game results, while there is no significant difference between the effect of 1-day rest and 2-day rest between games in that regard.

Kelly (2009) analyzed if there is a team specific injustice due to back-to-back scheduling games and he found that there is not. Each team faces away team disadvantage homogeneously. He also stated that the scheduling strategy in NBA is based on the minimization of travelling costs for all teams.

Haris Pojskic (2011) analyzed the home-court advantage in EuroLeague and found that in almost every performance metric, home team is better than the away team. The average winning percentage of the home team is 67% in the regular season, while it drops to 58% in playoffs. They developed a model to predict which team is the home team using team statistics, and accuracy of the model was approximately 70%.

Using Pearson's Chi-Square Test, Discriminant Analysis and Binary Logistic Regression, Pedro T. Esteves (2020) revealed that the chance of winning a game amplified significantly when teams have at least one-day rest compared to playing back-to- back games. Moreover, they found that even the risk of injury increases in back-to-back games. They also found that the significant decline is seen in the shooting percentage in back-to-back games.

2.2 Momentum and Form

Jeremy Arkes (2011) found evidence of momentum effect in NBA by evaluating win streaks of teams considering and adjusting home-court advantage, the strength of schedule and team strength. They analyzed three seasons in NBA. They found that every number of wins in the last five games increases the chance of winning the next game by 2% to 4%, regardless of team strength, home-court advantage, and strength of schedule on the last five games. For that reason, they concluded that momentum is a natural effect. Also, they stated that being a home or away team in the next game is not so important when the team catches a momentum.

Thomas T. Byrnes (2016) tested if momentum-based betting strategies gain consistent profit. Twelve regular seasons of NBA are analyzed and used in modelling. They stated that a momentum-based betting strategy generates significant profit. Strategy is betting on a team that had a winning streak or betting on the opposite team if the team had a losing streak. Streak is considered if a team had won or lost four or more games in a row. 4 is selected because it is long enough to secure momentum.

2.3 Comparison of NBA and European Leagues

Radivoj Mandic (2019) analyzed the NBA and EuroLeague statistics between the 2000-2017 seasons. Analytics are performed based on minutes per game to ensure a fair comparison between NBA and EuroLeague. A single match is 48 minutes long in NBA and 40 minutes long in EuroLeague. It's found that pace makes a significant

difference, which is the number of possessions per game. On average, basketball in NBA has a higher pace, which means it is faster. NBA is also better at blocks, assists, defensive rebounds, and free throw rate. Two-point field goal percentage and steals are higher in EuroLeague. It’s shown that the statistics are almost similar in the regular season and playoffs for the EuroLeague however there is a vast difference between NBA statistics for the regular season and playoffs.

CHAPTER 3

3 ANALYSIS OF FACTORS AFFECTING GAME RESULT

Several factors affect the total points scored by the teams and, consequently, the game result in a basketball game. Without a doubt, a major factor is the overall team strength which represents the level of quality of a team in terms of playing basketball. Overall, team strength is predominantly based on the players' talent, and secondly, it depends on the coordination and role sharing between these players concerning strategy of the team coach. The team with a better overall team strength is more likely to win the game. Measuring overall team strength is difficult. Moreover, overall team strength is not enough to explain all the game results. Other factors such as home-court advantage, team form and momentum, resting effect and player-based effects are also important and influential. All these factors are correlated and cannot be assessed separately in modelling. Therefore, the relation among these parameters shall be analyzed in detail.

3.1 Data Description

The data analyzed in the study covers 2016-2017, 2017-2018 and 2018-2019 regular seasons in NBA and TBL. Many statistics and metrics are used to analyze team standings, game results, box scores and player statistics. NBA's official web page (NBA, 2022), TBL's official web page (TBL/BSL, 2021), Basketball Reference (Basketball Reference, 2021) and TBLstat web pages (TBLstat, 2021) are used as the data sources. Microsoft Excel and Python programs are used for data collecting via web-scraping, data pre-processing, analysis, and modelling.

3.2 Overall Team Strength

Overall team strength is the average power of a team regardless of flexible factors such as home-court advantage, momentum, and injuries. Talent of the players and the strategic use of these players determine the overall team strength. Each player contributes to the team's strength in a different way. For example, some players are good at shooting, and some are good at rebounding. To measure overall team strength, win percentage and Four Factor metrics are used, which are introduced by Justin Kubatko (2007).

3.2.1 Winning Percentage

Winning percentage is the ratio of a number of wins to the number of games played in total. It is a straightforward but effective way of measuring team strength. In all basketball leagues, teams are ranked by the number of total wins obtained at the end of a regular season. Since each team plays the same number of full games, it also means that teams are ranked with respect to winning percentage. Using the number of total wins to rank the teams during the regular season might be misleading since the number of games played by teams can differ. For that reason, the winning percentage can be used to estimate the strength of a team roughly. Each team has a different schedule; thus winning percentage may be insufficient to explain the team strength at the beginning of the season. However, it becomes a more determinative factor when the schedule is getting closer to the end of the season.

The winning percentage metric was tested to understand if it is an accurate metric to predict the game results in 2016-17, 2017-18 and 2018-19 regular seasons for both NBA and TBL. For each game, excluding each team's first game in the season, the winning team is predicted by comparing win percentages of both teams until the next game to be played. Accuracies are calculated under four scenarios:

- Scenario 1: All games, home-court advantage is ignored

- Scenario 2: In all games, the home team is awarded an additional 10%-win percentage
- Scenario 3: Second half of the season, home-court advantage is ignored
- Scenario 4: Second half of the season, the home team is awarded an additional 10%-win percentage

In all scenarios, in the case of equality in winning percentage, the home team is assumed to be the winner. Figure 3 shows the accuracies under these four scenarios.

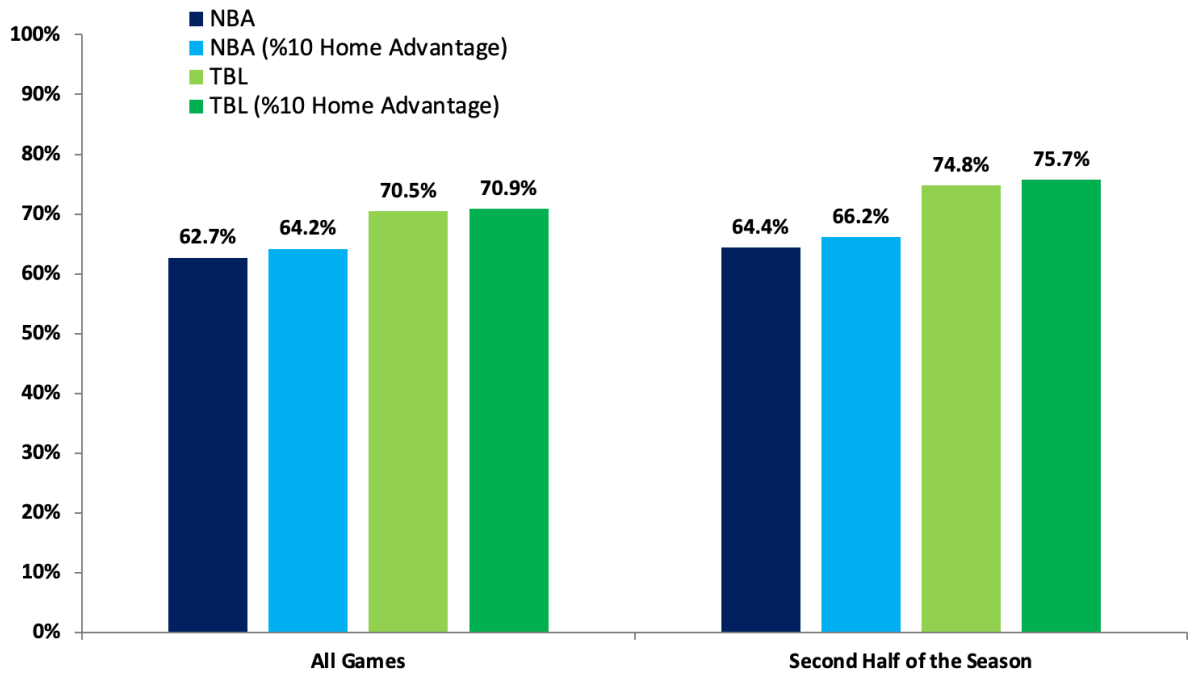


Figure 3: Win prediction accuracies based on different scenarios calculated by win percentages of teams

In Scenario 1, the average accuracy is 62.7% for NBA and 70.5% for TBL. In Scenario 2, accuracies are increased to 64.2% for NBA and 70.9% for TBL compared to Scenario 1, which means considering home-court advantage is necessary. Accuracies in Scenario 3 are 64.4% for NBA and 74.8% for TBL, which are higher than both Scenario 1 and Scenario 2, and it shows that winning percentage becomes a better indicator as the season proceeds. In Scenario 4, accuracies are 66.2% and

75.7% for TBL, which are the highest accuracies among all scenarios. In all the scenarios for both leagues, accuracy is higher than 60%, indicating that winning percentage is a proper metric to predict game results. It is also remarkable that in all scenarios, accuracies are higher for TBL compared to NBA. This is mainly related to the difference in the distribution of power of teams in the leagues and scheduling imbalances which will be mentioned in detail in the following sections. In TBL, the strength difference between the best teams and worst teams is quite high, while in the NBA strengths of the teams are relatively more balanced.

3.2.2 Four Factor

To analyze and model team strength in basketball, most of the studies use the Four Factors metrics, which are introduced by Dean Oliver in his "Four Factors of Basketball Success" article. The main idea behind Four Factor is that team strength can be better understood by decomposing it into four factors which are shooting, offensive rebounding, free-throw points and turnovers. To understand the essence of team strength, these four factors should be considered for both teams. For example, a team can have a high shooting performance, indicating that the team has a good offense. However, if the opposing team has a better shooting performance, then it means that the team has a poor defense against the opposing team. Thus, four factors are considered as eight factors in terms of modelling techniques.

The correlation matrix below shows how these four-factor metrics are related to the total number of wins in a season. Detailed explanation related to the correlation matrix is provided in the related sections.

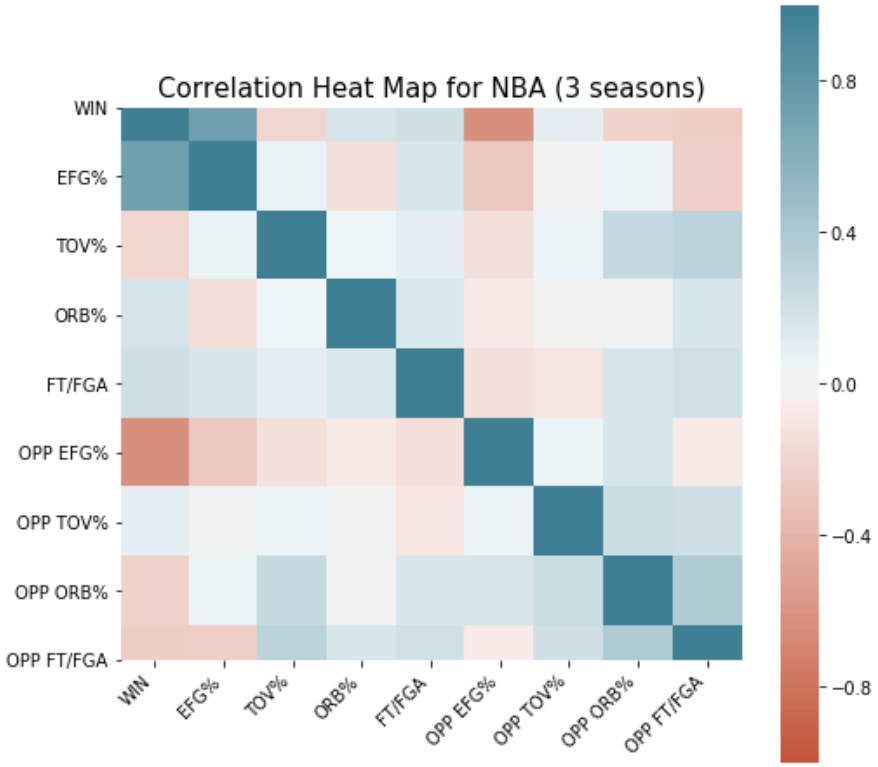


Figure 4: Correlation table for Four Factor metrics (for the team and opposite team) in NBA between 2016-2017 and 2018-2019 seasons

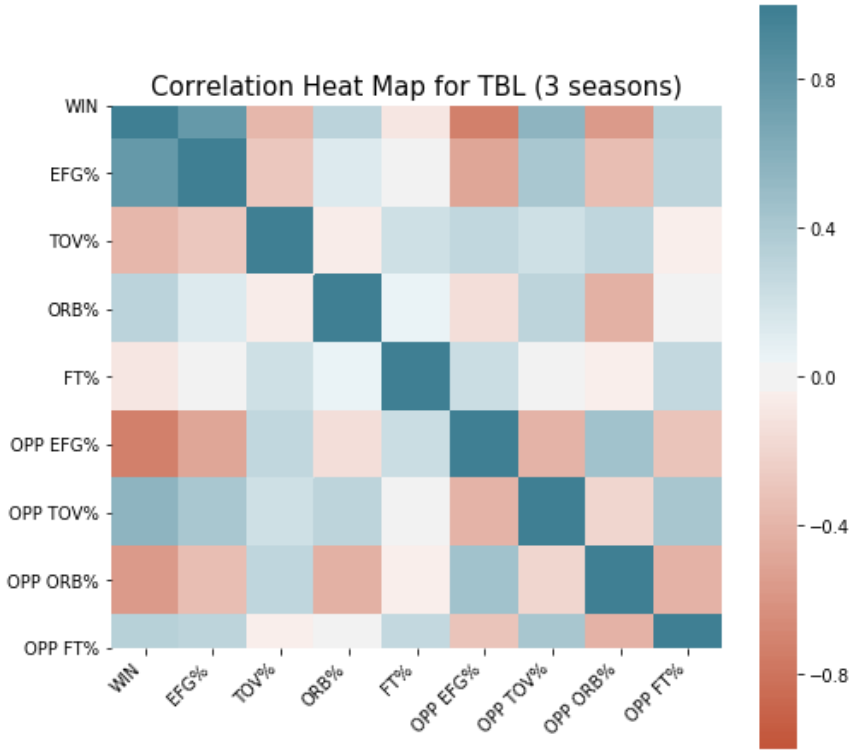


Figure 5: Correlation table for Four Factor metrics (for the team and opposite team) in TBL between 2016-2017 and 2018-2019 seasons

3.2.2.1 Shooting - (EFG%)

Shooting performance is considered the most significant factor in basketball since it determines the total team score. The shooting performance of a team depends on the shooting talents of the individual players and scoring by finding the open players and performing decent passing to end up with a good ball movement. To measure the shooting performance of a team, an effective field goal percentage (EFG%) metric can be used, which is the adjusted and more advanced version of field goal percentage (FG%).

Field goal percentage (FG%) is a primary performance metric for measuring shooting accuracy, and it is the number of successful shots per number of shot attempts.

$$\boldsymbol{FG}\% = \frac{FGM}{FGA}$$

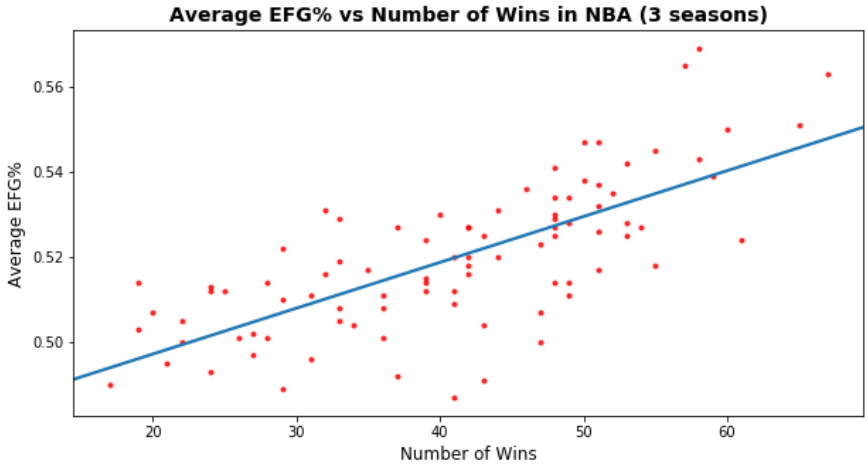
Where FGM is field goal made and FGA is field goal attempt.

Effective field goal percentage is an adjusted metric for field goal percentage (FG%) considering the value of 3 points made.

$$\boldsymbol{EFG}\% = \frac{FGM + \frac{3PM}{2}}{FGA}$$

Where 3PM is 3 points made.

EFG% is a more accurate metric than FG% for measuring shooting performance since 3 points are more valuable than 2 points. EFG% depends on the player shooting talent and ball movement and finding the open player.



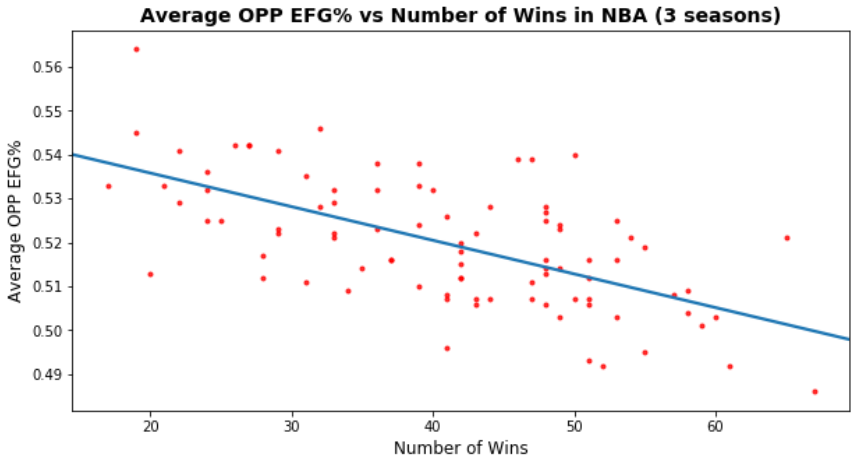
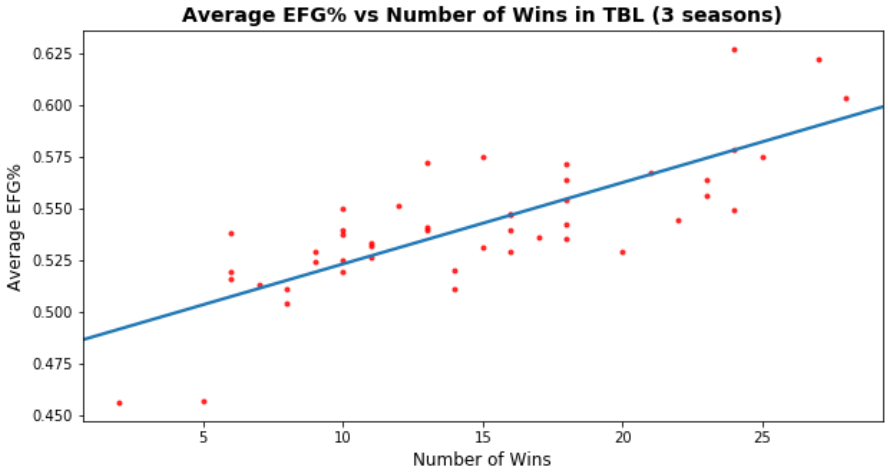


Figure 6: Number of wins relation with an effective field goal percentage in NBA between 2016-2017 and 2018-2019 seasons



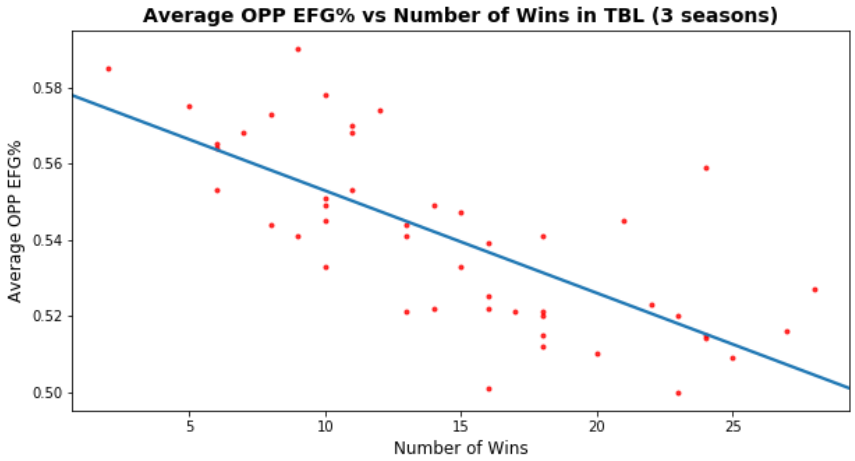


Figure 7: Number of wins relation with an effective field goal percentage in NBA between 2016-2017 and 2018-2019 seasons

In the 2016-2017, 2017-2018 and 2018-2019 NBA seasons, teams with the higher EFG% won in 81.1% of the games. In the same seasons, this percentage is 81.9% for TBL.

Table 1: Effect of EFG% on win percentage in NBA and TBL in 2016-17, 2017-18 and 2018-19 seasons

League	Seasons	Number of Games	Number of Games when Higher EFG% Team Won	Win Percentage (%)
NBA	2016-2017	1230	1013	82.3%
	2017-2018	1230	984	80.0%
	2018-2019	1230	997	81.1%
	Total	3690	2994	**81.1%**
TBL	2016-2017	240	189	78.8%
	2017-2018	240	191	79.6%
	2018-2019	210	185	88.1%
	Total	690	565	**81.9%**

The average EFG% of the winning teams is 55.1% during three seasons in NBA, while this number is 49.3% for losing teams. In the same seasons, the average EFG% for the winning teams is 57.8%, while it is 50.0% for the losing teams in TBL.

For both leagues, the t-test is applied to determine if EFG% is a significant factor affecting game results. The hypothesis is as follows:

H_0: The Winning Team and The Losing Team has the same EFG%

H_A: The Winning Team's EFG% > Losing Team's EFG%

P-values are found to be very close to 0 thus t-test results strongly support that EFG% is extremely important in determining the game results.

3.2.2.2 Rebounding – (DRB% - ORB%)

Rebounding is extremely important since it creates a new opportunity for the team to score.

Offensive rebound percentage (ORB%)

$$\boldsymbol{ORB\%} = \frac{ORB}{Opp\ DRB + ORB}$$

Defensive rebound percentage (DRB%)

$$\boldsymbol{DRB\%} = \frac{DRB}{Opp\ ORB + DRB} = \mathbf{1} - \boldsymbol{OPP\ ORB\%}$$

Table 2: Effect of ORB% on win percentage in NBA and TBL in 2016-17, 2017-18 and 2018-19 seasons

League	Seasons	Number of Games	Number of Games when Higher ORB% Team Won	Win Percentage (%)
NBA	2016-2017	1230	685	55.7%
	2017-2018	1230	683	55.5%
	2018-2019	1230	714	58.0%
	Total	3690	2082	**56.4%**
TBL	2016-2017	240	159	66.3%
	2017-2018	240	142	59.2%
	2018-2019	210	126	60.0%
	Total	690	427	**61.9%**

The average ORB% of the winning teams is 23.9% in 3 seasons of the NBA while this number is 22.0% for the losing teams. In the same seasons of TBL, the average EFG% for the winning teams is 31.1%, while it is 27.1% for the losing teams.

For both leagues, t-tests are applied to determine if offensive rebound percentage is significant to determine the game results. The hypothesis is as follows:

H_0: The Winning Team and Losing Team has the same ORB%

H_A: The Winning Team's ORB% > The Losing Team's ORB%

P-values were found as almost 0, which indicates that ORB% is a significant factor in winning the game.

3.2.2.3 Turnovers - (TOV%)

Turnover means losing control of the ball and as a result, the opposite team takes the position. Most of the time, turnover is worse than missing a shot since there won't be enough time to come back to defense.

Turnover percentage (TOV%) is defined as follows:

$$\boldsymbol{TOV}\% = \frac{TOV}{FGA + 0.44 * FTA + TOV}$$

TOV: *The number of turnovers*

FGA: *The number of field goal attempts*

FTA: *The number of free-throw attempts.*

Table 3: The effect of TOV% on win percentage in NBA and TBL in 2016-17, 2017-18 and 2018-19 seasons

League	Seasons	Number of Games	Number of Games when Higher TOV% Team Won	Win Percentage (%)
NBA	2016-2017	1230	678	55.1%
	2017-2018	1230	670	54.5%
	2018-2019	1230	690	56.1%
	Total	3690	2038	**55.2%**
TBL	2016-2017	240	142	59.2%
	2017-2018	240	146	60.8%
	2018-2019	210	130	61.9%
	Total	690	418	**60.6%**

The average TOV% of the winning teams is 12.3% in 3 seasons of the NBA while it is 12.8% for the losing teams. In the same seasons of TBL, the average TOV% of the winning teams is 14.3%, while it is 15.9% for the losing teams.

For both leagues, t-tests are applied to test if the turnover rate is an important factor in determining the game results. The hypothesis is as follows:

H_0: The winning team and the losing team has the same TOV%

H_A: The winning team's TOV% < The losing team's TOV%

A low p-value indicates that TOV% is a significant factor in winning the game.

3.2.2.4 Free Throws - (FT%)

Free throw is extremely important in basketball since it provides an easy scoring chance without any defensive pressure.

$$\boldsymbol{FT}\% = \frac{FTM}{FGA}$$

Table 4: The effect of FT% on winning percentage in NBA and TBL in 2016-17, 2017-18 and 2018-19 seasons

League	Seasons	Number of Games	Number of Games when Higher FT% Team Won	Win Percentage (%)
NBA	2016-2017	1230	723	58.8%
	2017-2018	1230	699	56.8%
	2018-2019	1230	737	59.9%
	Total	3690	2159	**58.5%**
TBL	2016-2017	240	108	45.0%
	2017-2018	240	125	52.1%
	2018-2019	210	110	52.4%
	Total	690	343	**49.7%**

The average FT% of the winning teams is 21.4% in 3 seasons of the NBA while it is 19.5% for the losing teams. In the same seasons of TBL, the average FT% for the winning teams is 22.9%, while it is 21.5% for the losing teams.

T-tests are applied to test if free throw rate is important to determine the game results. The hypothesis is as follows:

H_0: The Winning Team and The Losing Team has the same FT%

H_A: The Winning Team's FT% > The Losing Team's FT%

For the NBA, t-tests resulted in almost 0 p-value and thus it's strongly supported that FT% is extremely important in winning the game. For TBL, p-value is 0.051, thus FT% is not significant for 5% level.

In conclusion, all four-factor metrics are significantly important for the NBA, while for the TBL, only the free throw rate is not significantly important. Shooting is almost equally significant for NBA and TBL. Rebounding and turnovers are more significant in TBL compared to NBA. Meanwhile, the free throw rate is more significant in the NBA compared to TBL.

Table 5: Percentage of wins when the team has a better Four Factor Metrics in the NBA & TBL in 2016-17, 2017-18 and 2018-19 seasons

	NBA	**TBL**
Shooting (EFG%)	81.1%	81.9%
Rebounding (ORB%)	56.4%	61.9%
Turnovers (TOV%)	55.2%	60.6%
Free Throws (FT%)	58.5%	49.7%

3.3 Home Court Advantage

The home team has a significant advantage over the away team thanks to the support of the audience, and also home team players feel more comfortable playing at their home court since they are used to playing at that court. Fans' slogans, voices and reactions can indirectly affect the game's outcome in different ways. The fans can give a morale boost to the home team, and that can help the home team gain momentum. Moreover, fans may distract the away team and cause them to miss shots, especially in free throws. Finally, fans can affect the decision of referees on behalf of the home team. However, this scenario is scarce. The second major advantage of the home team other than the audience support is habits which makes players of the home team more comfortable than the players of the away team. Away team must travel to home team's city. Thus, their habits such as eating and sleeping are interrupted significantly, while home team players can easily follow their daily routines. Moreover, the home team is accustomed to the court's components such as the surface, the backboard, and the basketball hoop (David A. Harville, 2015).

The home-court advantage effect can be easily seen in both NBA and TBL. Throughout ten seasons between 2009-2010 and 2018-2019, the average winning percentage of home teams was **59.0%** in NBA and **59.4%** in TBL. In both leagues, home teams won approximately 3 out of every 5 games. Home team advantage is so significant that in both NBA and TBL playoffs, the teams with higher rank in the regular season are awarded more home court games than lower rank teams.

Table 6: Home vs away team statistics in NBA and TBL between 2009-2010 & 2018-2019 seasons

League	Season	Total Games	Home Win	Away Win	Home Winning %	Home Average Score	Away Average Win
NBA	2009/2010	1230	731	499	59.4 %	101.8	99.1
	2010/2011	1230	743	487	60.4 %	101.1	98.0

	2011/2012	990[1]	580	410	58.6 %	97.7	94.8
	2012/2013	1229	752	477	61.2 %	99.8	96.5
	2013/2014	1230	714	516	58.0 %	102.3	99.7
	2014/2015	1230	707	523	57.5 %	101.2	98.8
	2015/2016	1230	724	506	58.9 %	104.0	101.3
	2016/2017	1230	718	512	58.4 %	107.2	104.0
	2017/2018	1230	712	518	57.9 %	107.4	105.3
	2018/2019	1230	729	501	59.3 %	112.6	109.8
	10 seasons	12059	7110	4949	**59.0 %**	**103.6**	**100.9**
TBL	2009/2010	240	149	91	62.1%	79.2	76.1
	2010/2011	240	146	94	60.8%	79.2	76.1
	2011/2012	240	128	112	53.3%	79.1	77.3
	2012/2013	240	129	111	53.8%	77.6	76.2
	2013/2014	240	142	98	59.2%	79.3	76.2
	2014/2015	240	139	101	57.9%	79.7	78.0
	2015/2016	240	151	89	62.9%	81.5	78.3
	2016/2017	240	145	95	60.4%	82.1	78.6
	2017/2018	240	151	89	62.9%	81.8	78.6
	2018/2019	210[2]	128	82	61.0 %	80.9	78.4
	10 seasons	2370	1408	962	**59.4 %**	**80.2**	**77.5**

The points scored by the home team is 2.75 points more than the away team on average for both leagues. T-test shows that the difference between the scores of home and away teams is highly significant.

Table 7: T-test results for mean differences of average scores for home and away teams

	Mean Difference	**t-value**	**p-value**
NBA Average Score Home vs Away	2.76	17.03	~0.0
TBL Average Score Home vs Away	2.75	8.24	~0.0

[1] In 2011/2012 NBA regular season, there was a lockout, thus the number of games is less than 1230

[2] There was a 15 team in 2018/2019 season in TBL

The Chi-square test is applied to see if frequencies of winning for home and away teams are different. Test results indicate that number of win frequencies are different in both NBA and TBL for home and away teams.

Table 8: Chi-square test results considering the equal chance of winning (50%) for home and away teams

	Home Team	**Away Team**	**Chi-square**	**p-value**
NBA Number of Wins	7110	4949	387.3	~ 0.0
TBL Number of Wins	1520	1033	92.9	~ 0.0

The average win percentage of home teams in TBL is slightly higher than NBA; however, this is not significant. The P-value of the Z-test for the difference of proportions is 0.59 thus we do not reject the null hypothesis stating the difference of win percentages are different from 0.

The home team has a good advantage in both NBA and TBL however this effect might be different for each team. David A. Harville (2015) analyzed team-specific home team advantage, and the effect is found to be minor. Moreover, B.Jones (2008) also stated that team-specific home advantage is unreliable.

Table 9: NBA Team Specific Home and Away Win Percentages Between 2009-2010 – 2018-2019 seasons

Team	**Home Win %**	**Away Win %**	**Difference %**
ATL	0.62	0.44	0.18
BKN	0.43	0.31	0.12
BOS	0.64	0.49	0.15
CHA	0.53	0.32	0.21
CHI	0.60	0.47	0.13
CLE	0.57	0.40	0.17
DAL	0.60	0.45	0.15
DEN	0.67	0.41	0.26
DET	0.52	0.31	0.21
GS	0.71	0.55	0.16
HOU	0.70	0.52	0.18
IND	0.67	0.42	0.25

LAC	0.67	0.48	0.19
LAL	0.54	0.37	0.17
MEM	0.64	0.43	0.21
MIA	0.66	0.53	0.13
MIL	0.56	0.40	0.16
MIN	0.44	0.28	0.16
NO	0.53	0.35	0.18
NY	0.47	0.34	0.13
OKC	0.73	0.55	0.18
ORL	0.52	0.34	0.18
PHI	0.47	0.32	0.15
PHO	0.47	0.34	0.13
POR	0.68	0.44	0.24
SA	0.82	0.57	0.25
SAC	0.43	0.28	0.15
TOR	0.63	0.46	0.17
UTA	0.63	0.42	0.21
WAS	0.54	0.34	0.20

Table 10: TBL Team Specific Home and Away Win Percentages Between 2009-2010 & 2018-2019 seasons

Team	**Home Win (%)**	**Away Win (%)**	**Difference (%)**
Fenerbahçe	0.91	0.72	0.19
Anadolu Efes	0.86	0.74	0.12
Banvit	0.79	0.60	0.19
Galatasaray	0.78	0.55	0.23
Karşıyaka	0.75	0.44	0.31
Beşiktaş	0.72	0.52	0.20
Darüşşafaka	0.65	0.48	0.17
Tofaş	0.62	0.39	0.24
Gaziantep	0.57	0.34	0.23
Uşakspor	0.53	0.24	0.29
Türk Telekom	0.52	0.34	0.17

(Best ten teams are presented in terms of win percentage since 2 teams are changing in each season in TBL)

Home court advantage is also related to scheduling and the number of resting days. In NBA, teams have a tight and unbalanced schedule compared to TBL. In TBL,

every team plays one game per week, and the minimum number of rest days is 4, except for the teams that participated in European leagues such as EuroLeague or Eurocup. In Table 11, unbalanced resting for home and away teams can be easily seen.

Table 11: Resting days distribution for home and away teams in 2016-2017, 2017-2018 and 2018-2019 seasons of NBA

Day of Rest	Home Team	Away Team
0	12.7%	23.6%
1	65.8%	56.8%
2	16.3%	16.0%
3	3.6%	2.3%
3+	1.7%	1.3%

During the three seasons from 2016-2017 to 2018-2019, away teams played more back-to-back games than home teams. This is the result of the NBA scheduling system, which tries to minimize total travelled distance, thus when teams start a road trip, they play many games in a short period. That is why the schedule of visiting teams is much more stressful. They play the games on consecutive days frequently, making them more tired than their home team opponents (Oliver Entine, 2008). Playing back-to-back games may significantly decrease team performance since the players will be tired due to the lack of resting days. For that reason, home teams benefit from not only home-specific factors but also longer resting time. More details about the resting effect can be found in section 3.4.

3.4 Resting and Back-to-Back Games

Another critical factor affecting the game result is the number of resting days since it changes players' performance directly. In most of the leagues such as TBL, each team plays one game per week and minimum number of days of rest is 4 for both teams. Therefore, unbalanced schedules are not as important a factor in the TBL as in the NBA. Some of the teams in TBL such as Fenerbahçe, Anadolu Efes and Beşiktaş participate in European basketball leagues such as EuroLeague and

Eurocup. These teams generally play more than one game per week. However, even in such cases, no team is required to play back-to-back games in Europe in the last decade. Moreover, teams that participate in the European leagues are the only teams based on league average. That is why it can be assumed that there is no significant effect of resting on game results for TBL.

Table 12 clearly shows the significant effect of asymmetric rest days on NBA teams' winning. When teams play back-to-back games, and opposing teams rest at least one day, the winning percentage for the teams playing back-to-back games is 40.2% in 3 seasons between the 2016-2017 and 2018-2019 seasons. When teams rest precisely one day, and opposing teams rest for at least two days, the winning percentage increases to 50.6%, which is very close to the expected value of 50%. It means when both teams rest at least one day, it does not matter if there is a resting asymmetry. However, playing back-to-back games has a dramatic negative impact on the winning percentage.

Table 12: Relation Between Resting Days of Teams and Winning in NBA between 2016-2017 and 2018-2019 seasons

Resting Days Team 1	**Resting Days Team 2**	**Number of Wins Team 1**	**Number of Wins Team 2**	**Winning Percentage of Team 1**
0	1	292	417	41.2%
0	2	82	134	38.0%
0	3	13	25	34.2%
0	**1, 2 or 3**	**387**	**576**	**40.2%**
1	2	353	342	50.8%
1	3	62	63	49.6%
1	**2 or 3**	**415**	**405**	**50,6%**
2	3	15	20	42.9%

As mentioned in section 3.3, the number of rest days depends on whether a team is the home team or not due to the scheduling system of the NBA. Table 13 represents the resting day effect on winning considering home-court advantage.

Table 13: Resting Days and Winning Relation for Home and Away Teams

Resting Days for Away Team	Resting Days for Home Team	Number of Win Home Team	Number of Win Away Team	Home Team Win Percentage
0	0	108	65	62.4%
0	1	323	178	64.5%
0	2	102	46	68.9%
0	3	21	9	70.0%
1	0	114	94	54.8%
1	1	808	654	55.3%
1	2	187	135	58.1%
1	3	40	28	58.8%
2	0	36	32	52.9%
2	1	218	155	58.4%
2	2	69	43	61.6%
2	3	16	9	64.0%
3	0	4	4	50.0%
3	1	34	23	59.6%
3	2	6	4	60.0%
3	3	5	2	71.4%

Manner (2016) found that playing back-to-back games can decrease the total points scored by 1.85 and the winning probability by 10%. Another study found that average points scored changed by -1.77, -0.13 and +0.32 for resting days zero, one and two, respectively (Oliver Entine, 2008).

Table 14 shows the relation between the number of resting days and average points scored by home and away teams for NBA between 2016-2017 and 2018-2019 seasons. It can be easily seen that the average score of teams playing back-to-back games is lower than teams that rest for at least one day. Interestingly, the average points scored by the teams that rested precisely three days before the game is less than those that rested for two days. However, the sample size of the teams resting for three days is extremely small, thus it can be neglected.

Table 14: Resting Days and Winning Relation for Home and Away Teams

Resting Days	Average Score	Average Score – Home Team	Average Score – Away Team

0	106.1	107.9	105.1
1	107.9	109.0	106.6
2	108.6	109.8	107.4
3	107.9	109.5	105.4
3+	110.4	111.6	109.0

The T-test is applied to understand if playing back-to-back games while the opposing team rested for at least one day significantly decrease the points scored by home and away teams. P-values are 0.0274 and 0.001 for home and away teams, respectively, indicating that playing back-to-back games affects points scored significantly. The T-test is also applied to see if there is a significant effect of resting one day while opposing teams rest more than one day. P-values are greater than 0.05, stating that there is no significant effect of unbalanced resting days when both teams rest at least one day.

Table 15: Average scores by teams with respect to different resting day scenarios

Team 1 Resting Day	**Team 2 Resting Day**	**Team 1 Average Scores**	**Team 2 Average Scores**	**Difference**	**Number of Observation**
0	1	105,6	108,6	-3,0	709
0	2	107,0	110,6	-3,6	216
0	3	106,1	111,4	-5,3	38
0	3+	113,7	107,0	6,7	9
1	2	108,2	107,8	0,4	695
1	3	106,6	107,1	-0,5	125
1	3+	107,5	102,7	4,8	10
2	3	104,6	107,4	-2,8	35
2	3+	112,2	115,8	-3,6	5

Playing back-to-back games is significantly affecting the game results. Moreover, playing back-to-back games depends on scheduling. Thus, whether the teams are treated fairly while scheduling the games is a question of interest. Team-specific scheduling bias concerning back-to-back games is insignificant (Kelly, 2009).

Table 16: Home team winning percentage with respect to the number of consecutive away games of away team

Consecutive Away Games	Number of Games	Home Team Winning %
1	994	57.9%
2	565	58.7%
3	309	60.6%
4	155	58.7%
5	70	56.0%
6	27	57.4%
7	6	50.0%
8	3	75.0%

Table 17: Home team winning percentage with respect to the number of games in last ten days for both teams

Home Team Number of Games in 10 days	Away Team Number of Games in 10 days	Number of Games	Home Team Winning %
4	4	120	55.0%
4	5	266	63.5%
4	6	111	63.0%
5	4	277	60.6%
5	5	1075	59.3%
5	6	542	59.0%
6	4	97	46.4%
6	5	573	55.3%
6	6	292	57.5%

3.5 Team Form and Momentum

Occasionally teams can improve their performance significantly in a short period exceeding their average performance. As a result of increased performance, teams can gain momentum and win many games in a row. Momentum of a team can be triggered by the improvement in the individual performance of one or more players, the comeback of an important player after an injury, the addition of a new player to the team, a new coach, or adaptation of players in team basketball.

Table 18: Streak effect on winning percentage in NBA & TBL 2016-17, 2017-18 and 2018-19 seasons

League	Streak	Game Count	Winning Percentage
NBA	3 or more	1077	60.0%
	5 or more	384	64.3%
	10 or more	58	74.1%
TBL	3 or more	220	69.1%
	5 or more	105	77.1%
	10 or more	27	88.9%

Table 18 shows that when the team has a streak, the percentage of wins increases. It seems momentum and streak effect are more valid in TBL compared to NBA; however, this is mainly related to power imbalances in the league. In TBL, teams such as Anadolu Efes and Fenerbahçe tend to win most of the games every season; however, it is not the case in NBA.

3.5.1 Win Streak Example of Miami Heat in 2016-2017 Season

In approximately one month, the Miami Heat went from 11-30 to 24-30 by winning 13 consecutive games, arguably the most surprising story of the NBA in the 2016-2017 season. In the first half of the regular season, in other words, the first 41 games, the win percentage of the Miami Heat was only 26.8% and in the second half of the season, which is the remaining 41 games, the winning percentage dramatically improved to 73.2%.

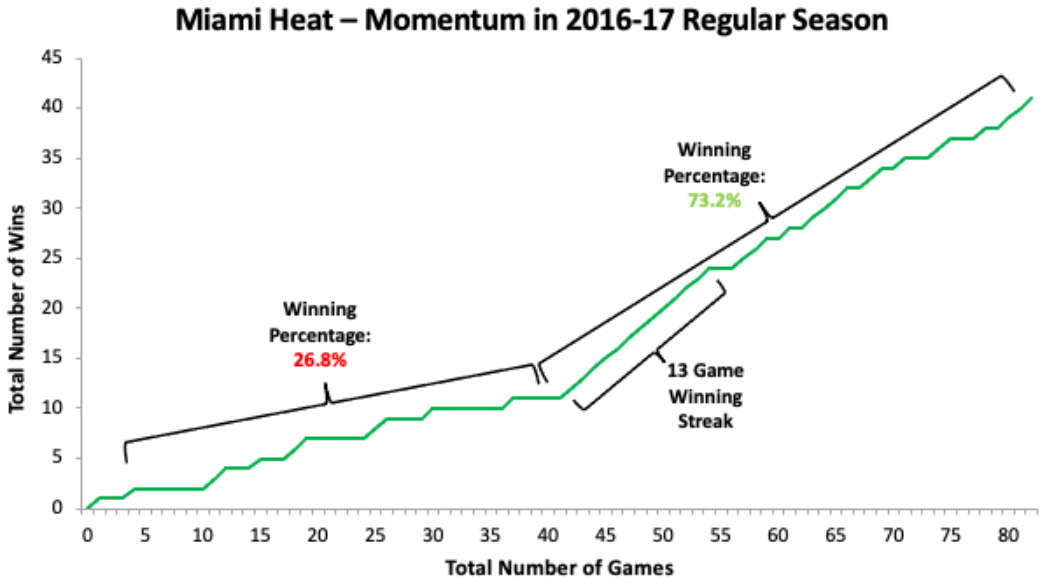


Figure 8: Number of Wins by Total Games Played for Miami Heat in 2016-2017 regular season

The performance of Miami Heat in the 2016-2017 season clearly shows that even below-average teams can take momentum without any changes in the team roster.

3.6 Individual Player Effect

Basketball is a team sport; however, the best player's effect in the team is highly significant, especially for the NBA. The talent of the best player and the strength of the team is somehow related. This relation is not linear and depends on the use of this best player and teammates. Sometimes a team with five decent players can lose to a team with one star player and four average players. This situation is explained in Figure 9.

Research Question: Which team is better?

Team A

Player	Rating
X_1	95
X_2	80
X_3	80
X_4	80
X_5	80
Average	83

Team B

Player	Rating
Y_1	85
Y_2	85
Y_3	85
Y_4	85
Y_5	85
Average	85

Figure 9: Research Question for Relation Between Overall Team Strength and The Strength of Individual Players

For that reason, overall team strength is not enough to explain the game results alone. Some superstars can change the game results significantly even though the overall team strength is lower than the opposing team.

There is no exact definition of "A superstar" in basketball since players have different talents. However, some undisputed superstars in the NBA, such as LeBron James, Stephen Curry, Kawhi Leonard, and Giannis Antetokounmpo. In TBL, some star players such as Shane Larkin, Vasilije Micic, and Jan Vesely played an excellent role for their teams to win games, although the superstar effect is slightly small compared to the NBA.

Superstar players can catch momentum during the game individually, especially in the clutch time, which is the last minutes of a close game, and change the odds in favor of their teams. Moreover, teams build their roster and strategy based on the superstar players. For those reasons, a form of a superstar player is extremely important in winning a game. If a superstar player gets injured and cannot play in the game, the team loses power significantly due to the lack of talented players and strategic problems.

To evaluate the individual performance of a player, basic metrics such as points per game, rebound per game, or assist per game can be used. Moreover, there is an advanced metric called Player Efficiency Rating (PER) to measure the whole performance of a player, considering points, rebounds, assists, steals, and many more stats. ESPN columnist John Hollinger developed the PER metric, and it is the most used metric to evaluate players' performances. PER is adjusted so that the league average equals 15. For that reason, a player with more than 25 PER average can be considered a superstar.

Most of the time, the player with the maximum PER in a team is the team's best player. Figure 10 shows the relation between the maximum PER reached by a player in a team and the number of wins by that team in a regular season.

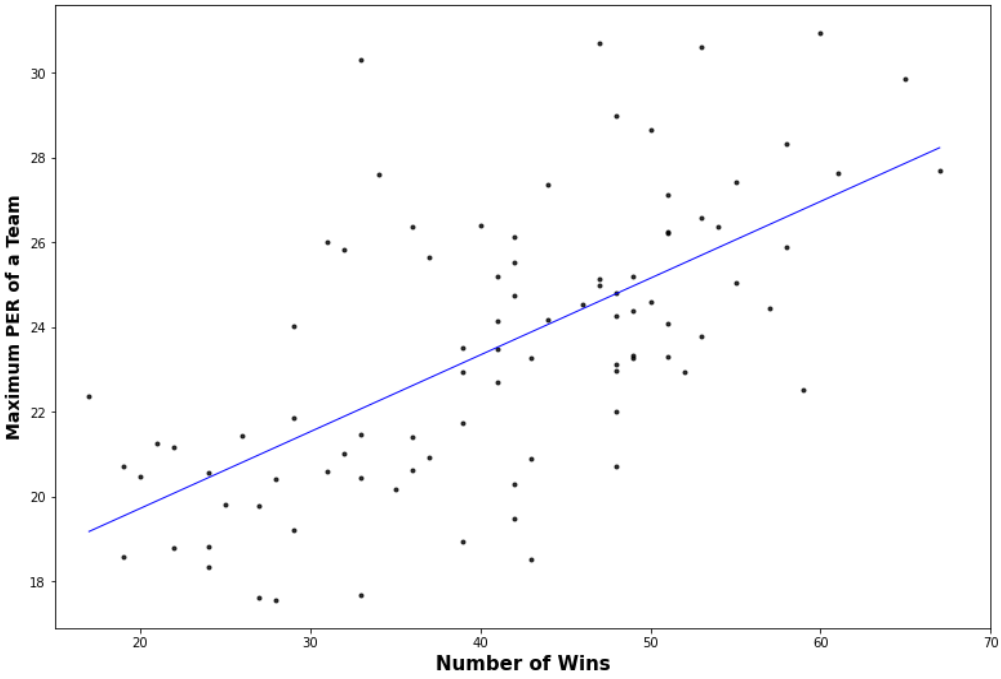


Figure 10: Number of wins by the teams vs maximum PER in teams in 2016-2017, 2017-2018 and 2018-2019 NBA regular seasons

The correlation between maximum PER of a team and the number of wins is **0.64.** This is relatively high when it's considered that basketball is a team game with five players. The maximum PER in 3 seasons of NBA is 30.95, achieved by Giannis Antetokounmpo from Milwaukee Bucks in the 2018/19 season. Bucks won 60 games

and finished the regular season as NBA leader. Table 19 shows the top 10 highest PER reached by players in 2016-2017, 2017-2018, and 2018-2019 NBA seasons and the ranking of these players' teams.

Table 19: Maximum PER of players and team rankings in NBA between 2016-17 and 2018-19 seasons

Player	Player Rank	PER	Season	Team	Number of Wins	Team Ranking
Giannis Antetokounmpo	1	30.95	2018-2019	Milwaukee Bucks	60	1
Russell Westbrook	2	30.70	2016-2017	Oklahoma City Thunder	47	10
James Harden	3	30.62	2018-2019	Houston Rockets	53	5
Anthony Davis	4	30.32	2018-2019	New Orleans Pelicans	33	22
James Harden	5	29.87	2017-2018	Houston Rockets	65	1
Anthony Davis	6	28.98	2017-2018	New Orleans Pelicans	48	8
LeBron James	7	28.65	2017-2018	Cleveland Cavaliers	50	6
Stephen Curry	8	28.32	2017-2018	Golden State Warriors	58	3
Kevin Durant	9	27.68	2016-2017	Golden State Warriors	67	1
Kawhi Leonard	10	27.62	2016-2017	San Antonio Spurs	61	2

There are 30 teams in the NBA and most of the teams which hold one of the best players in the league finished the regular season in the top 5. There are some exceptions, such as New Orleans Pelicans, who finished the regular season as 22nd even they had Anthony Davis in the 2018-2019 season. The team underperformed because Anthony Davis missed 26 games due to injuries in that season.

Points per game is also an effective metric to evaluate individual performances and distinguishing superstars. Table 20 shows the best scorers in the 2016-2017, 2017-2018, and 2018-2019 NBA seasons and the ranking of these player's teams.

Table 20: Maximum points scored per game by the players and the ranking of their teams - NBA

Player	Player Rank	Points/ Game	Season	Team	Number of Wins	Team Ranking
James Harden	1	36.1	2018-2019	Houston Rockets	53	5
Russell Westbrook	2	31.6	2016-2017	Oklahoma City Thunder	47	10
James Harden	3	30.4	2017-2018	Houston Rockets	65	1
James Harden	4	29.1	2016-2017	Houston Rockets	55	3
Isaiah Thomas	5	28.9	2016-2017	Boston Celtics	53	4
Anthony Davis	6	28.1	2017-2018	New Orleans Pelicans	48	8
Paul George	7	28.0	2018-2019	Oklahoma City Thunder	49	9
Anthony Davis	8	28.0	2016-2017	New Orleans Pelicans	34	21
Giannis Antetokounmpo	9	27.7	2018-2019	Milwaukee Bucks	60	1
LeBron James	10	27.5	2017-2018	Cleveland Cavaliers	50	6

The correlation coefficient between maximum points per game in a team and the number of wins is 0.49 for the NBA while it is -0.43 for the TBL, which shows the significant difference in the use of star players and building game strategy in NBA and TBL. Figure 11 shows the negative relation between the highest scorer in a team and team performance.

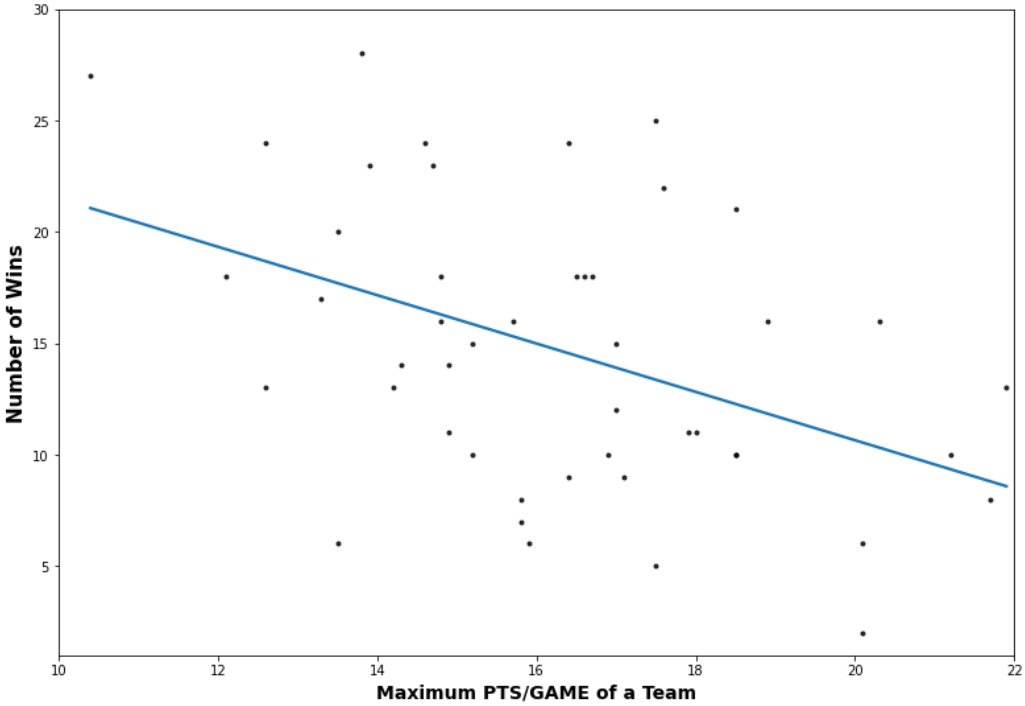


Figure 11: Maximum points per game and number of wins by teams in 2016-2017, 2017-2018 and 2018-2019 TBL regular seasons

Table 21 shows the top 10 scorers in 3 seasons, namely 2016-17, 2017-18, and 2018-19 for the TBL and their teams' performances. There are 16 teams in 2016-17 and 2017-18 seasons and 15 teams in the 2018-19 season in TBL. Rankings of the top scorers' teams are pretty low, which contradicts the situation in the NBA.

Table 21: Maximum points per game by the players and ranking of their teams - TBL

Player	Player Rank	PTS	Season	Team	Number of Wins	Team Ranking
Manny Harris	1	21.9	2018-2019	Bahçeşehir	13	9
Kenneth Hayes	2	21.7	2016-2017	Büyükçekmece	8	12
Ricky Ledo	3	21.2	2017-2018	YeşilGiresun	10	11
Sylven Landesberg	4	20.3	2016-2017	Türk Telekom	16	7
Terrell Holloway	5	20.1	2016-2017	İstanbul Belediyesi	6	13

Jerod Shorter	6	20.1	2017-2018	TED Koleji	2	16
Davon Jefferson	7	18.9	2018-2019	Gaziantep Basketbol	16	7
Caleb Green	8	18.5	2016-2017	Trabzonspor	10	13
Jordan Theodore	9	18.5	2018-2019	Banvit	21	5
Erving Walker	10	18.5	2017-2018	Büyükçekmece	10	12

It is essential to understand that the negative correlation between the points scored by the top scorers and the number of wins by their teams in TBL does not indicate that the top scorers negatively impact their teams. They are still the best players who make the most significant contribution to their teams. It shows that teams should build strategies to allow each player to share the score if they want to be successful. In NBA, teams that build their team strategy based on top scorers tend to win more games. However, in TBL, high-quality teams make their team strategy to allow the contribution of any player rather than focusing on top scorers. The main reason for the difference in the approach of these two leagues is that superstars in NBA are one-of-a-kind players who can change the game dynamics ultimately. In TBL, even if you have an above-average player, he cannot carry the team alone.

3.6.1 Trades & Transactions

If a player transfers to another team, it can affect both the old and the new team for good or bad, depending on the team's abilities and role. Due to the nontrivial relation between talent and the use of players with the power of a team, sometimes the effect of trades and transactions cannot be foreseen beforehand. However, the effect of trades and transactions can be easily seen for the star players. Figure 12 shows the importance of LeBron James, who is arguably the best player in the NBA, for Cleveland Cavaliers.

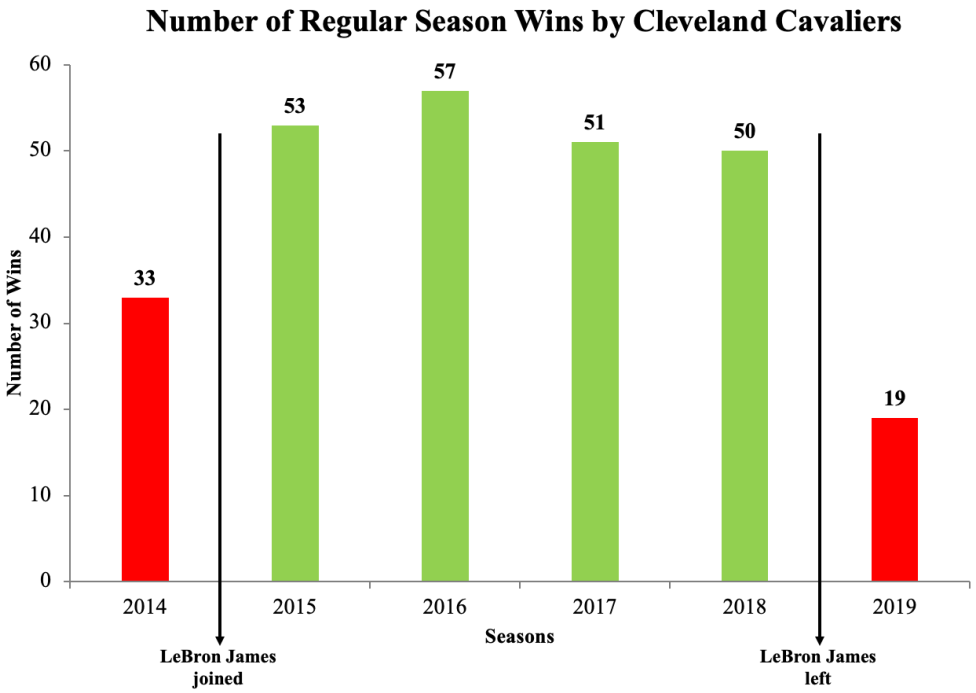


Figure 12: Effect of LeBron James on Cleveland Cavaliers between 2013-2014 and 2018-2019 NBA regular season

In the 2013-2014 season, Cleveland Cavaliers won 33 games out of 82 and ranked 22nd within 30 teams. After LeBron James started to play for Cleveland, the number of wins increased by almost 60% in the 2014-2015 season compared to the previous year. The team ranked 7th with a 53 regular-season victory. The team won 57, 51 and 50 games in the following seasons and ranked 3rd, 5^{th,} and 6^{th,} respectively. Cleveland won the NBA championship in the 2015-2016 season. After James left, Cleveland was able to win only 19 games and ranked 28th. There are some other factors that affect the performance of Cleveland. However, dramatic changes shown by the statistics imply that a single player can make a significant difference for a team.

3.6.2 Injury

Injury is frightening for the players and teams. It is one of the most significant chance factors in sports and unfortunately, it dramatically affects the whole season. Basketball has become a more physical game in the last few years. Players developed new contact moves and used their bodies to gain an advantage. For example, they

fight for the position to get a rebound, draw contact in the air while shooting the ball, and use their forearms and elbows to counteract defenders, and they all lead to injuries (Mark C. Drakos, 2010).

The absence of a star player due to injury decreases the chance of winning dramatically. Figure 13 shows the difference in winning percentages when star players played and did not play due to injury in 2016-2017, 2017-2018, and 2018-2019 NBA regular seasons.

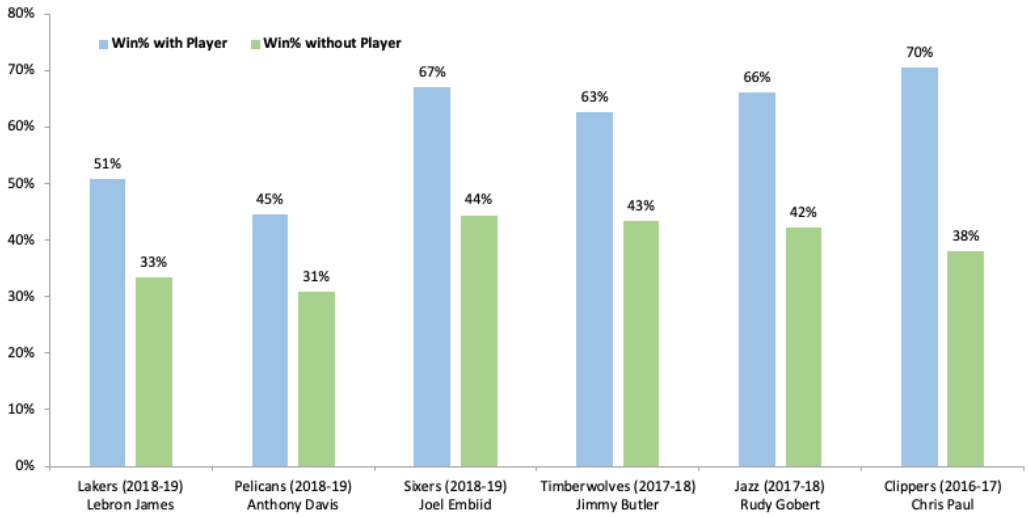


Figure 13: Effect of injury and resting of some notable players in NBA

3.6.3 The Momentum of a Player

Most of the time, the momentum of a team is triggered by the improvement in one player's performance. It is usual for players to over-or under-perform during a long season based on physical and mental conditions. To understand the effect of momentum of players on team performance, some players are selected, and their performances are analyzed. Each month, the best players in conferences (west and east) are awarded the player of the month award in the NBA during the regular season. In 2016-17, 2017-18, and 2018-19 regular seasons, 33 players of the month

award were given in total. The monthly winning percentages of the teams are calculated and compared. It seems that the winning percentages of the teams are directly related to whether a player from that team is selected as the player of the month or not. The average win percentage is 9.5% higher for the months in which a player in the team receives a player of the month award compared to the months with no award. For example, in 2016-17 NBA regular season, Damian Lillard is one of the best examples of how an individual performance can make the team gain momentum. Lillard won the player of the month award in March that season. The winning percentage of Lillard's team, Portland Trail Blazers, was 81.3% (13-3) in March and 42.4% (28-38) in other months. In March, points per game of Lillard was 28.8, which is 2.3 points higher than other months in the 2016-17 season.

3.7 Minute Sharing & Entropy

While the use of talented, super players is extremely important to winning the game, minute sharing between players is also critical because it affects players' tiredness and motivation. Players cannot play the whole game indefatigably thus players should be used efficiently. Moreover, if players play for extreme time, they do not only get tired, but also this may cause them to get injured. Some teams prefer to have a balanced minute sharing, while some prefer to limit the minutes of bench players as much as possible. Both strategies have some advantages and disadvantages. Teams with many players who got reasonably enough time on the court may lack quality on the game. On the other hand, teams that only focusing on the star players may face a severe problem in case of injury or tiredness of a star player.

To measure the minute sharing profile in a team, the entropy formula is used in some studies, including (Radivoj Mandic, 2019) and (Ge Cheng, 2016). Entropy is the metric for the amount of information in a variable developed by Claude Shannon. The formula for measuring the minute sharing is shown below:

$$Entropy(team) = -\sum_{i=1}^{n} p_i * \log(p_i)$$

While i is the player, n is the number of players in a team and, p_i is calculated as follows:

$$p_i = \frac{minutes\ of\ player\ i}{total\ minutes}$$

When the entropy is high, it means minute sharing is balanced. Mandic found that entropy has been increasing in the NBA consistently in recent years, which shows that teams are giving more chances to bench players than in the past.

To understand the effect of minute sharing on team success, correlation coefficients are calculated using maximum minutes per game among all players in a team, entropy, and number of total wins for each team. The correlation coefficients show that the effect of minute sharing on winning is completely different in NBA and TBL. Correlation coefficients are shown in Table 22.

Table 22: Correlation Coefficients for Minute Sharing Analysis

	Correlation Coefficients with Number of Total Wins	
Minute Sharing Variable	**NBA**	**TBL**
Maximum Minutes per Game	0.37	-0.71
Entropy	-0.11	+0.64

There is a significant relation between maximum minutes per game and number of wins, however the relation is positive in NBA with 0.37, while it is negative and stronger in terms of magnitude in TBL with -0.71. Again, this result shows the difference in star player usage in TBL and NBA, same as the maximum points per game. Entropy is inversely proportional with maximum minutes per game. It does not seem to be a proper variable for explaining the game result in NBA since the correspondence coefficient is only -0.11, it might be a decent variable for TBL.

CHAPTER 4

4 MODELLING THE GAME RESULT PREDICTION

In the Chapter 3, possible factors that affect the basketball game results are analyzed and some findings are presented for NBA and TBL. With the help of these findings, machine learning based prediction models are developed to see if these metrics are effective indicators to explain basketball game results. Models are used to predict the game results in the NBA and TBL for 2016-2017, 2017-2018, and 2018-2019 regular seasons.

4.1 Modeling Approach and Assumptions

Predicting the winning team in a game in which two teams are facing with each other is a binary classification problem with two different classes. Several machine learning models are developed using the related inputs, which are generated based on the findings in the descriptive analytics. A summary of the modeling approach can be found in Figure 14.

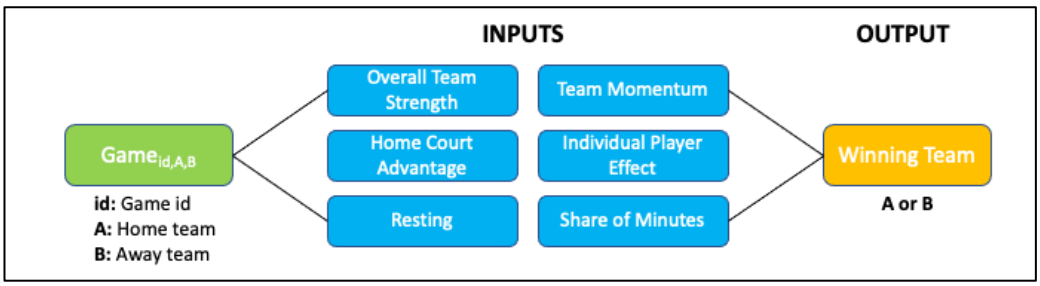


Figure 14: Summary of modeling methodology

The output of the classification model is constituted as follows:

$$Outcome\ y_n = \begin{cases} 1\ if\ home\ team\ wins\ the\ game\ n \\ 0\ if\ home\ team\ loses\ the\ game\ n \end{cases}$$

There are six major input categories: overall team strength, home court advantage, team momentum, resting, team momentum and share of minutes. Under these six major categories, several input variables are developed and used in the models.

The first 70% of the regular season games are selected as a train data set, and remaining games are used for the test. It is assumed that the seasons are independent of each other since there are significant changes in the rosters of teams between two seasons, and as a result strength of teams can change dramatically.

4.1.1 Input Scenarios

Three different input scenarios are used to compare model performances, find the best modelling approach, and see if using new metrics is helpful to increase accuracy. Scenarios are explained below:

- **Input Scenario 1:** Input set, which includes only classical and basic team statistics variables such as points per game and rebound per game. The input set does not have any advanced team statistics such as Four Factor or player-related statistics.
- **Input Scenario 2:** All created advanced team statistics and player-related statistics such as Four Factor, resting days, consecutive number of wins, points per game for a maximum scorer on the team are used in this scenario.
- **Input Scenario 3:** Only selected variables from Input Scenario 2 are used. Approximately half of all variables are chosen as input set; 25 for NBA and 20 for TBL. The feature selection method which will be explained in section 4.1.3 is used. The main aim of this scenario is to understand the most significant features for the game result prediction.

4.1.2 Set of Input Variables

To create an input set for modeling, some variables are selected from existing data, and new variables are generated based on the results from the literature review and results from Chapter 3. All data preparation processes are performed using Microsoft

Excel and Python. For NBA, some of the advanced variables are used already, such as Four Factor metrics however most of the variables are created for the first time. For TBL, almost every variable except in Input Scenario 1 is created for the first time.

Table 23: Input Set for Scenario 1

No	INPUT VARIABLES	NBA	TBL
1	Home Pts/Game	X	X
2	Away Pts/Game	X	X
3	Home Reb/Game	X	X
4	Away Reb/Game	X	X
5	Home Ast/Game	X	X
6	Away Ast/Game	X	X
7	Home Stl/Game	X	X
8	Away Stl/Game	X	X
9	Home Blk/Game	X	
10	Away Blk/Game	X	
11	Home To/Game	X	X
12	Away To/Game	X	X

Table 24: Input Set for Scenario 2 & 3

No	INPUT VARIABLE	NBA	TBL
1	Home Win%	X	X
2	Away Win%	X	X
3	Home EFG%	X	X
4	Away EFG%	X	X
5	Home OPP EFG%	X	X
6	Away OPP EFG%	X	X
7	Home ORB%	X	X
8	Away ORB%	X	X
9	Home OPP ORB%	X	X
10	Away OPP ORB%	X	X
11	Home FT%	X	X
12	Away FT%	X	X
13	Home OPP FT%	X	X
14	Away OPP FT%	X	X
15	Home TOV%	X	X
16	Away TOV%	X	X
17	Home OPP TOV%	X	X
18	Away OPP TOV%	X	X

19	Home Win Streak	X	X
20	Away Win Streak	X	X
21	Home Win in Last 10 Game	X	
22	Away Win in Last 10 Game	X	
23	Home Win in Last 5 Game		X
24	Away Win in Last 5 Game		X
25	Home - Rest Days	X	
26	Away - Rest Days	X	
27	Consecutive Home Games	X	
28	Consecutive Away Games	X	
29	Home - Games in 10 Days	X	
30	Away – Games in 10 Days	X	
31	Home – Away Games in 10 Days	X	
32	Away – Away Games in 10 Days	X	
33	Home – Entropy	X	X
34	Away - Entropy	X	X
35	Home – Max Minutes	X	X
36	Away – Max Minutes	X	X
37	Home – Max Minutes Play	X	X
38	Away – Max Minutes Play	X	X
39	Home – Max Pts/Game	X	X
40	Away – Max Pts/Game	X	X
41	Home – Max Pts/Game Play	X	X
42	Away – Max Pts/Game Play	X	X
43	Home – 2nd Max Pts/Game	X	X
44	Away – 2nd Max Pts/Game	X	X
45	Home – 2nd Max Pts/Game Play	X	X
46	Away – 2nd Max Pts/Game Play	X	X
47	Home – Max PER/Game	X	
48	Away – Max PER/Game	X	
49	Home – 2nd Max PER/Game	X	
50	Away – 2nd Max PER/Game	X	
51	Home – Max Pts/Game Last 3 Games	X	X
52	Away – Max Pts/Game Last 3 Games	X	X
53	Home – Max Pts/Game Last 3 Games Play	X	X
54	Away – Max Pts/Game Last 3 Games Play	X	X

Table 24 shows all potential variables for Input Scenario 3. The selected variables vary for all seasons and models with respect to the feature selection method.

It is also important to note that some of the variables are correlated. However, most of the modern machine learning algorithms do not suffer from multicollinearity

(2013). For that reason all of the inputs are used together in Input Scenario 2, assuming model performances do not suffer from the relation between independent variables.

4.1.3 Input Selection for Input Scenario 3

In general, the feature selection process is performed to eliminate redundant variables or multicollinearity problems. Moreover, it can be used to determine the most important variables for understanding cause-effect relationship. Since each variable is selected carefully according to descriptive analysis and literature in this study, feature selection is not used to reduce the number of variables and increase accuracy. The main aim of feature selection is to determine most significant factors.

For feature engineering, Sequential Forward Selection (SFS) method is used to determine the most important factors and input sets for scenario 3. Sequential Forward Selection is a feature selection method that selects the most crucial **k** variables from the input set where **k** is an arbitrary number. The basic idea is to add one feature at a time based on accuracy performance until the predetermined size **k** is reached. K value is chosen as 25 for NBA and 20 for TBL, which are approximately half of the input variables. The steps of the SFS algorithm are given below:

Input set: $X_1, X_2, \ldots X_n$

Step 1: Define desired size k (where $k < n$)

Step 2: Initialize iteration $i = 1$, $j = 1$ and selected input set $\mathbf{X} = \{\}$

Step 3: Add X_j to X if it is not already included in X and called it Xnew

Step 4: Develop a model with Xnew and calculate accuracy

Step 5: $j = j+1$ and go to Step 3, if $j = n$ go to Step 6

Step 6: Determine the X_j which gives the best accuracy in Step 4

Step 7: Add selected Xj to X

Step 8: $i = i+1$

Step 9: Stop when $i = k$

Each model and seasons have different input selection results under SFS, which will be mentioned in the results and conclusion parts.

4.1.4 Hyperparameter Tuning and Cross-Validation

Hyperparameters are one of the most important factors determining machine learning algorithms' performances. They must be appropriately set so that the model fits training data well and predict the test data with high accuracy. Hyperparameter tuning is a process that should be performed carefully to eliminate overfitting or underfitting.

There are two standard methods for hyperparameter tunning in machine learning algorithms: Grid Search and Random Search. In this study, the Random Search method is used because it is faster and more efficient than Grid Search. The basic logic behind methods and differences are presented in Figure 15.

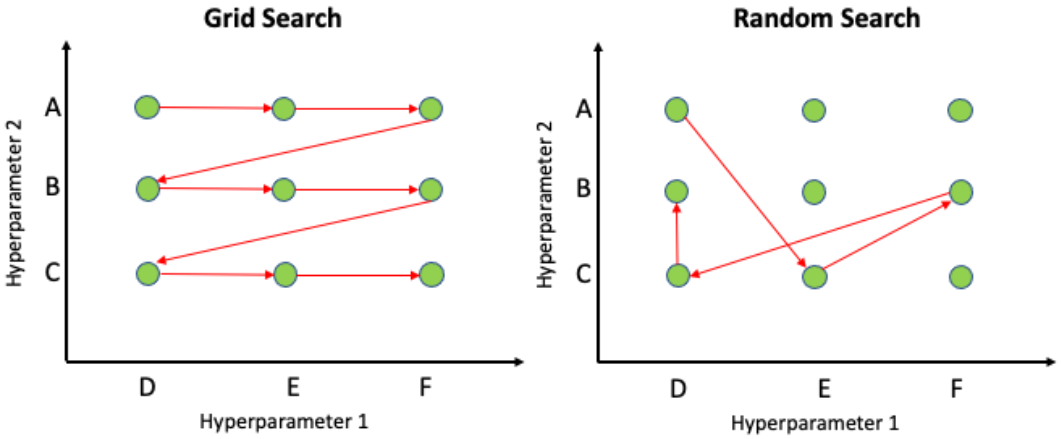


Figure 15: Logic of Grid Search and Random Search

During the hyperparameter tuning process, validation is critical in machine learning modeling to avoid overfitting. Validation should be performed carefully during hyperparameter tuning otherwise models can easily overfit since several scenarios are tried. K-fold Cross Validation is the most popular method for validation of the model. Steps in the K-fold CV method are as follows; split training data into K

equally sized pieces, fit the model using the K-1 parts as train data, predict and calculate the model accuracy for the remaining part, and do the same for K-times iteratively, calculate the average accuracy. In this study, 10-fold CV is used.

For this purpose RandomizedSearchCV() function from the Sklearn package in Python is used. For each machine learning model, some of the most significant hyperparameters are selected to tune, and tuned hyperparameters are explained in section 4.2.

4.1.5 Performance Evaluation Criteria

In general, the confusion matrix is calculated, and accuracy, precision, recall, and F1 score are used for classification problems. F1 score is mostly used for problems with unbalanced data where observation of one category is relatively rare compared to other categories such as detecting rare illness. However, for basketball games, this is not the case. Home and away team wins are balanced in the data, and predicting the home win or away win is equally important. That is why accuracy is the best metric to evaluate the performance of models. Calculation of accuracy can be seen in Figure 16.

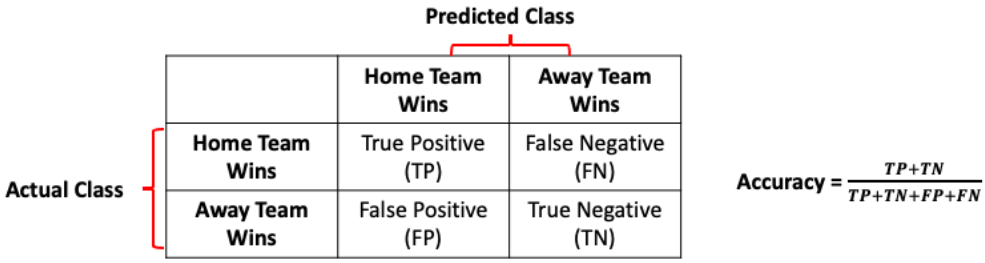


Figure 16: Accuracy Calculation for Classification Problems

4.2 Classification Models and Results

Machine Learning algorithms can be classified under three major categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning has two types: regression and classification. Classification models are used

to predict when the outputs are categorical such as regular e-mail (0) or spam e-mail (1).

In Figure 17 the machine learning categories and some popular algorithms are shown.

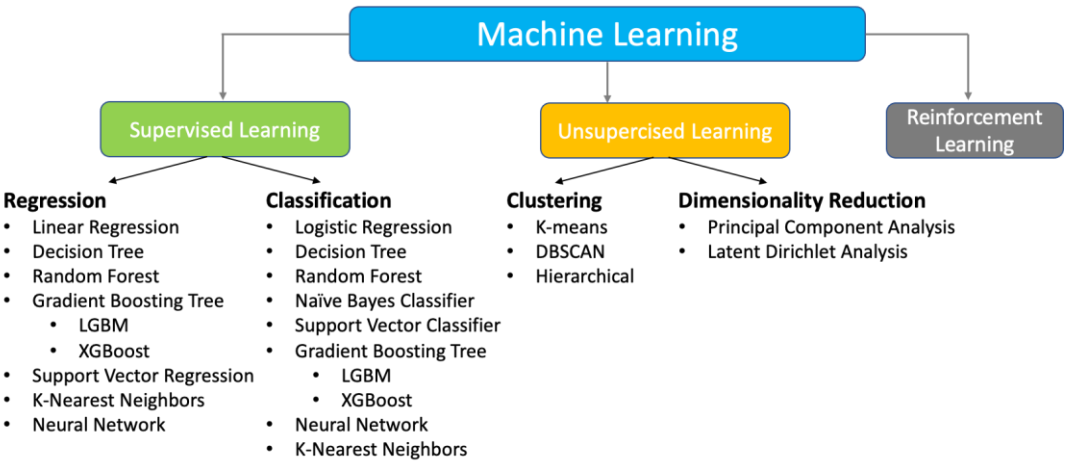


Figure 17: Popular Machine Learning models with related learning problems

In this study Logistic Regression, Support Vector Classifier, Decision Tree, Random Forest, Naïve Bayes Classifier, K-Nearest Neighbors, LGBM, XGBoost and Neural Network models are used. Moreover, the ELO Rating model is developed to test if it can compete with other models in terms of accuracy.

4.2.1 ELO Rating Model with Home Court Advantage

ELO rating is a proper way to measure team strength in the middle of the season with an incomplete schedule for each team since it considers the strength of the schedule. The basic logic behind the ELO rating is giving a same initial score for each team and updating this score based on the result of the game, a score of the team, and score of the opposite team. The mathematical expressions and usage of ELO Rating can be found below.

The commonly used initial score is 1500, thus at the start of the league or tournament, each team has given a 1500 ELO Rating. To update these ratings, firstly expected win probability is calculated based on the teams' ratings based on the formula below.

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

E_A: Expected win probability of team A

R_A: Rating of team A before the game

R_B: Rating of team B before the game

Moreover, if there is a home-court advantage in the league, unlike chess, the rating system can be modified by adding a fixed rating score to the home team. For example, FiveThirtyEight, the sport web site that used ELO Rating to predict win probabilities of games and championship in the NBA, gives 100 additional rating points to the home team (Silver & Fischer-Baum, 2015). In this case, the rating of the home team is increased by 100 points, and the expected win probability is calculated accordingly. After the game is ended, ratings are updated based on the predicted win probability, calculated before the game and game result, using the formula below.

$$R_{A,new} = R_A + K * (S_A - E_A)$$

$R_{A,new}$: Rating of team A after the game

R_A: Rating of team A before the game

S_A: 1 if team A wins, 0 if loses

E_A: Expected win probability of team A

K: Maximum adjustment per game

K value is usually taken as 16 or 32 in the related works. In the model, K is taken as 16. Moreover, to find a proper value of additional home court scores, many trials are made. Expected win probabilities are used to predict the game result. In Figure 18,

the performance of the ELO Rating model and the relation of this performance with different levels of additional home court rating for TBL and NBA are given.

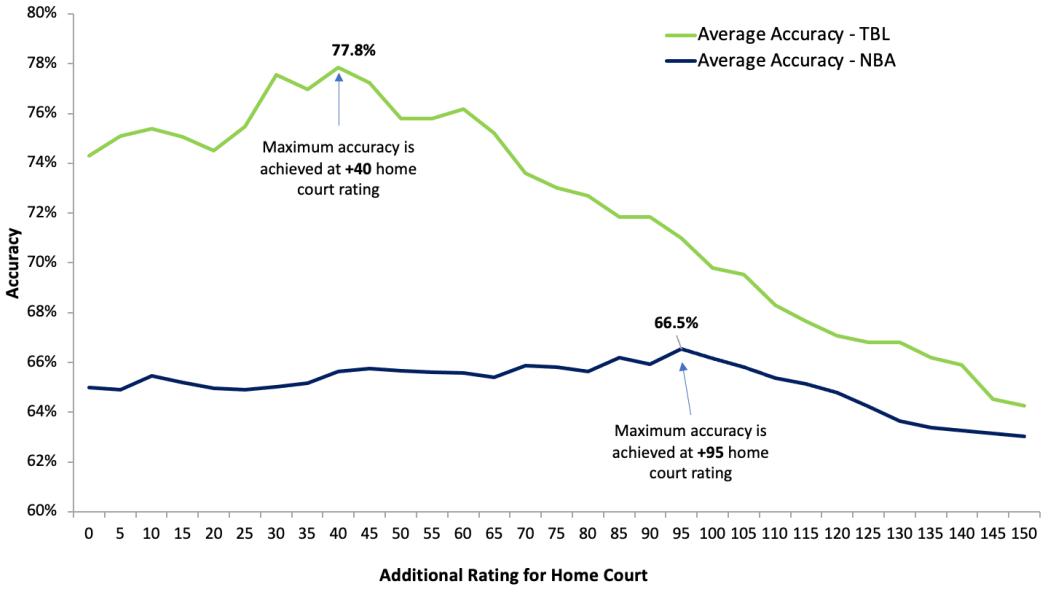


Figure 18: Average ELO Rating model performance for different values for home-court advantage in NBA and TBL in 2016-17, 2017-18 and 2018-19

Results of the ELO rating model show that general prediction accuracy in TBL is significantly higher than in NBA. The main reason for this significant difference is that the Elo rating model is based on calculating the strength of teams and assuming the strongest team will win considering home-court advantage. In TBL, the strengths of teams are more unbalanced compared to NBA thus it is considerably easier to predict the game result using the ELO rating model. Moreover, Figure 18 shows how accuracy is changing by fixed rating score given to the home team. For TBL, the optimum value for home-court advantage rating is found to be +40 while it is +95 for NBA. In TBL, after +60 home advantage ratings, accuracy starts to decline dramatically due to the unbalanced strengths of teams in the league.

4.2.2 Logistic Regression

Logistic regression is one of the most popular and most straightforward models in machine learning used for classification problems. The logic behind the model is very similar to linear regression. The difference is that in logistic regression, the output of the model must converge to predefined classes such as 0 or 1. In our problem, 1 represents the home team win, while 0 represents the away team win. For this kind of convergence, logistic regression uses a logistic function known as an S-shaped curve. The graphical expression of logistic regression with one independent variable is presented in Figure 19 as an example for this problem.

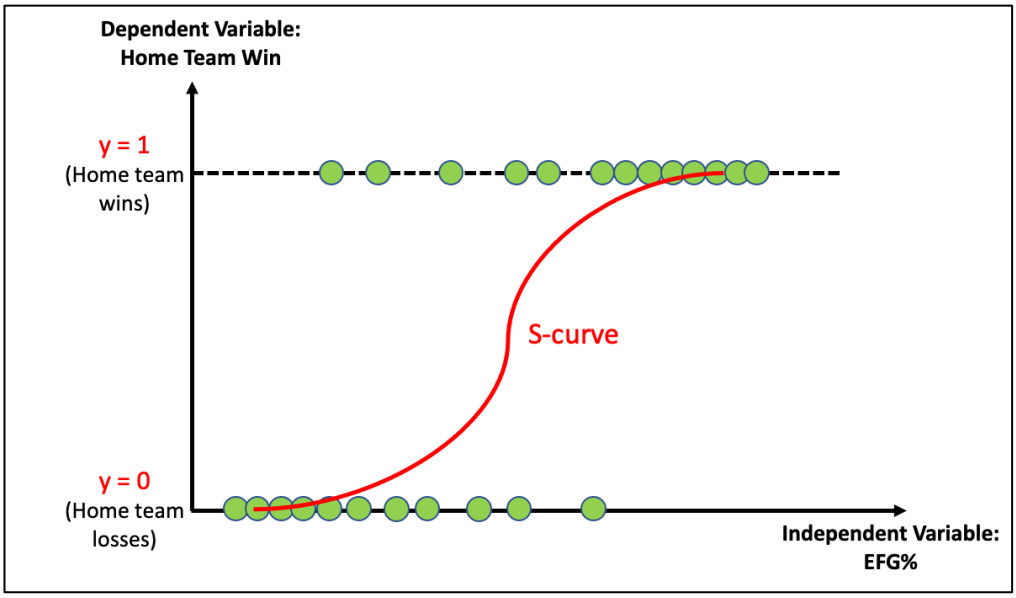


Figure 19: Mathematical Logic Behind Logistic Regression with Example

Using the logistic function, the model calculates the probability of belonging to classes for each observation and assigns a class for them if the calculated probability is bigger than the defined threshold value, commonly taken as 0.5, which is the case in our model. For modelling LogisticRegression() function from linear_model under Sklearn package is used. Accuracies of the Logistic Regression model under different scenarios are presented in Table 25.

Table 25: Accuracies for Logistic Regression Model

League & Season	Accuracies for Input Scenario 1	Accuracies for Input Scenario 2	Accuracies for Input Scenario 3
NBA 2016-17	60.2%	66.7%	62.9%
NBA 2017-18	60.2%	70.2%	68.3%
NBA 2018-19	58.8%	65.6%	62.1%
NBA Average	**59.7%**	**67.5%**	**64.4%**
TBL 2016-17	%68.1	83.3%	79.2%
TBL 2017-18	%70.8	68.1%	70.8%
TBL 2018-19	%68.3	74.6%	76.2%
TBL Average	**%69.1**	**75.3%**	**75.4%**

4.2.3 Gaussian Naive Bayes Classifier

Naive Bayes Classifier is used for classification problems by calculating the conditional probabilities based on Bayes Theorem. The Naive Bayes model assumes that each variable is independent and contributes to the outcome evenly. Gaussian Naive Bayes Classifier is a particular version of Naive Bayes Classifier, which assumes that continuous variables are normally distributed.

For modeling GaussianNB function from the naive_bayes package under Sklearn is used. There are two hyperparameters in the Gaussian Naive Bayes Classifier: prior probabilities of the classes and portion of the most significant variance. These hyperparameters are taken as the default, so that model calculates the prior probabilities from the training data.

Table 26: Accuracies for Gaussian Naive Bayes Classifier

League & Season	Accuracies for Input Scenario 1	Accuracies for Input Scenario 2	Accuracies for Input Scenario 3
NBA 2016-17	65.0%	64.2%	62.1%
NBA 2017-18	61.0%	68.3%	66.7%
NBA 2018-19	55.8%	64.2%	62.1%
NBA Average	**60.6%**	**65.6%**	**63.6%**
TBL 2016-17	66.7%	86.1%	80.6%

TBL 2017-18	70.8%	63.9%	63.9%
TBL 2018-19	60.3%	82.5%	81.0%
TBL Average	**65.9%**	**77.5%**	**75.2%**

4.2.4 K-Nearest Neighbors (KNN)

K-Nearest Neighbors algorithm tries to find the most similar k samples in the data for desired prediction sample point. KNN model can be used for both regression and classification. Model calculates the Euclidean distance between each data samples and selects the smallest k samples. Then takes the average value for regression or count the highest occurrence for classification. K is the most critical hyperparameter in this model and tuned in this study.

Table 27: Accuracies for K-Nearest Neighbors

League & Season	Accuracies for Input Scenario 1	Accuracies for Input Scenario 2	Accuracies for Input Scenario 3
NBA 2016-17	60.7%	65.9%	59.1%
NBA 2017-18	64.0%	66.4%	65.6%
NBA 2018-19	59.1%	63.1%	63.4%
NBA Average	**61.3%**	**65.1%**	**62.7%**
TBL 2016-17	72.2%	73.6%	69.4%
TBL 2017-18	73.6%	68.1%	56.9%
TBL 2018-19	58.7%	63.5%	58.7%
TBL Average	**68.2%**	**68.4%**	**61.7%**

4.2.5 Support Vector Classifier (SVC)

Support Vector Machine algorithm can be both used for regression and classification problems; however mostly it is used for classification problems due to its high capacity of separating the classes via hyperplanes created by support vectors and Kernel Function. Support vectors are selected from the samples by the model.

Support Vector Classifier has several key hyperparameters, including cost parameter, kernel function, and gamma parameter. The radial basis kernel function

is used in the model, which is widely used in most of the studies. Cost parameter and gamma parameter are determined via random search method. Cost parameter represents the penalty cost for misclassification. When it gets higher, it means the model tries to minimize miscalculated samples as much as possible by changing the hyperplane with a risk of overfitting. Gamma parameter defines how far the samples are considered as support vectors. Low gamma value means far away sample points from the hyperplane are also considered. The details of the cost and gamma hyperparameters can be found in Appendices. For modeling SVC function from SVM package under Sklearn is used. The results for all seasons and input scenarios are shown in Table 28.

Table 28: Accuracies for Support Vector Classifier

League & Season	Accuracies for Input Scenario 1	Accuracies for Input Scenario 2	Accuracies for Input Scenario 3
NBA 2016-17	61.5%	65.0%	63.4%
NBA 2017-18	65.5%	68.6%	67.8%
NBA 2018-19	58.0%	67.5%	59.6%
NBA Average	**61.3%**	**67.0%**	**63.6%**
TBL 2016-17	73.6%	81.9%	75.0%
TBL 2017-18	72.2%	62.5%	62.5%
TBL 2018-19	68.3%	69.8%	58.7%
TBL Average	**71.4%**	**71.4%**	**65.4%**

4.2.6 Decision Tree

The decision tree model is a tree-based algorithm like Random Forest, LGBM, and XGBoost. It is a rule-based algorithm that splits the data using nodes in a tree structure until it is isolated on the desired level.

Tuned hyperparameters are the criterion for split, maximum depth, the minimum number of samples for split, and maximum features. Usually, the Gini impurity index or entropy is used for the criterion for the split. Maximum depth is the limit for the longest branches in a Tree structure. If the number of samples in a node is less than the “minimum number of samples for split” hyperparameter, then the model does not

split this node anymore. Maximum features are the number of features to consider for determining the best split criteria. The details of hyperparameter tuning can be found in Appendices. For modelling DecisionTreeClassifier() function from tree package under Sklearn is used. The results for all seasons and input scenarios are shown in Table 29.

Table 29: Accuracies for Decision Tree

League & Season	Accuracies for Input Scenario 1	Accuracies for Input Scenario 2	Accuracies for Input Scenario 3
NBA 2016-17	59.9%	62.6%	60.2%
NBA 2017-18	61.2%	64.8%	62.3%
NBA 2018-19	57.2%	62.6%	54.5%
NBA Average	**59.4%**	**63.3%**	**59.0%**
TBL 2016-17	59.7%	75.0%	59.7%
TBL 2017-18	65.3%	62.5%	58.3%
TBL 2018-19	66.7%	71.4%	69.8%
TBL Average	**63.9%**	**69.6%**	**62.6%**

4.2.7 Random Forest

The Random Forest model is based on combined individual Decision Trees. Random Forest aims to get the advantage of the majority of votes and develop a more robust model since it is possible to fail in the single Decision Tree model.

Tuned hyperparameters are maximum depth, number of estimators, the minimum number of samples for split, and minimum number of samples for leaf. The maximum depth and a minimum number of samples for the split are the same with the ones in the Decision Tree. A number of estimators is a number of Decision Trees to be used for prediction. A minimum number of samples for leaf is the least required number of samples at the leaf node, which is the node at the bottom of the tree. The details of hyperparameter tuning can be found in Appendices. For modelling RandomForestClassifier() function from ensemble package under Sklearn is used. The results for all seasons and input scenarios are shown in Table 30.

Table 30: Accuracies for Random Forest

League & Season	Accuracies for Input Scenario 1	Accuracies for Input Scenario 2	Accuracies for Input Scenario 3
NBA 2016-17	57.5%	64.5%	61.0%
NBA 2017-18	65.9%	69.6%	67.8%
NBA 2018-19	59.1%	64.5%	63.1%
NBA Average	**60.8%**	**66.2%**	**64.0%**
TBL 2016-17	70.8%	79.2%	70.8%
TBL 2017-18	73.6%	68.1%	62.5%
TBL 2018-19	65.1%	77.8%	74.6%
TBL Average	**69.8%**	**75.0%**	**69.3%**

4.2.8 Light Gradient Boosting Machine (LGBM)

LGBM algorithms combine the tree structure with a gradient boosting framework. Gradient Boosting method also combines the set of decision trees; however unlike Random Forest, Gradient Boosting builds additive and dependent trees using the performance of the previous tree. The summary and difference between Decision Tree, Random Forests, and Gradient Boosting Trees can be found in Figure 20.

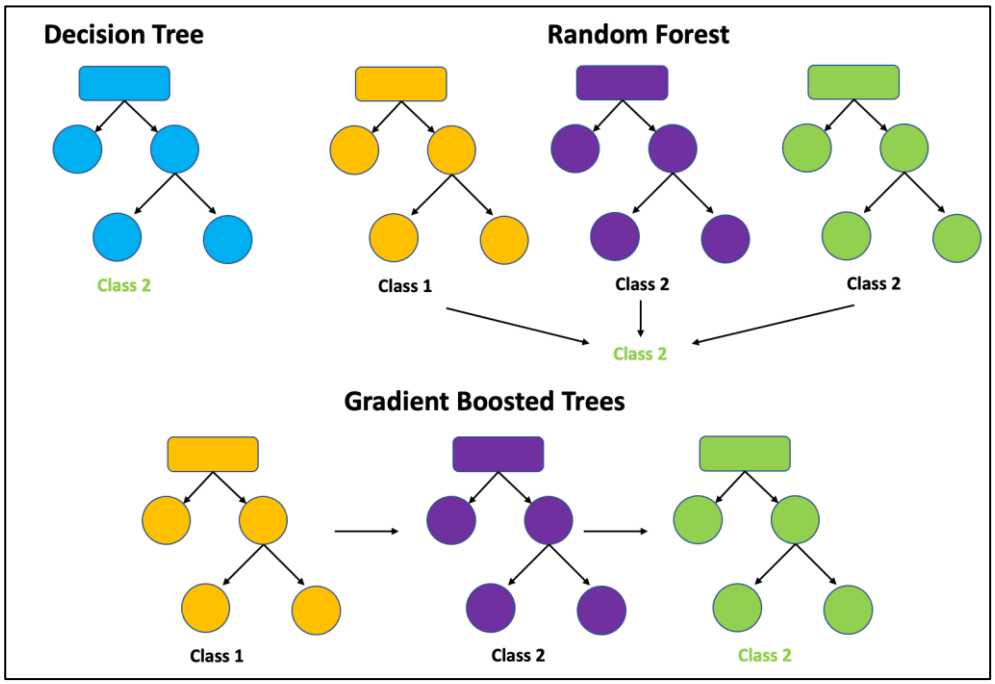


Figure 20: Decision Tree vs Random Forest vs Gradient Boosted Trees

LGBM grows trees vertically (leaf-wise) rather than horizontally (level-wise) used by other boosting algorithms. For that reason, LGBM is mainly known for its high speed among machine learning algorithms.

LGBM has several hyperparameters, including maximum number of leaves, minimum leaf weight, learning rate, L1 and L2 regularization terms to avoid overfitting, which are tuned in this model. LGBMClassifier() function is used in lightgbm package. The results for all seasons and input scenarios for LGBM are shown in the Table 31.

Table 31: Accuracies for LGBM

League & Season	**Accuracies for Input Scenario 1**	**Accuracies for Input Scenario 2**	**Accuracies for Input Scenario 3**
NBA 2016-17	61.8%	64.5%	63.7%
NBA 2017-18	64.2%	68.8%	61.5%
NBA 2018-19	59.3%	62.1%	59.1%
NBA Average	**61.8%**	**65.1%**	**61.4%**
TBL 2016-17	69.4%	79.2%	70.8%
TBL 2017-18	59.7%	68.1%	69.4%
TBL 2018-19	73.0%	76.2%	71.4%

TBL Average	**67.4%**	**74.5%**	**70.5%**

4.2.9 Extreme Gradient Boosting (XGBoost)

XGBoost is an improved version of the gradient boosting algorithm just like LGBM. The main difference between XGBoost and LGBM is how they compute the best split for trees. XGBoost uses a histogram-based algorithm, while LGBM uses a gradient-based one-side sampling technique.

XGBoost has several hyperparameters, including the tuned hyperparameters: learning rate, maximum depth of the tree, number of trees to build, minimum loss reduction for split (gamma), subsample ratio of training samples to prevent overfitting. XGBClassifier() function is used from xgboost package. The results for all seasons and input scenarios obtained from the XGBoost model are shown in the Table 32.

Table 32: Accuracies for XGBoost

League & Season	Accuracies for Input Scenario 1	Accuracies for Input Scenario 2	Accuracies for Input Scenario 3
NBA 2016-17	59.9%	63.1%	58.3%
NBA 2017-18	67.5%	69.1%	64.2%
NBA 2018-19	59.9%	64.5%	62.6%
NBA Average	**62.4%**	**65.6%**	**61.7%**
TBL 2016-17	72.2%	76.4%	70.8%
TBL 2017-18	72.2%	69.4%	73.6%
TBL 2018-19	74.6%	77.8%	71.4%
TBL Average	**73.0%**	**74.5%**	**71.9%**

4.2.10 Artificial Neural Network

An artificial Neural Network is a Deep Learning model which mimics human neuron's structure to connect inputs and output. In this structure, there are input layers, hidden layers, and output layers and related nodes in each layer. Nodes in

hidden layers have weight and activation functions to map inputs to output. Weights are determined iteratively using forward and backpropagation.

In this model, the Keras library is used. Two hidden layers are used, with 20 nodes in the first layer and 10 nodes in the second layer. The results for all seasons and input scenarios obtained from ANN model are shown in the Table 33.

Table 33: Accuracies for ANN

League & Season	**Accuracies for Input Scenario 1**	**Accuracies for Input Scenario 2**	**Accuracies for Input Scenario 3**
NBA 2016-17	62.1%	63.1%	59.9%
NBA 2017-18	56.6%	66.1%	64.5%
NBA 2018-19	60.2%	62.9%	65.3%
NBA Average	**59.6%**	**64.0%**	**63.2%**
TBL 2016-17	75.0%	83.3%	77.8%
TBL 2017-18	62.5%	66.7%	75.0%
TBL 2018-19	58.7%	69.8%	74.6%
TBL Average	**65.4%**	**73.3%**	**75.8%**

Summaries and findings from all model results can be found in Chapter 5.

CHAPTER 5

5 CONCLUSION

Various findings are obtained related to important factors for basketball game result, the difference in the dynamics of NBA and TBL from both descriptive analysis (Chapter 3) and prediction results (Chapter 4).

5.1 Summary of Descriptive Analysis

In Chapter 3, several factors and their relations with the game result are examined to understand how effective they are in basketball games and how their effects differ from NBA to TBL. Significant findings from the descriptive analysis are summarized below:

- Winning percentage is a simple but effective metric to explain team strength and game results in basketball. The winning percentage seems to be a better indicator in TBL compared to NBA. There is approximately an 10% difference in accuracies between NBA and TBL when the winning percentage is used as an only input for the game result. Moreover, winning percentage is becoming a more valid indicator as the season progresses.

- Four Factor metrics effectively measure team strength for both TBL and NBA, except for the free throw rate for TBL. Especially EFG% is the most important factor since it measures the shooting performance, and approximately 80% of the games are won by the teams who have a better EFG% than opposing teams in both NBA and TBL.

- It can be said that home-court advantage is a more critical factor in TBL, even if the winning percentage of home teams is approximately the same in both leagues with 59%. Because away teams are playing more back-to-back games

than home teams in NBA, which makes home teams more favorable. In TBL, scheduling is not a significant factor since each team plays one game per week, and the minimum number of resting days is 4. The average winning percentage of home teams in TBL is 59.4%, and it comes only from home-court advantage, while this number is 59.0%, and it includes both effect of home-court advantage and resting in NBA. Schedule and resting are critical factors in the NBA since the game load is not the same for the teams in short periods. When teams play back-to-back games, and opposite teams rest at least one game, the winning percentages of the teams are only 40%.

- Momentum and winning streak are important factors for both NBA and TBL. When teams have a winning streak for five or more games, the winning percentage of the next game is 64.3% for the NBA and 77.1% for TBL, which shows that the winning streak is an excellent indicator for power and momentum of the team. Moreover, momentum is not only valid for powerful teams, but also even weaker teams can gain momentum for a short period.

- One of the significant differences between NBA and TBL is the effect of individual players. The correlation between the performance of the best scorer of a team and team success is positive in NBA (+0.49), while it is negative in TBL (-0.43). This major difference shows that superstars in NBA are highly effective on my own to win the games, and teams are building their roster by putting them in the center of a team. Player-centered teams in TBL, however, are not successful as the teams who distributed the roles more equally and hierarchy between players are not so sure. Of course, it does not mean that star players are not influential in TBL; it only means they are not enough to carry a team to become successful. As the individual player effect is a more important factor in NBA, attention should be paid to transactions, injuries, and player-based momentum.

- The minute sharing effect is significant for TBL; however, it is not the case for NBA, and the level of impact is not so high compared to other factors. In TBL, the correlation coefficient between a number of wins and maximum minutes per game in a team is -0.71 and +0.64 for entropy. This supports that successful teams have a balanced role sharing, rather than focusing on stars in TBL, unlike NBA.

5.2 Summary of Prediction Results

There are numerous prediction scenarios under two leagues, three seasons, and 10 different models thus there are many things to compare and make a comment out of them. To make a comparison, home court advantages are taken as +50 points for TBL and +100 points for NBA in the ELO Rating Model.

There is a significant difference between NBA and TBL in terms of predictability. The comparison of maximum obtained accuracies of predictions models for NBA and TBL are presented in Figure 21.

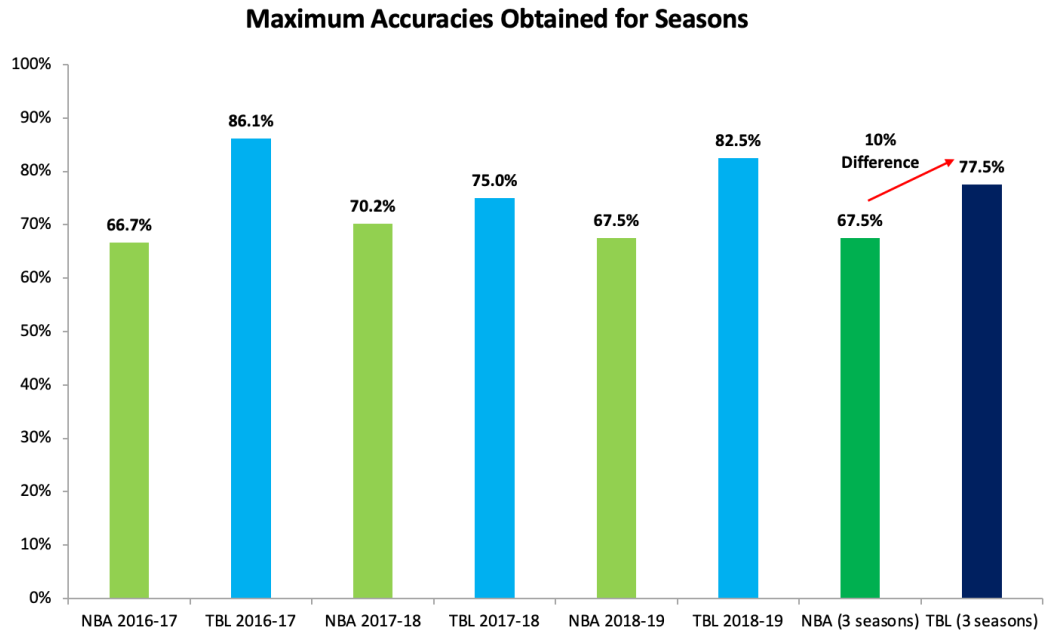


Figure 21: Comparison of best models for NBA and TBL by each season

TBL is more predictable than the NBA since the best-obtained accuracy for all seasons is approximately 10% higher in TBL than in the NBA. The major reason of the difference in predictability comes from the difference between the balance of power in these leagues. In TBL, there are teams at the top of the ladder, such as Fenerbahçe and Anadolu Efes, which makes it easier to predict the game result regardless of the home court or scheduling.

Moreover, variances of accuracies between seasons can be seen in Figure 21. For example, the best model accuracy was only 75.0% in the 2017-18 regular season in TBL, while this number was 86.1% in the 2016-17 season. This shows that there might be dramatic changes in dynamic and predictability between seasons, especially in TBL.

Additionally, result comparison for different input scenarios is presented in Figure 22.

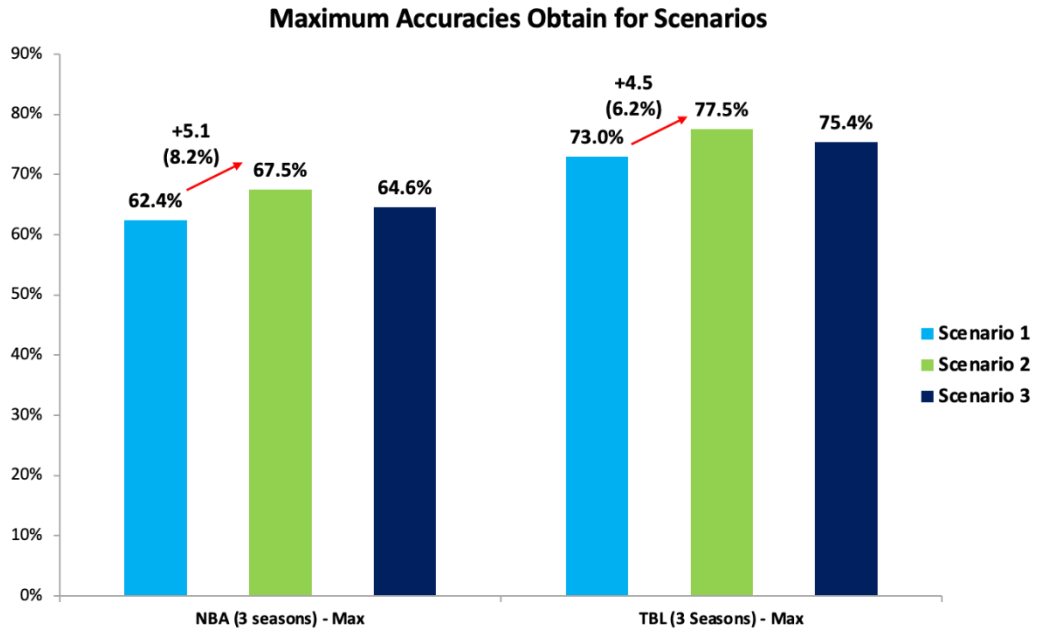


Figure 22: Comparison of best models for NBA and TBL under input scenarios

As shown in Figure 22, using advanced metrics (input scenario 2) increased the accuracy by approximately 4% to 5% compared to using basic team metrics in input scenario 1.

For model comparison, the best models for each season are presented in Figure 23.

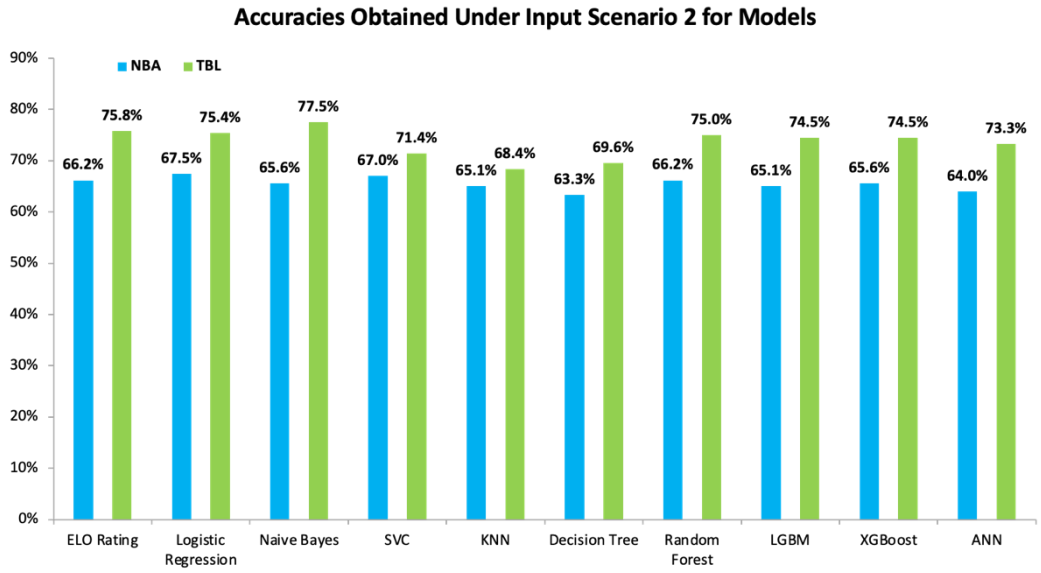


Figure 23: Comparison of models for NBA and TBL under input scenario 2

Logistic Regression and Naive Bayes perform better than other complex machine learning algorithms. Training size is small to learn better for other machine learning algorithms such as Random Forest or LGBM. Moreover, Decision Tree and KNN models give the worst performance considering both NBA and TBL.

Several important features are selected for input scenario 3 for each season. To determine most important variables, variables are ranked by a number of times they are picked for input scenario 3. Figures 24 and 25 show the most 25 features for NBA and top 20 features for TBL, respectively.

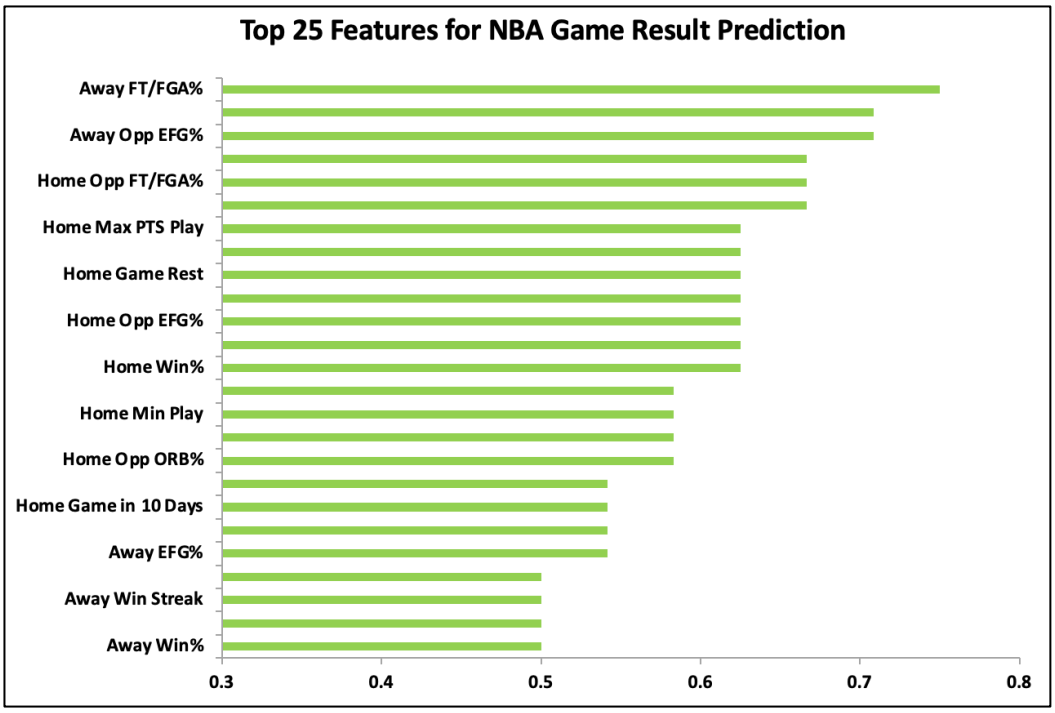


Figure 24: Top 25 Features for NBA based on SFS and most appearances

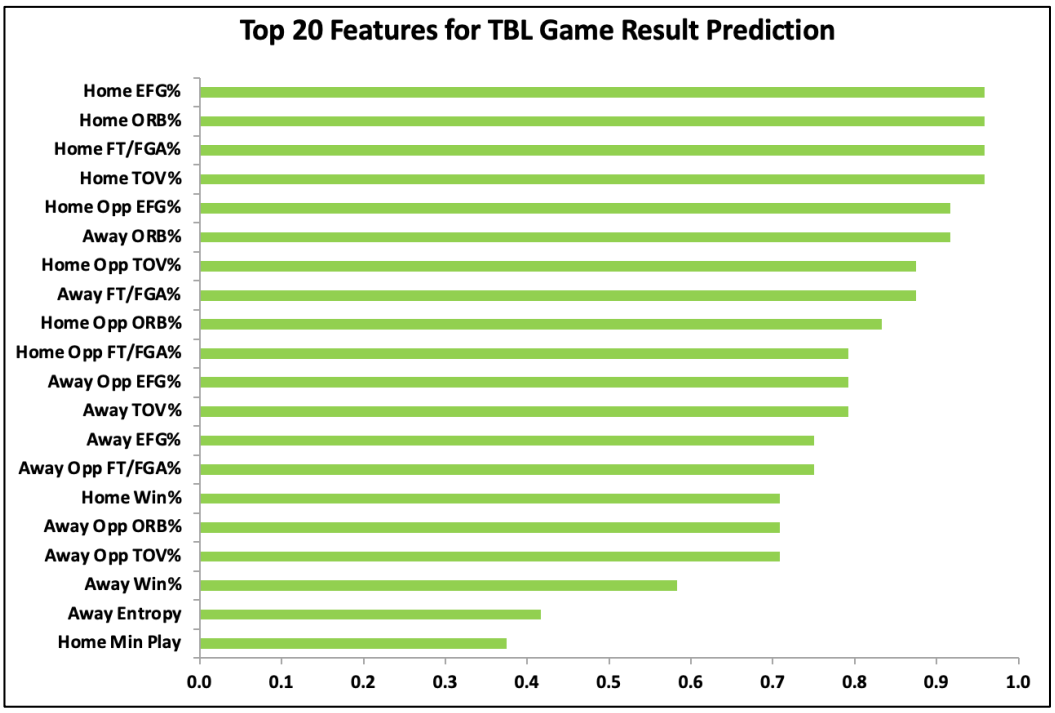


Figure 25: Top 20 Features for TBL based on SFS and most appearances

CHAPTER 6

6 FUTURE WORK

Future studies can expand the scope of the prediction of basketball game results by trying new input and modeling scenarios. Some of the future work recommendations are given below:

- New advance metrics can be generated as inputs to see if model performances increase. ELO rating for each team can be calculated and used as an input. Moreover, the importance of the game is not the same for teams thus it might be included to model using new metrics. For example, if a team has to win to make it to the playoffs, and the opposite team does not have such pressure, this affects the odds in favor of teams with playoff desire. This asymmetric situation can be modeled by generating and using a playoff probability metric.
- Analysis and game result prediction can be performed for other remarkable leagues like EuroLeague or Spanish Basketball League. By doing so, the effect of significant factors such as home court advantage can be compared with NBA and TBL.
- Game result prediction during the game can be studied using in-game data. This problem has some different challenges since the game is a continuous process with 48 or 40 minutes. Prediction models can use other inputs such as point difference, time remaining, in-game momentum, foul trouble, etc.
- Ensemble models can increase the accuracies. For that reason, ensemble models can be tried to test if this method increased the accuracy.
- Finally, play-off games can be modeled since dynamics of the regular season and playoffs are different. The main challenge in play-off game modeling is total number of games is too small compared to regular-season games.

REFERENCES

B.Jones, M. (2008). A Note on Team-Specific Home Advantage in the NBA. *Quantitative Analysis in Sports*.

Baghal, T. (2012). Are the "Four Factors" Indicators of One Factor? An Application of Structural Equation Modeling Methodology to NBA Data in Prediction of Winning Percentage. *Quantitative Analysis in Sports*.

Barry Schwartz, S. F. (1977). The Home Advantage. *Social Forces*.

Basketball Reference. (2021). Retrieved from https://www.basketball-reference.com/

Bernard Loeffelholz, E. B. (2009). Predicting NBA Games Using Neural Networks. *Quantitative Analysis in Sports*.

Carlin, B. P. (1996). Improved NCAA Basketball Tournament Modeling via Point Spread and Team Strength Information. *The American Statistican*.

David A. Harville, M. H. (2015). The Home-Come Advantage: How Large is it, and Does it Vary From Team to Team. *The American Statistician*.

Dragan Miljkovic, L. G. (2010). The Use of Data Mining for Basketball Matches Outcomes Prediction. *8th International Symposium on Intelligent Systems and Informatics.* Serbia: IEEE.

Dormann, Carsten F (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance *Ecography 27-46*

Elo, A. (1978). *The Rating of Chessplayers, Past & Present.* New York: FIDE.

Fadi Thabtah, L. Z. (2019). NBA Game Result Prediction Using Feature Analysis and Machine Learning. *Springer*.

Ge Cheng, Z. Z. (2016). Predicting the Outcome of NBA Playoffs Based on the Maximum Entropy Principle. *Entropy*.

Grand View Research. (2021). *Sports Analytics Market Size, Share & Trends Analysis Report By Component (Software, Service), By Analysis Type (On-field, Off-field), By Sports (Football, Cricket, Basketball, Baseball), And Segment Forecasts, 2021 - 2028.*

Haris Pojskic, V. S. (2011). Modelling Home Advantage in Basketball At Different Levels Of Competition. *Acta Kinesiologica*.

Jeremy Arkes, J. M. (2011). Finally, Evidence for a Momentum Effect in the NBA. *Quantitative Analysis in Sports*.

Jones, M. B. (2007). Home Advantage in the NBA as a Game-Long Process. *Quantitative Analysis in Sports*.

Justin Kubatko, D. O. (2007). A Starting Point for Analyzing Basketball Statistics. *Quantitive Analysis in Sports*.

Kelly, Y. J. (2009). The Myth of Scheduling Bias With Back-to-Back Games in the NBA. *Sports Economics*.

Manner, H. (2016). Modeling and forecasting the outcomes of NBA basketball games. *Quantitative Analysis in Sports*.

Manuela Cattelan, C. V. (2012). Dynamic Bradley-Terry modelling of sports tournaments. *Royal Statistical Society*.

Mark C. Drakos, B. D. (2010). Injury in the National Basketball Association: A 17-Year Overview. *Sports Health*.

Markets and Markets. (2020). *Sports Analytics Market by Sports Type (Individual and Team), Component, Application (Performance Analysis, Player Fitness and Safety, Player and Team Valuation, and Fan Engagement), Deployment Model, and Region - Global Forecast to 2024.*

Mavridis George, T. E. (2017). The inside game in World Basketball. Comparison between European and NBA teams. *International Journal of Performance Analysis in Sport*.

NBA. (2022). *NBA Stat*. Retrieved from https://www.nba.com/stats/

Oliver Entine, D. S. (2008). The Role of Rest in the NBA Home-Court Advantage. *Quantitative Analysis in Sports*.

Paul Kvam, J. S. (2006). A Logistic Regression/Markov Chain Model for NCAA Basketball. *Wiley InterScience*.

Pedro T. Esteves, K. M. (2020). Basketball performance is affected by the schedule congestion: NBA back-to-backs under the microscope. *European Journal of Sport Science*.

Ping-Feng Pai, L.-H. C.-P. (2016). Analyzing basketball games by a support vector machines with decision tree model. *The Natural Computing Applications Forum*.

Radivoj Mandic, S. J. (2019). Trends in NBA and Euroleague basketball: Analysis and comparison of statistical data from 2000 to 2017. *Plos One*.

Silver, N., & Fischer-Baum, R. (2015, May 21). *FiveThirtyEight*. Retrieved from https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/

Stump, M. (2017). Statistical Analysis of Momentum in Basketball.

TBL/BSL. (2021). Retrieved from https://www.tbf.org.tr/ligler/bsl-2021-2022

TBLstat. (2021). *TBLstat*. Retrieved from http://www.tblstat.net/

Thomas T. Byrnes, J. A. (2016). The Effect of Momentum on the NBA Point Spread Market. *The Sport Journal*.