8-2019

SUCCESSFUL SHOT LOCATIONS AND SHOT TYPES USED IN NCAA MEN’S DIVISION I BASKETBALL

Olivia D. Perrin
Northern Michigan University, operrin@nmu.edu

SUCCESSFUL SHOT LOCATIONS AND SHOT TYPES USED IN NCAA MEN'S DIVISION I BASKETBALL

By

Olivia D. Perrin

THESIS

Submitted to
Northern Michigan University
In partial fulfillment of the requirements
For the degree of

MASTER OF SCIENCE

Office of Graduate Education and Research

August 2019

SIGNATURE APPROVAL FORM

SUCCESSFUL SHOT LOCATIONS AND SHOT TYPES USED IN NCAA MEN'S DIVISION I BASKETBALL

This thesis by Olivia D. Perrin is recommended for approval by the student's Thesis Committee and Associate Dean and Director of the School of Health & Human Performance and by the Dean of Graduate Education and Research.

Committee Chair: Randall L. Jensen Date

First Reader: Mitchell L. Stephenson Date

Second Reader: Randy R. Appleton Date

Elizabeth Wuorinen Date
Associate Dean & Director of the School of Health & Human Performance

Dr. Lisa S. Eckert Date
Dean of Graduate Education and Research

ABSTRACT

SUCCESSFUL SHOT LOCATIONS AND SHOT TYPES USED IN NCAA MEN'S DIVISION I BASKETBALL

By

Olivia D. Perrin

The primary purpose of the current study was to investigate the effect of court location (distance and angle from basket) and shot types used on shot success in NCAA Men's DI basketball during the 2017-18 season. A secondary purpose was to further expand the analysis based on two additional factors: player position (guard, forward, or center) and team ranking. All statistical analyses were completed in RStudio and three binomial logistic regression analyses were performed to evaluate factors that influence shot success; one for all two and three point shot attempts, one for only two point attempts, and one for only three point attempts. Results indicated that guards are most likely to score as distance increases, when compared to forwards and centers. In addition, jump shots are most likely to be utilized successfully for every one-meter increase in distance, when compared to hook shots, tip shots, lay ups, and dunks. Results also indicated that, for further distances, the probability of shot success increases as angle decreases. The probability of shot success was also shown to be significantly influenced by team rank, with higher ranking teams having higher probabilities of shot success, although the magnitude of this effect was small and not practically relevant.

KEYWORDS: logistic, field goal, regression, collegiate.

DEDICATION

This thesis is dedicated to my parents, Joe and Carol. The foundation of every new endeavor has always been unwavering support from the two of you.

ACKNOWLEDGMENTS

The author wishes to thank her committee – Dr. Randall Jensen, Dr. Randy Appleton, and Mitchell Stephenson. Dr. Jensen allowed me to pursue my research passions and provided encouragement and feedback throughout the project. Dr. Appleton and Mitchell Stephenson offered valuable advice, made themselves available to answer any questions of mine, and also provided encouragement and feedback.

This thesis follows format requirements specified by Northern Michigan University's School of Health and Human Performance and the Journal of Quantitative Analysis in Sports, whose guidelines can be accessed at the link below.

https://www.degruyter.com/view/j/jqas#callForPapersHeader

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

Akaike Information Criteria .... AIC

Chi-Square Statistic .... χ^2

Degrees of Freedom .... df

Division I .... DI

Field Goal .... FG

International Basketball Federation .... FIBA

Key Performance Indicators .... KPI

Machine Learning .... ML

National Basketball Association .... NBA

Odds Ratio .... OR

Principal Component Analysis .... PCA

Rating Percentage Index .... RPI

Standard Error .... SE

The National Collegiate Athletic Association .... NCAA

Unstandardized Beta Coefficient .... β

Women's National Basketball Association .... WNBA

CHAPTER I: JOURNAL MANUSCRIPT

Introduction

Basketball is a court-based sport; characterized by intermittent high intensity efforts, during which players are required to repeatedly perform fast movements in association with unique technical actions according to specific tactics (Conte et al. 2018). The game is played as five versus five, where each player is categorized into one of the following five positions: point guard, shooting guard, small forward, power forward, and center. At the end of the game, the team that has scored the most points is declared the winner. Commonly known as a field goal, the non-free-throw shot is the primary way of scoring points in a game and is one of the most frequent and important technical elements in basketball (Erčulj and Štrumbelj 2015). The resulting points of a field goal are either two or three, depending on the location on the court where the shot was taken from. Players utilize different techniques when shooting; the choice of which depends on various factors such as distance away from the basket and player type (Erčulj and Štrumbelj 2015).

In the United States, basketball is played at several levels including high school, college, semi-professional, and professional. The National Collegiate Athletic Association (NCAA) is the pinnacle of collegiate basketball in America and can further be categorized into three levels: Division I (DI), Division II, and Division III. A range of talent exists between the three divisions, where the DI level typically contains the most sought after recruits from high school who are expected to be the most talented players. Each collegiate team plays a number of conference and non-conference games throughout the regular season, vying for an opportunity to be one of sixty-eight teams selected to compete in their division's NCAA tournament. The eventual winner of this tournament is crowned as the NCAA National Champion for their respective division.

Recent growth in the application of data analytics to basketball settings has expanded the investigation of quantifying players' tactical and technical in-game demands as well as game success in the men's game at the NCAA level. Akers, Wolff, & Buttross (1992) investigated factors that are important for winning in Division I Men's basketball, reporting that two point field goal percentage, turnovers, free throw percentage, steals, and rebounds were the most critical. Conte et al. (2018) agreed with these findings, but also reported defensive rebounds, free throws attempted, free throw rate, effective field goal percentage, and offensive rating as important factors in determining the outcome of a game. Many of these shooting-related, performance variables are known as key performance indicators (KPI) (Garcia et al. 2013; Gómez et al. 2008). Shot location and shot type impact shooting-related variables, which highlights the importance of shot location and type in regards to game outcome. However, the vast majority of published literature on shot location and shot type exists only at the professional level.

At the professional level, guards tend to play farther from the basket and also shoot more often from distance, while centers tend to play closer to the basket and are more likely to perform a dunk or tip in shot (Erčulj and Štrumbelj 2015; Miller and Bartlett 1996). The work of Harmon, Lucey, and Klabjan (2016) built off the above statement, noting that National Basketball Association (NBA) centers tend to have the highest shooting percentage, given that many of their shot attempts are close to the basket. In regards specifically to shot location, Harmon, Lucey, and Klabjan (2016) determined that the probability of making a shot decreases as distance away from the basket increases. A second study reported that, across many competition levels, more successful teams on average attempt fewer three point field goals (Erčulj and Štrumbelj 2015). With no research to date on shot location and shot type at collegiate level, little is known how the above findings translate to the college game.

The primary purpose of the current study was to investigate the effect of court location and shot types used on made field goals in NCAA Men's DI basketball during the 2017-2018 season. A secondary purpose of the current study was to further expand the analysis based on two additional factors: player position (guard, forward, or center) and team ranking at the end of the regular season. The exploration of shot location and shot type allows for a deeper understanding of key performance variables and enhances the ability to explain the way in which they shape game outcomes at the collegiate level.

Methods

Data Acquisition

Participants of the current study were basketball players who participated in NCAA Men's DI basketball during the 2017-18 season. All data were publicly available online and were obtained from a dataset created by the NCAA on Google Cloud Platform (https://console.cloud.google.com/marketplace/details/ncaa-bb-public/ncaa-basketball?pli=1). Therefore, the current study was exempt from requiring consent of participants. Approval for this study was granted by the Human Subjects Institutional Review Board of Northern Michigan University, Marquette, Michigan, USA (HS19-1044). This free online dataset contains several tables of data about NCAA basketball players, games, and teams. The tables for games consist of play-by-play data, box score data, and final scores. Basketball shot data was extracted from the play-by-play table by selecting the following variables from the dataset: *game id*, *team market*, *team basket*, *event coordinate x*, *event coordinate y*, *shot made*, *shot type*, *and position*. Data from these variables were filtered to only reflect shots (made or missed) that were taken during the 2017-18 regular season in DI Men's basketball. As a result, 218,696 basketball shots from the 2017-18 season representing 333 out of 351 DI Men's basketball teams were extracted.

Data Reduction and Preparation

All data reduction and preparation was completed in RStudio (Version 1.2.1335, RStudio, Inc., Boston, MA, USA). A total of 927 shots were excluded from the dataset. Of these shots, eight hundred of them were excluded due to errors in shot classification where court location did not align with the classification of a two or three point shot attempt, respectively. Three shots were excluded due to missing values for multiple variables. Close examination of the data revealed a skewed frequency of games per team. Therefore, to avoid a biased model, only teams with ten or more reported games were included. The remaining 124 shots were excluded due to the shot attempts coming from beyond the half court line. After these changes, a total of 185,253 shots from 131 teams remained. An illustration of the distribution of these shots across shot distance (Figure 1) and angle (Figure 2). Shot frequencies across position and shot type are shown in Table 1.

Locations for each shot were extracted as x, y coordinates (*event coordinate x*, *event coordinate y*). The x coordinate was reported as the location of the play in number of inches from the "left" baseline, while the y coordinate was reported as the location of the play in number of inches from "top" sideline. For simplicity purposes, all coordinate data were converted from Cartesian coordinates to polar. The origin was transposed for each shot to its respective net using *team basket* and all values were converted from inches to meters. The polar location of each shot was reported in the form of two new variables: *distance* (meters) (Equation 1) and *angle* (radians) (Equation 2). Finally, all angles were converted from radians to degrees to increase sensitivity in the model.

Equation 1. Polar distance calculation

$$\text{Distance} = \sqrt{x^2 + y^2}$$

Equation 2. Polar angle calculation

$$\text{Angle} = \tan^{-1}\left(\frac{y}{x}\right)$$

In addition to the variables above, another variable (*team rank*) was created, which contained the ranking of each team at the conclusion of the regular season (February 25th, 2018) according to the Rating Percentage Index (RPI). RPI was the official metric used for team ranking by the NCAA for the 2017-18 season. All team ranking data were publicly available online and obtained from: https://www.teamrankings.com/ncaa-basketball/rpi-ranking/rpi-rating-by-team?date=2018-02-25. The variable *shot made* was left as is and used as the dependent variable in the analyses. The variable *shot made* contains two levels: "0" and "1" where "0" represents a missed shot and "1" represents a made shot. For this variable, level "0" (missed shot) was set as the reference category for each of the respective analyses. The reference level is the level to which every other level is compared against. Detailed explanations of the independent variables of interest and their associated levels are provided below.

Independent Variables of Interest

The current study focused on the following independent variables:

- *Shot type*– divided into five levels (jump shot set as reference category):

- Jump shot: This occurs when a player jumps in the air and shoots the ball above their head. This is the most common type of shot used when shooting from distance, but can also be utilized when a player is near the basket.
- Hook shot: This occurs when a player shoots the ball turned approximately perpendicular to the basket by bringing the arm farther away from the basket up overhead in a sweeping motion, extending the shoulder movement and flexing the wrist.
- Layup: This is a one-handed shot that occurs when a player releases the ball after an upwards motion of the arm. This shot is typically executed close to the basket by jumping off one leg and bouncing the ball off the backboard.
- Tip-shot: This occurs when a player leaps into mid-air and tips the ball into the basket on a rebound.
- Dunk: This occurs when a player slams the ball down through the basket with their hands above the rim. Only players with sufficient height or vertical jump are able to execute this shot.

The following two independent variables, which may influence shot type selection, were also of interest:

- *Location*– With respect to the basket, the location on the court where the shot was taken. Location was split into the following two variables for the analyses:
    - Distance: The distance (in meters) away from the basket.

    - Angle: The angle (degrees) away from the midline (the imaginary line that divides the court in half from rim to rim).

- *Position* – Players of different positions have different roles in basketball (Dežman, Trninić, and Dizdar 2001), motor skills (Erčulj et al. 2009), and anthropometric dimensions (Erčulj and Štrumbelj 2015; Sampaio et al. 2006). Therefore, different shooting tendencies are expected among different player positions. Position was divided into three levels (Guard was set as the reference category):
    - Guard: This position generally facilitates scoring opportunities for other teammates, as well as for themselves. They primarily handle the ball on offense.
    - Forward: This position generally possesses quickness and strength, attacking the basket from the "wing" location (outside and near baseline). They are sometimes interchangeable with the guard position.
    - Center: This position generally operates inside of the three point line and close to the basket.

Statistical Analysis

All statistical analyses were completed in RStudio. All categorical variables were temporarily converted to numeric to assess multi-collinearity in a correlation matrix. All pairs yielded small correlation coefficients (less than 0.25), confirming the absence of multi-collinearity. All categorical variables (*position*, *team rank*, *shot type*, and *shot made*) were then converted to factors for the analyses. Three independent binomial logistic regression analyses were performed, where *shot made* was used as the dependent variable in all models. The first binomial logistic regression evaluated factors influencing shot success across both two and three point shot attempts, whereas the second and third models evaluated only two and only three point shot attempts, respectively. The variable *three point shot* was used to accurately identify instances of three point shot attempts. The variable *shot type* was excluded from the three point only model as jump shots were used exclusively in this model. An exploratory process was utilized where multiple combinations of variables and interactions were assessed. The most appropriate option for each respective model was selected based on model performance and parsimony. In regards to model performance, the selection process for each model was implemented through the completion of a likelihood ratio test and Wald test, as well as the assessment of values for the Akaike Information Criteria (AIC), log likelihood, and McFadden Pseudo R^2. Results for both the likelihood ratio test and Wald test are expressed as a Chi-Square statistic (χ^2) with their associated degrees of freedom (df) and p-value. Odds ratios with 95% confidence intervals were also calculated and reported. Odds ratios and associated 95% confidence intervals for interaction terms were manually calculated (Equation 3, Equation 4) using an adjusted standard error (Equation 5).

Equation 3. Odds Ratio calculation for interaction terms

$$OR_{\beta 1+\beta 2} = e^{\beta 1+\beta 2}$$

Equation 4. 95% Confidence Interval calculation for interaction terms

$$95\%\ CI = e^{OR_{\beta 1+\beta 2} \pm \left(1.96 \times SE_{\beta 1+\beta 2}\right)}$$

Equation 5. Standard Error calculation for interaction terms

$$SE_{\beta 1+\beta 2} = \sqrt{\left(SE_{\beta 1}^2 + SE_{\beta 2}^2 + 2 * Cov(\beta 1, \beta 2)\right)}.$$

The model equations for the selected binomial models are presented below in Equation 6, Equation 7, and Equation 8:

Equation 6. Binomial model 1 equation for all two and three point shots

$$\begin{aligned} Ln\left(\frac{Pi}{1-Pi}\right) = {} & \beta_0 + \beta_1 Distance_i + \beta_2 Forward_i + \beta_3 Center_i + \beta_4 Hook\ shot_i + \beta_5 Tip\ shot_i \\ & + \beta_6 Lay\ up_i + \beta_7 Dunk_i + \beta_8 Angle_i + \beta_9 Team\ Rank_i \\ & + \beta_{10}(Distance_i \times Forward_i) + \beta_{11}(Distance_i \times Center_i) \\ & + \beta_{12}(Distance_i \times Hook\ shot_i) + \beta_{13}(Distance_i \times Tip\ shot_i) \\ & + \beta_{14}(Distance_i \times Lay\ up_i) + \beta_{15}(Distance_i \times Dunk_i) \\ & + \beta_{16}(Distance_i \times Angle_i) \end{aligned}$$

Equation 7. Binomial model 2 equation for all two point shots

$$\begin{aligned} Ln\left(\frac{Pi}{1-Pi}\right) = {} & \beta_0 + \beta_1 Distance_i + \beta_2 Forward_i + \beta_3 Center_i + \beta_4 Hook\ shot_i + \beta_5 Tip\ shot_i \\ & + \beta_6 Lay\ up_i + \beta_7 Dunk_i + \beta_8 Angle_i + \beta_9 Team\ Rank_i \\ & + \beta_{10}(Distance_i \times Forward_i) + \beta_{11}(Distance_i \times Center_i) \\ & + \beta_{12}(Distance_i \times Hook\ shot_i) + \beta_{13}(Distance_i \times Tip\ shot_i) \\ & + \beta_{14}(Distance_i \times Lay\ up_i) + \beta_{15}(Distance_i \times Dunk_i) \end{aligned}$$

Equation 8. Binomial model 3 equation for all three point shots

$$Ln\left(\frac{Pi}{1-Pi}\right) = \beta_0 + \beta_1 Distance_i + \beta_2 Forward_i + \beta_3 Center_i + \beta_4 Angle_i + \beta_5 Team\ Rank_i$$

Results

As shown in Table 2, all variables and interaction terms were significant ($p < 0.05$) for the first model with all two and three point shot attempts. When interpreting the odds ratios of the interaction between distance and position, results indicated that forwards were 7.17% less likely to make a shot for every one-meter increase in distance when compared to the reference category of guard, adjusting for shot type, angle, and team rank. Centers were 8.68% less likely to make a shot for every one-meter increase in distance compared to the position of guard, when adjusting for shot type, angle, and team rank.

In relation to the odds ratios for the interaction between distance and shot type, a player utilizing a hook shot was 30.58% less likely to make the shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. A player utilizing a tip shot was 34.08% less likely to make the shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. Results also indicated that a player utilizing a layup was 54.18% less likely to make the shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. A player utilizing a dunk was 53.59% less likely to make the shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank Team rank was also statistically significant, although the magnitude of the effect was small. In relation to the interaction between distance and angle, results indicate that for further distances the probability of shot success increases as angle decreases.

As shown in Table 3, all variables and interactions were significant ($p < 0.05$) with the exception of the interaction between distance and the position of center ($p > 0.05$). When interpreting the odds ratios of the interaction between distance and position, results indicated that forwards were 5.49% less likely to make a two point shot for every one-meter increase in distance when compared to the reference category of guard, adjusting for shot type, angle, and team rank. In relation to the odds ratios for the interaction between distance and shot type, a player utilizing a hook shot was 28.99% less likely to make a two point shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. A player utilizing a tip shot was 32.84% less likely to make a two point shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. Results also indicated that a player utilizing a layup was 53.00% less likely to make a two point shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. A player utilizing a dunk was 52.74% less likely to make a two point shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. Team rank and angle were also statistically significant, although the magnitude of these effects were small.

As noted in Table 4, all variables were significant ($p < 0.05$) with the exception of the position of center ($p > 0.05$). When interpreting the odds ratio for distance, results indicated that a player is 12.34% less likely to make a three point shot for every one-meter increase in distance. In relation to the odds ratio for the position of forward, players in this position are 7.49% less likely to make a three point shot when compared to the reference category of guard, adjusting for distance, angle, and team rank. Angle and team rank were also statistically significant, although the magnitude of these effects were small.

Discussion

The primary purpose of the current study was to investigate the effect of court location and shot types used on made field goals in NCAA Men's DI basketball during the 2017-2018 season. A secondary purpose of the current study was to further expand the analysis based on two additional factors: player position (guard, forward, or center) and team ranking at the end of the regular season. Court location, shot type, and player position were all shown to significantly influence the probability of shot success. In addition, team rank and angle were also shown to significantly influence the probability of shot success, although the tangible impact of these changes were minimal.

As shown in Table 2, team rank was statistically significant at the $p < 0.0001$ level. For this variable, the odds ratio was reported as 0.9994. Results from Tables 2 and 3 show similar odds ratios for team rank; 0.9994 and 0.9992, respectively. Although the odds ratios for this variable are all very close to 1, they can still be interpreted as greater numbers for ranking decrease the probability of shot success (greater numbers for team rank indicate a lower ranking team). However, this change is minimal given that for every one unit increase in team rank, the probability of shot success only decreased by approximately 0.06 – 0.08%. Although it generally could be expected that a lower ranking team may shoot at a lower percentage than a higher ranking team, the magnitude of this effect was very small and lacks practical significance. Mikołajec, Maszczyk, and Zając (2013) investigated factors influencing NBA team rank, concluding that shooting-related variables such as offensive efficiency and third quarter points per game were factors that significantly influenced team rank. Although the magnitude of the effect in the current study was small, the findings of Mikołajec, Maszczyk, and Zając (2013) provides some support for the significant relationship identified between team rank and probability of shot success. Additionally, it is possible that the dependent variable used in the current study was too broad to see meaningful changes for team rank. Given the results of Mikołajec, Maszczyk, and Zając (2013), it may be possible that team rank is more likely to influence only selective aspects of shooting performance as opposed to shot success. Furthermore, another consideration when interpreting this relationship is that the NCAA replaced the use of RPI as their metric for measuring team rank following the 2017-18 season. Future analyses utilizing the newly adopted ranking system may yield different results.

In the current study, location was comprised of distance and angle from the basket. When considering angle, past literature has reported no effect on the probability of basketball shot success (Erčulj and Štrumbelj 2015). However, as noted in Tables 2 and 3, the results of the current study identified that angle significantly influenced the probability of shot success. Interestingly, the results from each of the two models contradict each other. For the second model, when looking at only two point shots, results indicated that every one unit increase in angle decreases the probability of shot success by 0.11%. For the third model, when looking only at three point shots, results indicated that every one unit increase in angle increases the probability of shot success by 0.13%. Although these results are conflicting, it is important to note that the magnitude of the effect of angle for both models is minimal. Therefore, from a practical standpoint, this information is not meaningful to players and coaches.

In the first model, which looked at all shots, a significant interaction between distance and angle was present. As seen in Figure 2, results of this interaction indicated that, at farther distances, the probability of shot success decreases as angle increases, adjusting for team rank, position, and shot type. Therefore, it may be suggested that, as distance increases, players experience higher probabilities of shot success if they take their shots closer to the midline and away from the baseline. However, it is important to note that the 95% confidence interval for this interaction crosses 1 (0.9319-1.0023), which means that this interaction should be interpreted with caution (Tan and Tan 2010) and may not be practically relevant for players and coaches.

Previous literature focused on basketball has documented that court location, among several other factors, influences a player's choice of shot type used (Erčulj and Štrumbelj 2015). Additionally, it has also been reported that the jump shot is the most common shot used when a player shoots from distance (Erčulj and Štrumbelj 2015). The results of the first and second model (shot type was excluded in the third model) in the current study support both of the above findings. For tip shots, dunks, hook shots, and lay ups, the probability of shot success decreased for every one-meter of increase in distance compared to a jump shot, adjusting for player position, angle, and team rank. This indicates that jump shots were utilized most successfully when players took shots from distance, which aligns with the findings of Erčulj and Štrumbelj (2015) that the jump shot is the most common shot used when a player shoots from distance. The increased difficulty of shot making from distance is emphasized by the results of the current study given that the probability of shot success decreased with distance for all shot types when compared to a jump shot. These findings are in agreement with Harmon, Lucey, and Klabjan (2016), who reported that as distance away from the basket increases, the probability of making a shot decreases. Moreover, knowing that increased distance decreases the probability of shot success for some shot types more or less than others, the findings of the current study strengthens the account of Erčulj and Štrumbelj (2015), who noted that court location significantly influences a player's choice of shot type used. In a practical sense, this aligns with what is commonly known about basketball since shot types such as dunks, layups, and tip-ins can only be utilized closer to the basket. In other words, the limits in regards to distance for these two shot types are restricted to a player's physical capabilities including their ability to leap horizontally and vertically. For example, it would be common to see a jump shot attempted from the free throw line area, but not a layup, dunk, or tip in as the basket would be extremely difficult to reach when jumping from that far away.

When looking more specifically at how each shot type interacts with distance, further discussion is warranted. As shown in Table 2, when considering all shots, results from the first model indicated that a player utilizing a layup was 54.18% less likely to make the shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. Furthermore, a player utilizing a dunk was 53.59% less likely to make the shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. Although both of these interactions were statistically significant, these results do not provide any additional information above what is already commonly understood within the game of basketball. As discussed above, for either a layup or a dunk it is required for players to be at or very near to the basket for the execution of the shot. Consequently, it is expected that the probability of shot success for these shots would significantly decrease as distance increases. As a result, this information is not practically relevant to coaches or players. As shown in Table 3, odds ratios from the second model looking at only two point shots were similar in magnitude and direction for these two shot types and their respective interactions with distance. This comes as no surprise considering that, as discussed above, most all dunks and layups are executed close to the basket or relatively short distances.

As shown in Table 2, results from the first model indicated that a player utilizing a hook shot was 30.58% less likely to make the shot for every one-meter increase in distance compared to a jump shot when adjusting for position, angle, and team rank. Results of the same interaction from the second model align closely with the first model at 28.99%. Although hook shots can be taken farther away from the basket, it is possible that in comparison to a jump shot, these results speak primarily to the difficulty of the hook shot, especially as distance increases. Erčulj and Štrumbelj (2015) noted that the technique of a hook shot is rarely practiced in the NBA like it used to be, which has resulted in a lower relative frequency of the shot over time. These findings may support the lack of popularity and overall use of the shot at the collegiate level, which may indicate that collegiate players don't practice the technique of a hook shot often. If this was the case, players may have used inferior technique for the hook shot during a game, resulting in many missed shots, especially those with a higher difficulty at a greater distance.

In regards to player position in basketball, it has been noted in past literature that players in the NBA, at the position of center, tend to play closer to the basket and attempt many of their shots close to the basket (Erčulj and Štrumbelj 2015; Harmon, Lucey, and Klabjan 2016). It has also been reported that guards at the professional level play farther from the basket and also shoot more often from distance (Erčulj and Štrumbelj 2015; Miller and Bartlett 1996). In an analysis of basketball shots from players across different competition levels, Erčulj and Štrumbelj (2015) noted that guards were the most accurate shooting followed by forwards and finally centers, with respect to distance.

The results of both the first and second models in the current study are in agreement with the findings of Erčulj and Štrumbelj (2015), demonstrating that guards had the highest probability of shot success as distance increases. As shown in Table 2, results from the first model including all shots indicated that forwards were 7.17% less likely to make a shot for every one-meter increase in distance when compared to guards, adjusting for shot type, angle, and team rank. Centers were 8.68% less likely to make a shot for every one-meter increase in distance compared to guards, when adjusting for shot type, angle, and team rank. Unlike the first model, only the interaction between distance and the position of forward was significant in the second model. However, the interaction between distance and the position of center in the second model was trending in the direction of statistical significance, which is similar to the findings for this interaction in the first model. This suggests that centers are less successful shooters as distance increases even when three point attempts are not taken into consideration. Interestingly, the results between the two models for the interaction between distance and the position of forward were only slightly different; as the results from the two point shot model indicated that forwards were 5.49% less likely to make the shot for every one-meter increase in distance when compared to guards, adjusting for shot type, angle, and team rank. This suggests that, at closer distances, forwards experience probabilities of shot success that are close to that which guards exhibit.

As mentioned above, both the first and second model indicated that guards are the most likely to make shots as distance increases. A potential factor explaining the superior performance of guards' shot success from distance may be linked with their ability to maintain consistent shooting form as distance increases when compared to centers, as noted by Miller and Bartlett (1996). It is suggested that their ability to adapt to varying distances may be due to the increased frequency of shooting from distance for this position (Miller and Bartlett 1996). The findings of Miller and Bartlett (1996), when considered alongside the work of Harmon, Lucey, and Klabjan (2016) as well as Erčulj and Štrumbelj (2015), demonstrates that guards shoot more often from distance and therefore develop strategies to allow them to be consistently successful in these locations, when compared to other positions such as centers.

As mentioned previously, the third model examined all variables except shot type and the data for this model included three point shots only. The results for angle and team rank from this model were briefly discussed above in previous sections, leaving position and distance, which both warrant discussion. When looking at position, results indicated that forwards were 7.48% less likely to make a three point shot than guards. The center position was not significantly different from the guard position ($p > 0.05$). In regards to distance, results of this model indicated that for every one-meter increase in distance behind the three point line, a player is 12.34% less likely to make that three point shot. This result is especially interesting given that the NCAA recently approved a rule change which will move the NCAA three point line back from 6.32 meters to 6.75 meters, effective immediately for the upcoming 2019-20 season (Johnson 2019). Based on the results of the third model, it may be suggested that players will make less three point shots than previous years. The results of this model also indicate that, when considering position, forwards may struggle in the upcoming year behind the arc more than guards. This is practically relevant information for players and coaches and may suggest that shooting a high percentage from behind the three point line in the coming years may require players to regularly practice this shot to get used to shooting from farther distances.

A limitation of the current study was the uneven distribution of shot data between teams in DI Men's basketball. For this reason, only teams with shot data for ten or more games were included in the analysis. As a result, many teams from Men's DI were not represented in the data and even some entire conferences were excluded. Therefore, the results may not be applicable for all teams in the division. Another limitation of the current study was that, although the sample size was large (n = 185,377), all shot data used in the analysis were only from one season (2017-18). Each year in NCAA basketball, the landscape of players change considerably as many new players arrive as freshmen and other players graduate or may even leave to play in the NBA. As a result, it may be possible to see varied results when analyzing data from different seasons.

One other limitation of the current study was that temporal aspects of the game were not analyzed. It has been documented in previous literature that temporal aspects of the game such as pace or time left on the shot clock influence shot location and shot type selection (Erčulj and Štrumbelj 2015; Skinner 2012). Given this information, it is possible that if some of these temporal aspects were analyzed, these variables could have helped provide greater explanatory power in regards to the probability of shot success. The dataset used for the current study also did not include tactical factors such as defensive pressure on the ball or defensive strategies used against offensive players. These factors have been shown to influence court locations used to take shots as well as shot success (Csapo and Raab 2014; Gomez, Gasperi, and Lupo 2016). As mentioned above, the inclusion of factors such as these may have strengthened the model, providing another layer of understanding of shot success when considering the locations and shot types used at the collegiate level.

Conclusion

In conclusion, the current study demonstrates that the probability of shot success in NCAA Men's DI basketball was significantly influenced by shot type, court location (distance, and angle), position, and team rank. Although these variables were statistically significant, they may not be practically significant to implement in a real-world team sports environment, as the magnitude of the effect observed in some variables (team rank, angle) were minimal. Overall, the results of the current study indicated that guards were the most successful shooters from distance, most often utilizing a jump shot, which had the highest probability of shot success from distance when compared to all other shot types. These findings align with previous literature at the professional level. This suggests that shooting characteristics of collegiate players are similar to that of professional players, which may indicate that appropriate offensive strategies are utilized at the collegiate level, with respect to developing players for the professional level. Future research investigating factors that influence the probability of shot success should incorporate temporal and tactical aspects of the game, which may have the potential to further enhance the understanding of successful shooting at the collegiate level.

Tables and Figures

Table 1. Shot frequencies according to shot type and position.

	Total	Jump shot	Hook shot	Tip shot	Lay up	Dunk
Total	185,253	113,578	4,054	2,272	56,717	8,632
Guard	113,538	78,788	245	634	31,532	2,339
Forward	63,091	31,871	2,860	1,365	21,860	5,135
Center	8,624	2,919	949	273	3325	1,158

Table 2. Results relating to the binomial logistic regression model 1 for all two and three point shot attempts (dependent variable = shot made [0=missed, 1=made])

Independent Variables	β (SE)	χ^2	*p*	OR (95% CI)
(Intercept)	-0.3623 (0.0247)	214.5053	<0.0001	0.6961 (0.6631-0.7306)
Distance	-0.0348 (0.0042)	67.3384	<0.0001	0.9658 (0.9579-0.9739)
Forward	0.2149 (0.0186)	134.0038	<0.0001	1.2398 (1.1955-1.2857)
Center	0.3316 (0.0390)	72.2840	<0.0001	1.3932 (1.2908-1.5041)
Hook shot	0.9636 (0.1043)	85.3776	<0.0001	2.6210 (2.1378-3.2176)
Tip shot	1.1390 (0.1025)	123.4099	<0.0001	3.1243 (2.5588-3.8254)
Lay up	1.5450 (0.0274)	3176.7878	<0.0001	4.6890 (4.4439-4.9480)
Dunk	3.1490 (0.0764)	1699.8304	<0.0001	23.3165 (20.0955-27.114)
Angle	-0.0024 (0.0004)	42.3150	<0.0001	0.9976 (0.9969-0.9983)
Team Rank	-0.0006 (0.0001)	90.0221	<0.0001	0.9994 (0.9992-0.9995)
Distance*Forward	-0.0396 (0.0040)	96.8256	<0.0001	0.9283 (0.8951-0.9627)
Distance*Center	-0.056 (0.0124)	20.2860	<0.0001	0.9132 (0.8806-0.9471)
Distance*Hook shot	-0.3303 (0.0473)	48.8601	<0.0001	0.6942 (0.6693-0.7199)
Distance*Tip shot	-0.3819 (0.0969)	15.5472	<0.0001	0.6592 (0.6354-0.6840)
Distance*Lay up	-0.7457 (0.0161)	2139.6175	<0.0001	0.4582 (0.4418-0.4752)
Distance*Dunk	-0.733 (0.0668)	120.5384	<0.0001	0.4641 (0.4475-0.4813)
Distance*Angle	0.0006 (0.0001)	74.8052	<0.0001	0.9664 (0.9319-1.0023)
Model Performance		χ^2	*p*	*df*
Likelihood ratio test	-	17,313.0000	<0.0001	16
Wald Test	-	13,342.2400	<0.0001	16
Log Likelihood	-118,950.3000	-	-	-
McFadden Pseudo R^2	0.0678	-	-	-
Akaike Information Criterion	237,935.000	-	-	-

Note: R Programming code: [glm(formula = Shot made ~ Distance*Shot Type + Distance*Position + Distance*Angle + Team Rank, family = binomial(link = logit)]. 95% CI is 95% confidence interval, β is the unstandardized beta coefficient, OR is the odds ratio, SE is the standard error, χ^2 is the Wald's Chi-Square. Statistical significance accepted at <0.05. All statistics reported herein use 4 decimal places in order to maintain statistical precision.

Table 3. Results relating to the binomial logistic regression model 2 for all two point shot attempts (dependent variable = shot made [0=missed, 1=made])

Independent Variables	β (SE)	χ^2	*p*	OR (95% CI)
(Intercept)	-0.4072 (0.0304)	178.8906	<0.0001	0.6655 (0.6269-0.7064)
Distance	-0.0179 (0.0076)	5.5272	0.0187	0.9822 (0.9677-0.997)
Forward	0.2179 (0.0227)	92.4098	<0.0001	1.2434 (1.1894-1.2999)
Center	0.3085 (0.0464)	44.1560	<0.0001	1.3615 (1.2433-1.4915)
Hook shot	0.9517 (0.1062)	80.3712	<0.0001	2.5901 (2.1048-3.1912)
Tip shot	1.1376 (0.1043)	118.9626	<0.0001	3.1192 (2.5458-3.8322)
Lay up	1.53 (0.0333)	2113.5167	<0.0001	4.6181 (4.3267-4.9296)
Dunk	3.1482 (0.0787)	1599.5200	<0.0001	23.2952 (19.9842-27.2119)
Angle	-0.0011 (0.0003)	20.5028	<0.0001	0.9989 (0.9984-0.9993)
Team Rank	-0.0006 (0.0001)	43.5336	<0.0001	0.9994 (0.9993-0.9996)
Distance*Forward	-0.0386 (0.0088)	19.0707	<0.0001	0.9451 (0.931-0.9593)
Distance*Center	-0.0383 (0.0203)	3.5495	0.0596	0.9454 (0.8816-1.0137)
Distance* Hook shot	-0.3245 (0.0479)	45.8600	<0.0001	0.7101 (0.6995-0.7208)
Distance* Tip shot	-0.3801 (0.0971)	15.3194	<0.0001	0.6716 (0.6609-0.6825)
Distance* Lay up	-0.7371 (0.0173)	1815.6121	<0.0001	0.47 (0.463-0.4771)
Distance* Dunk	-0.7315 (0.0671)	118.8100	<0.0001	0.4726 (0.4656-0.4797)
Model Performance		χ^2	*p*	*df*
Likelihood ratio test	-	13,086.0000	<0.0001	15
Wald Test	-	10,053.3000	<0.0001	15
Log Likelihood	-74,1999.4300	-	-	-
McFadden Pseudo R^2	0.0810	-	-	-
Akaike Information Criterion	148,431.0000	-	-	-

Note: R Programming code: [glm(formula = Shot made ~ Distance * Shot Type + Distance * Position + Angle + Team Rank, family = binomial(link=logit))]. 95% CI is 95% confidence interval, β is the unstandardized beta coefficient, OR is the odds ratio, SE is the standard error, χ^2 is the Wald's Chi-Square. Statistical significance accepted at <0.05. All statistics reported herein use 4 decimal places in order to maintain statistical precision.

Table 4. Results relating to the binomial logistic regression model 3 for all three point shot attempts (dependent variable = shot made [0=missed, 1=made])

Independent Variables	β (SE)	χ^2	*p*	OR (95% CI)
(Intercept)	0.3905 (0.1416)	7.6066	0.0058	1.4777 (1.1201-1.9514)
Distance	-0.1317 (0.0185)	50.4100	<0.0001	0.8766 (0.8453-0.9090)
Forward	-0.0778 (0.0191)	16.5893	<0.0001	0.9252 (0.8911-0.9604)
Center	-0.1167 (0.0798)	2.1374	0.1438	0.8899 (0.7599-1.0393)
Angle	0.0013 (0.0003)	16.3458	<0.0001	1.0013 (1.0007-1.0020)
Team Rank	-0.0008 (0.0001)	48.9580	<0.0001	0.9992 (0.999-0.9994)
Model Performance		χ^2	*p*	*df*
Likelihood ratio test	-	180.1400	<0.0001	5
Wald Test	-	178.4100	<0.0001	5
Log Likelihood	-44,725.7400	-	-	-
McFadden Pseudo R^2	0.0020	-	-	-
Akaike Information Criterion	89,463.0000	-	-	-

Note: R Programming code: [glm(formula = Shot made ~ Distance + Angle + Position + Team Rank, family = binomial(link=logit))]. 95% CI is 95% confidence interval, β is the unstandardized beta coefficient, OR is the odds ratio, SE is the standard error, χ^2 is the Wald's Chi-Square. Statistical significance accepted at <0.05. All statistics reported herein use 4 decimal places in order to maintain statistical precision.

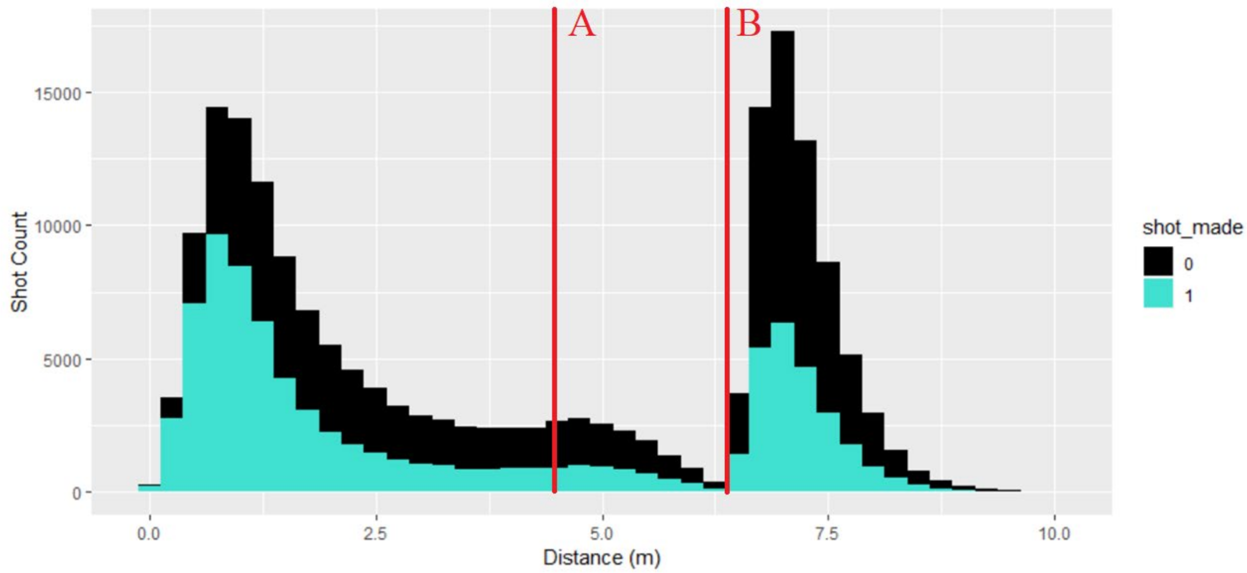

Figure 1. Distribution of shots across varying distances. Red vertical lines indicate the location of the free-throw line (A) and three point line (B).

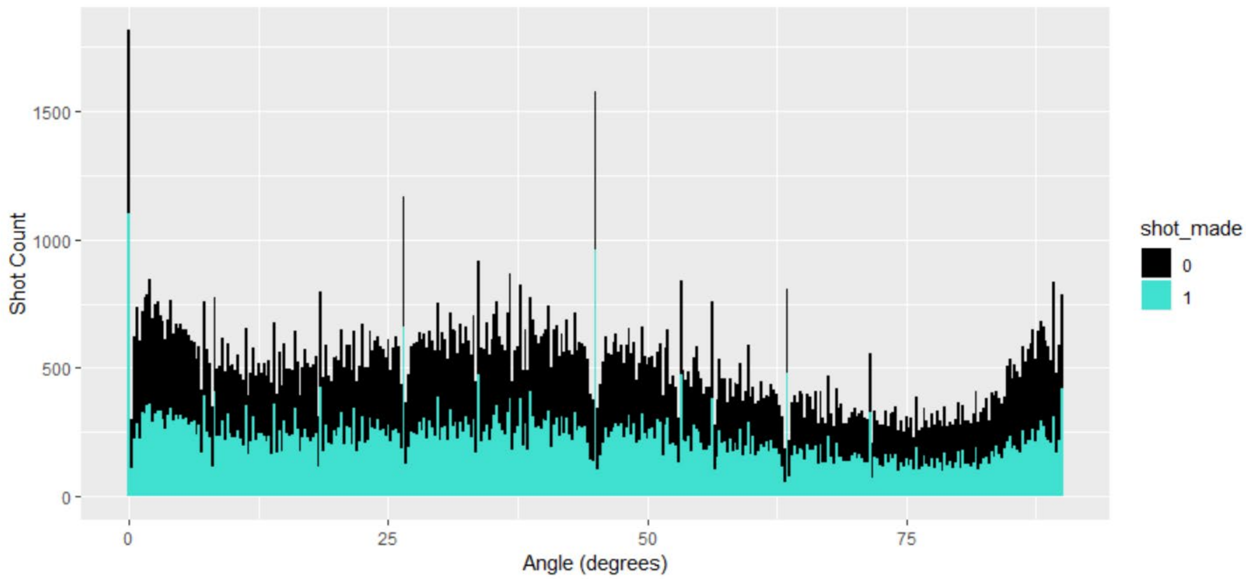

Figure 2. Distribution of shots across varying angles. A shot from zero degrees is at the midline and a shot from ninety degrees is near the baseline.

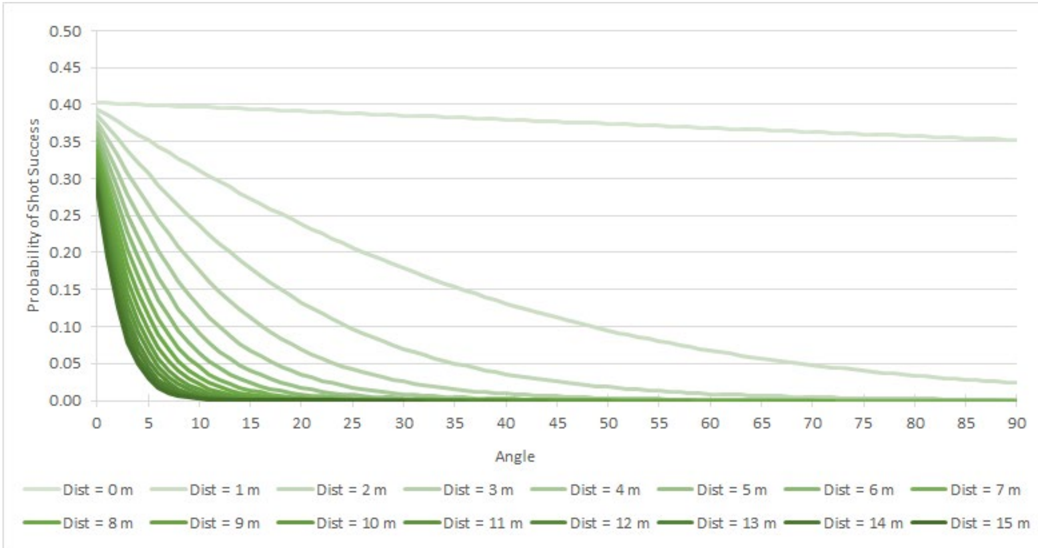

Figure 3. Distribution of shot probabilities corresponding to various shot distances (0-15 meters) across a range of shot angles. The figure above represents a guard utilizing a jump shot.

CHAPTER II: LITERATURE REVIEW

Basketball

General Background

The game of basketball was invented by James Naismith at Springfield College in Massachusetts in 1891 (Rains 2011). Naismith wrote down thirteen rules, hammered two peach baskets to the gymnasium balcony, and instructed his class to play nine versus nine with a soccer ball (Horger 2001; Naismith 1996). The object of the game was to throw the soccer style ball overhead and land it in the basket. Less than a year later in 1892, the bottoms of the peach baskets were removed to allow the ball to fall through the basket and the first public basketball game was played in Springfield, Massachusetts (Naismith 1996). In 1906, metal hoops, nets, and backboards were introduced along with the first version of a basketball (Rains 2011). Over the next four decades, the game rapidly increased in popularity and the game's rules were slowly adapting toward many of the modern rules in the game today. A series of small-scale professional and college leagues were formed early on, but basketball became firmly established at the collegiate level when the National Collegiate Athletic Association (NCAA) officially formed in 1939 as well as at the professional level when the National Basketball Association (NBA) took shape in 1946 (Horger 2001). A well-established women's professional league known as the Women's National Basketball Association (WNBA) was later founded in 1996 (Horger 2001). During these critical years of growth for the sport, the game was also growing at the international level, which eventually led to the first international organization known as the International Basketball Federation (FIBA) (Naismith 1996). The formation of FIBA allowed the game of basketball to officially be introduced to the Olympics in 1936, where the USA took the Gold medal in Berlin (Rains 2011).

Today, the game is played as five versus five where each of the five players are listed as one of three commonly known positions: guard, forward, or center (Erčulj and Štrumbelj 2015). However, these general positional categories can be broken down into five specific positions: point guard, shooting guard, small forward, power forward, and center (see Figure 4) (Naismith 1996). The game is played on a rectangular court and the winner is the team that scores the most points in their respective baskets. The court size for NBA, Women's National Basketball Association (WNBA), and NCAA is 28.65 meters in length and 15.24 meters in width, while FIBA is 28 meters and 15 meters, respectively. Two points are scored when a player shoots and makes a basket anywhere inside the three point line and three points are scored when a player shoots and makes a basket anywhere beyond the three point line. Either of these types of made baskets are known as a successful field goal (FG). The three point line for NCAA is on a radius of 6.32 meters from the basket as opposed to 6.75 meters (6.60 meters in the corners) for WNBA and FIBA. The three point line is farthest from the basket in the NBA at 7.24 meters (6.70 meters in the corners). Single points can also be scored as a result of a made free throw, which occurs after a player is fouled by an opposing player while attempting to shoot the ball. The number of free throws granted are determined by the location of the shot attempt and each free throw shot is taken at the free throw line, which is centered and fifteen feet from the basket at all levels. Free throws may also be granted after a non-shooting foul; this occurs when the fouling team has committed their seventh foul of the half and for each subsequent foul within the half. The ball is

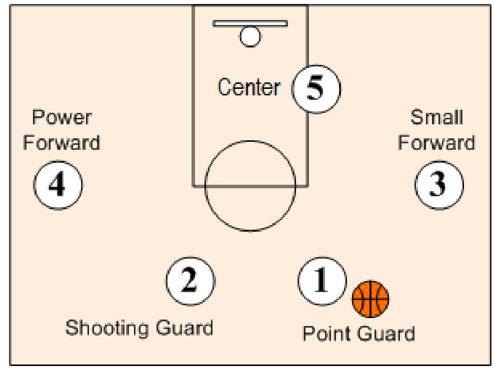

Figure 4. Basketball Player Positions. Image licensed by CC BY-SA 3.0 available from Wikimedia Commons (https://creativecommons.org/licenses/by-sa/3.0/),

advanced up and around the court by either dribbling or passing. Games are played in the format of four, twelve minute quarters in the NBA and four, ten minute quarters in the WNBA. At the college level, NCAA Men's games are played in the format of two, twenty minute halves and in the format of four, ten minute quarters for NCAA Women's games.

NCAA Basketball

The NCAA is the pinnacle of collegiate basketball in America and can further be categorized into three levels: Division I (DI), Division II, and Division III. The three divisions boast a range of talent, in which the DI level typically contains the most sought after recruits from high school who are expected to be the most talented players. At the DI level, every team belongs to a conference, which is not the case in Divisions II and III where some independent institutions exist. An NCAA basketball season usually begins with preseason games in October with the regular season starting soon after in November. The regular season generally runs from November to early March. At this point in time, teams that performed well enough to make their conference tournament will face off for their respective conference championships. After all conference championships are decided, a committee from the NCAA selects sixty-eight teams for each Division to compete in the NCAA Tournament. The eventual winner of this tournament is crowned as the NCAA National Champion for their respective Division.

Measurement of Basketball Performance

Key Performance Indicators (KPI) are defined as a selection or combination of action variables that aim to define specific aspects of performance, which most closely relate to a successful outcome (Hughes and Bartlett 2002; Ortega, Villarejo, and Palao 2009). KPI's are most commonly used to assess the performance of a team or an individual within that team, but they can also be used from a comparative perspective with opponents or other athletes and teams (Hughes and Bartlett 2002). Presenting KPI's in isolation can be deceiving and result in a misleading, distorted perception of a team or individual's performance (Hughes and Bartlett 2002; Mikolajec, Maszczyk, and Zając 2013). However, even when presented in combination, it is difficult to identify precisely why a team did or did not produce a successful outcome (García et al. 2013; Hughes and Bartlett 2002; Mikolajec, Maszczyk, and Zając 2013).

Reported across men's and women's continental championships (Madarame 2018a, 2018b), NBA (Mikolajec, Maszczyk, and Zając 2013; Pai, ChangLiao, and Lin 2017; Zimmermann 2016), NCAA (Zimmermann 2016), Olympic level (Leicht, Gomez, and Woods 2017), and the Spanish Basketball League (García et al. 2013), the most common basketball KPIs contribute to numerous aspects of performance such as scoring, offense, and defense. However, in the early years of the sport, the most straightforward way to describe a team in regards to their success was scoring (Zimmermann 2016). As a result, the first measurable variables noted as KPI's mainly revolved around scoring points; either scoring points on offense or preventing the opponent from scoring (Zimmermann 2016). These KPI's are still a part of what are largely accepted today and include: field goals made, three-point field goals made, free throws made, offensive rebounds, turnovers, defensive rebounds, steals, blocks, points scored per game, and points allowed per game (García et al. 2013; Leicht, Gomez, and Woods 2017; Madarame 2018a, 2018b; Mikolajec, Maszczyk, and Zając 2013; Pai, ChangLiao, and Lin 2017; Zimmermann 2016).

The previously mentioned variables remain as valuable KPI's, but there are limits to their expressiveness since they are raw numbers (Zimmermann 2016). For example, knowing that a team collected twenty rebounds in a game makes it difficult to consider the value for this variable, good or poor, considering that we do not know how many total rebounds there were to be had. As a result of this limited knowledge, many of today's most common basketball KPI's are presented in the form of either “rate”, “percentage”, or “efficiency” (Mikolajec, Maszczyk, and Zając 2013; Zimmermann 2016). Rate is considered to be a fixed ratio between two things, while percentage is a part of a whole expressed in hundredths, and efficiency refers to producing desired results with little or no waste.

The combination of widely accepted basketball KPI’s today influence major aspects of today’s game such as the scoring aspect and technical aspect. The KPI’s can be broken down into two categories: offensive and defensive. The KPI’s that belong to the offensive category primarily revolve around the team’s shot locations, shot types, and their ability to score points. A list of these KPI’s and their abbreviations are shown below in Table 5.

Table 5. A summary table of the most commonly used offensive basketball key performance indicators

Key Performance Indicator (KPI)	**Abbreviation**
Field Goals Attempted	FGA
Field Goals Made	FGM
Field Goal Percentage	FG%
Two point Field Goals Attempted	2PA
Two point Field Goals Made	2FGM
Two point Field Goal Percentage	2P%
Two point Field Goal Attempt Rate	2PA Rate
Three point Field Goals Attempted	3PA
Three point Field Goals Made	3FGM
Three point Field Goal Percentage	3P%
Three point Field Goal Attempt Rate	3PA Rate
Free Throws Attempted	FTA
Free Throws Made	FTM
Free Throw Percentage	FT%
Free Throw Rate	FT Rate
Point Difference*	PD
Points Scored	PTS
Effective Field Goal Percentage	eFG%
Offensive Efficiency	OE
Adjusted Offensive Efficiency	AdjOE
Turnovers	TO
Turnover Rate	TO Rate
Turnover Percentage	TO%
Offensive Rating	ORtg
Team Ball Possessions	TBP
Assist Percentage	AST%
Assist Turnover Ratio	AST/TO
Assists	AST
Offensive Rebounds	OR
Offensive Rebounding Rate	ORR
Offensive Rebounding Percentage	OR%
Total Rebounds*	REB

KPI relevant to both offensive and defensive categories

The KPI's that belong to the defensive category primarily reflect to the team's ability to thwart scoring chances. These KPI's are shown below in Table 6.

Table 6. A summary table of the most commonly used defensive basketball key performance indicators

Key Performance Indicator (KPI)	Abbreviation
Defensive Efficiency	DE
Adjusted Defensive Efficiency	AdjDE
Defensive Rating	DRtg
Fouls Committed	Fouls
Points allowed	PTSA
Point Difference*	PD
Steals	STL
Blocks	BLK
Defensive Rebounds	DR
Defensive Rebounding Percentage	DR%
Defensive Rebounding Rate	DRR
Total Rebounds*	REB
***KPI relevant to both offensive and defensive categories**	

Shot Location and Type

Among the most fundamental decisions to make in a basketball game are where and how to take a shot. These decisions impact a myriad of offensive KPIs listed in Table 5, most notably field goals made, field goal percentage, and points scored. Shooting locations and shot types are chosen in a fast-paced manner by players during each team's offensive possessions during the game. There are several different locations to shoot from and types of shots to utilize; the choice of which is influenced by several factors such as distance from the basket, player type, and player skills (Erčulj and Štrumbelj 2015).

Types of shots can be separated into five common categories: jump shot, layup, tip in, dunk, and hook shot (Erčulj and Štrumbelj 2015). Additional details about shot type known as the subtype of a shot exist; common shot subtypes are fade away, floating, pull up, turn around, alley-oop, driving, finger roll, put back, and reverse. Jump shots followed by lay ups are known to be the most common types of shots seen across different levels of competitive basketball (Erčulj and Štrumbelj 2015). Given that player types have different roles in basketball (Dežman, Trninić, and Dizdar 2001), motor skills (Erčulj et al. 2009), and anthropometric dimensions (Erčulj and Štrumbelj 2015; Sampaio et al. 2006), differences in basketball shooting tendencies are expected.

Analytics in Sport

Notational analysis is a technique used to produce a permanent record of events, which can later be analyzed and used to provide feedback (Carling, Reilly, and Williams 2008). The use of notational analysis to examine performance in team sports is well-established in previous literature (Franks and Hughes 2004; O'Donoghue 2009; Ortega, Villarejo, and Palao 2009) and is utilized to inform the training process (Hughes and Bartlett 2002; Ortega, Villarejo, and Palao 2009). Early forms of notational analysis were done by hand, usually by an assistant coach, and was shown to be an easy, adaptable method (Carling, Reilly, and Williams 2008; Franks and Hughes 2004). However, this method is extremely time consuming and requires constant attention from organization staff. In an effort to reduce time spent recording events during games as well as human error, the utilization of video footage became a staple for notational analysis (Carling, Reilly, and Williams 2008). While this allows the viewing of important in-game events after the fact, it still requires a person to physically tally the instances of these events making human error remain an inevitable factor (Carling, Reilly, and Williams 2008; Franks and Hughes 2004). Computerized notational analysis was next to emerge, which raised the bar from the rudimentary use of video footage (Carling, Reilly, and Williams 2008). This type of notational analysis allows the user to simply click a button on a computer or tablet to identify the occurrence of an event, simultaneously updating event totals, and continuously building the game's timeline of events in real time (Carling, Reilly, and Williams 2008; Hughes and Bartlett 2002). This method requires costly equipment and software, but the ability to view the timeline of events as they happen in real time gives coaches an opportunity to quickly adjust strategy and feedback (Franks and Hughes 2004).

Outside of a few examples of collecting statistics to advise baseball strategies in the 1960's and 70's (Lindsey 1963; Rees, Rakes, and Deane 2015), notational analysis has been historically undervalued. Most early sport analysis and game predictions were qualitative; based on sports commentators, former players, or coaches using instinct, past experience, or gut-feelings (Cao 2012; Gerrard 2016; Leung and Joseph 2014; Mondello and Kamke 2014; Rees, Rakes, and Deane 2015). These analyses were often delivered and heavily discussed before televised sporting events and were considered "expert predictions" (Cao 2012; Leung and Joseph 2014). However, built on a foundation of anecdotal evidence, bias, and subjectivity, the accuracy of these claims were highly variable (Cao 2012; Leung and Joseph 2014; Mondello and Kamke 2014; Rees, Rakes, and Deane 2015). The use of quantitative methods to inform sport strategies, decision-making, and make game predictions was long an unimagined thought from the general population until the release of the best-selling book (2003) and popular movie (2011), *Moneyball* (Lewis 2004). While it is not the earliest account of applying analytics to sport, it is largely credited as the catalyst for introducing every day sports fans and the broader sports community to the array of potential benefits of the use of quantitative methods (Cao 2012; Fry and Ohlmann 2012).

With advancements in notational analysis and a peaking interest in using quantitative methods to improve team success, considerable amounts of data have become available to sporting organizations in the professional and college ranks, subsequently triggering explosive growth in the field of sports analytics. (Cao 2012; Gerrard 2016; Leung and Joseph 2014; McCullagh 2010). Consequently, this access to large amounts of data at a rapid rate has highlighted our ability to suddenly collect and maintain data as well as our inability to quickly turn it into useful information (Haghighat, Rastegari, and Nourafza 2013; Padhy, Mishra, and Panigrahi 2012). Thus, a multitude of advanced methods utilizing quantitative analysis to advise in game strategies, inform personnel needs, or make predictions now exist.

Advanced Quantitative Methods

Machine learning (ML) is an application of Artificial Intelligence that provides systems the ability to learn from data, identify patterns, and make decisions with minimal human intervention (Bunker and Thabtah 2017; Freitag 2000). It has shown promise in the domains of classification and prediction, particularly in sport (Bunker and Thabtah 2017). This is highlighted in previous literature for an array of sports including weight training/lifting (Novatchkov and Baca 2013), running (Kugler et al. 2011), golf (Eskofier et al. 2011), soccer (Buursma 2011; Faria et al. 2010; Hucaljuk and Rakipović 2011; Min et al. 2008), basketball (Ángel Gómez et al. 2008; Mikolajec, Maszczyk, and Zając 2013; Parejo et al. 2013; Zimmermann, Moorthy, and Shi 2013), and baseball (Smith, Lipscomb, and Simkins 2007). Machine learning methods are either supervised or unsupervised (see Figure 5).

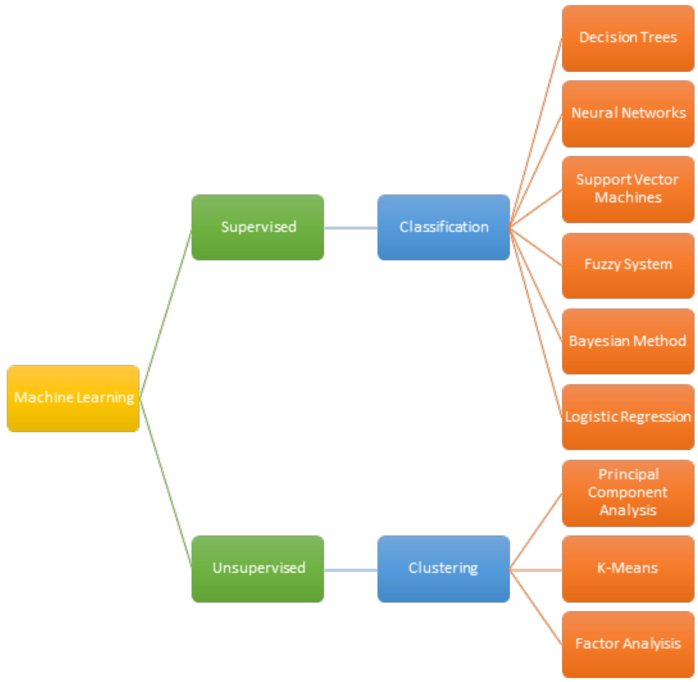

Figure 5. Methods Used for Result Prediction in Sport.

In supervised learning, a predictive model is developed based on both input and output data whereas in unsupervised learning, data are grouped and based only on input data (Bunker and Thabtah 2017).

<u>Supervised Learning Methods</u>

Supervised learning methods fall under the general umbrella of classification, using training data and test data to predict a target variable (Bunker and Thabtah 2017). Common methods that fall under this classification include decision trees, neural networks, support vector machines, fuzzy systems, the Bayesian method, and logistic regressions (see Figure 5 above).

Neural networks consist of a number of interconnected neurons within specific layers that constantly adjust weights, that contribute to the final prediction, and are one of the most commonly used ML approaches to sport prediction problems (Bunker and Thabtah 2017; Haghighat, Rastegari, and Nourafza 2013). The weights associated with the interconnected neurons constantly change to accomplish higher levels of predictive accuracy, which can sometimes lead to overfitting or wasting of computing resources (Bunker and Thabtah 2017; Mohammad, Thabtah, and McCluskey 2014). However, neural networks are appealing due to their flexibility when defining classification variables (Bunker and Thabtah 2017). Neural networks were used for the prediction of results from the National Football League in the late 1990's and early 2000's (Kahn 2003; Purucker 1996), which are examples of very early ML techniques used in professional football.

While the path to final predictions are hidden in neural networks, decision trees result in a set of rules which clarify the final result (Mariscal, Marbán, and Fernández 2010). A decision tree first poses questions about certain features of the data and then classifies it appropriately. Each of these questions asked are subsets of a node; and each of the interior nodes then direct to a child node, which is simply a possible answer to the posed question (Haghighat, Rastegari, and Nourafza 2013). This creates a "tree" shape from the top node down to each leaf, which is considered any node that is childless; i.e. has no further connections (Haghighat, Rastegari, and Nourafza 2013). The main advantages of decision trees are that they are computationally fast, can handle large amounts of data on different measurement scales, and they make no statistical assumptions (Pal and Mather 2003). Other advantages of decision trees are that software to develop them is readily available over the internet; and they are generally considered to be easier to interpret than some other ML algorithms (Friedl and Brodley 1997; Pal and Mather 2003). However, disadvantages of decision trees are overfitting and sampling errors (Rokach and Maimon 2005; Srivastava et al. 2002), which can lead to less than satisfactory results. Decision trees have previously been used to inform optimal end-game strategies in basketball (Annis

2006) and one-on-one attacking and defender interactions in hockey (Morgan, Williams, and Barnes 2013).

Another well-established supervised learning method is a support vector machine. A support vector machine can be used for both classification and regression purposes and is formally defined by a separating hyperplane (Haghighat, Rastegari, and Nourafza 2013). A hyperplane separates two groups of points and is at equal distance from the two. The algorithm searches for the optimal separating hyperplane, which then acts as a decision boundary between the two classes that are being examined (Haghighat, Rastegari, and Nourafza 2013). The main advantages of using support vector machines, are that they are not prone to overfitting, as well as their impressive capability to produce a complex, non-linear decision boundary (Guyon et al. 2002; Haghighat, Rastegari, and Nourafza 2013). They are considered advantageous to neural networks due to the absence of spurious local minima within the optimization procedure as well as the fact that there are very few parameters to tune/adjust (Hearst et al. 1998). Major disadvantages of support vector machines include lengthy training time for large-scale problems and the difficulty of implementation (Haghighat, Rastegari, and Nourafza 2013; Platt 1999). Support vector machines have previously aided in basketball outcome prediction (Cao 2012).

Fuzzy system, or fuzzy logic, was developed for systems and problems that require complicated mathematical analyses (Haghighat, Rastegari, and Nourafza 2013). While variables from traditional binary sets either have a value of zero or one, fuzzy logic can find truth values between zero and one as well as accurately describe complex, irrational phenomena (Tavana et al. 2013). Its ability to aid in identifying indefinite and complex phenomena is a massive upside, but training complexity and the need to finely tune a large number of parameters are major drawbacks (Liang and Mendel 2000). Recently, fuzzy logic has been heavily used in the examination of cricket player performance (Curtis, Kelly, and Craven 2009; Curtis 2010; Singh, Bhatia, and Singh 2011)

The Bayesian Method is one of the most famous supervised ML classification techniques (Haghighat, Rastegari, and Nourafza 2013). Major advantages of the Bayesian Method are that it is relatively simple and works well on data with high levels of noise or various unrelated features (Haghighat, Rastegari, and Nourafza 2013). This is because it is a probabilistic prediction model, which assumes that all features are conditionally independent of the target variable (Haghighat, Rastegari, and Nourafza 2013). However, the Bayesian method does have significant drawbacks: many of the current approaches are needlessly data-inefficient and they do not take advantage of small-scale properties of differentiable functions near local optima (Lizotte 2008). Nevertheless, this method has been previously implemented to predict future winners in NBA games (Miljković et al. 2010).

Another very well-known tool for classification problems is Logistic Regression. It is similar to linear regression in that it depends on a linear combination of features, which are eventually mapped to a certain value between zero and one (Ye 2003). First, the odds of characteristics of each group are estimated and then cut-off points are determined, which results in the appropriate categorization of certain features (Haghighat, Rastegari, and Nourafza 2013). Logistic regressions are advantageous due to the nature of their simple calculations and interpretations, which generally produce reliable results (Kantardzic 2011). However, overestimation and difficulty in predictor selection for the model present as drawbacks (Steyerberg et al. 2000; Van Houwelingen and Le Cessie 1990). Logistic regressions have been used widely to aid in the answering of questions in the realm of sport, especially in the prediction of soccer matches (Buursma 2011).

Unsupervised Learning Methods

Unsupervised learning methods fall under the general category of clustering (see Figure 5 above). The aim of clustering is to quickly pass through (i.e. assess) data and gain first order knowledge by partitioning data points into like groups, or clusters (Ding and He 2004). Although there are an abundance of methods used in cluster analysis, a few common methods are K-means, Principal component analysis (PCA), and Factor Analysis.

K-means is one of the most common methods used in clustering, and is appreciated for its simplicity and relative efficiency (Ding and He 2004). In this method, centroids are used to represent clusters through the optimization of the squared error function (Ding and He 2004). The number of clusters (K) must be known beforehand and supplied as a parameter, which is seen as a disadvantage for this method (Pena, Lozano, and Larranaga 1999; Ray and Turi 1999). K-means cluster analysis has previously been used to classify game pace in Olympic level basketball (Sampaio, Lago, and Drinkwater 2010).

Although there are distinct differences, PCA and Factor Analysis are often discussed in the same breath. This is because they are mainly used as dimension-reduction procedures, which means that they can identify a small group of variables (often called factors) that explain most of the total (PCA) or common (Factor Analysis) variation from the full set of original variables (Bryant and Yarnold 1995). However, despite this similarity, there is a fundamental difference between these two methods of data-reduction: Factor analysis is a measurement model of a latent, or inferred variable, while PCA is instead, a linear combination of variables (Anderson 1962; Bryant and Yarnold 1995). Talent identification in sport has been previously studied using PCA (Douda et al. 2008) and Factor Analysis (Morris 2000; Verma 2016) methods.

Spatial Analysis

In broad terms, spatial analysis refers to the quantitative study of phenomena that are located in space (Bailey and Gatrell 1995). In the context of sports analytics, space typically refers to the playing area, i.e. a basketball court or playing field. Spatial analysis techniques provide a layer of depth to conventional sport evaluation methods by revealing key space-based variations in player and team performances. A wide range of techniques exist; some of which utilize spatial coordinates to determine locations or the use of Voronoi diagrams to allow for spatial partitioning of an area into cells with specific associations (Fonseca et al. 2012). Spatial analysis has been used in basketball to determine players with the best shooting range (Goldsberry 2012) and in soccer to analyze the location and outcomes of direct free kicks at the World Cup level (Alcock 2010). This type of analysis has also been used to explore serving locations in tennis (Hizan, Whipp, and Reid 2015) as well as shooting performance in futsal (Vilar et al. 2013).

Basketball-Specific Methods

Examinations of KPI's and game success in the NBA is well-established in previous literature (Mikolajec, Maszczyk, and Zając 2013; Miljković et al. 2010; Teramoto and Cross 2010; Zimmermann 2016). Teramoto et al. (2010) used logistic regression analysis to determine that effective field goal percentage and turnovers were most critical for NBA game success. Miljković et al. (2010) examined chances of winning in the NBA, as the home or away team, by way of the Bayesian Method and noted that the model predicted 67% of 778 games correctly. Mikołajec, Maszczyk, & Zając (2013) employed the Factor Analysis method and reported that fouls committed, steals, and offensive efficiency are most critical to the final result and the team's rank. Parejo et al. (2013) used K-means cluster analysis to determine that assists, steals, total rebounds, blocks, and fouls received are the most significant contributors to the final score for amateur teams in a Spanish league. However, Garcia et al. (2013) found conflicting results for professional teams in the Spanish league using the same cluster analysis, noting that defensive rebounds, two point field goals made, and three point field goals made were most critical to the final score. Gomez et al. (2008) also examined the Spanish league using K-means and were in agreement with Garcia et al. (2013), reporting two and three point field goals made were critical factors. However, Gomez et al. (2008) also noted that assists, steals, and turnovers as important factors, which is in agreement with Parejo et al. (2013), on amateur teams in Spain. Trawinski (2010) used fuzzy logic in an attempt to predict basketball game outcomes in the Spanish league, but were unable to predict game outcomes with high accuracy. Ivanković et al. (2010) used a neural network to examine factors associated with winning in the First B basketball league for men in Serbia, reporting defensive rebounds and two point field goals made were the most influential.

NCAA Men's basketball has been studied (Akers, Wolff, and Buttross 1992; Conte et al. 2018), but less so than at the professional level. Akers, Wolff, & Buttross (1992) investigated factors that are important for winning in DI Men's basketball by way of regression analyses, reporting that two point field goal percentage, turnovers, free throw percentage, steals, and rebounds were the most critical. Conte et al. (2018) is in agreement, but also reported defensive rebounds, free throws attempted, free throw rate, effective field goal percentage, and offensive rating as important factors.

The previous literature highlights many shooting-related variables as key performance indicators in basketball, indicating the importance of shot type and shot location in game outcome. While the literature based around shot location and type in basketball is not overly abundant, most of which is established already, revolves around the professional level.

In regards to shot location at the professional level, Harmon, Lucey, and Klabjan (2016) determined that the probability of making a shot decreases as distance away from the basket increases. A second study reported that, across many competition levels, more successful teams on average attempt fewer three point field goals (Erčulj and Štrumbelj 2015). With consideration to position, guards tend to play farther from the basket and also shoot more often from distance, while centers tend to play closer to the basket and are more likely to perform a dunk or tip in shot (Harmon, Lucey, and Klabjan 2016; Miller and Bartlett 1996). The work of Harmon, Lucey, & Klabjan (2016) built off the above statement, noting that NBA centers tend to have the highest shooting percentage, given that many of their shot attempts are close to the basket.

Conclusion

With an explosion in the field of sports analytics, advanced quantitative methods have been utilized to analyze basketball performance. As a result, key performance indicators across competition levels are well-established. Many of the key performance indicators are shooting-related, highlighting the importance of shot location and type in regards to game outcome. Limited research has been published on basketball shot location and type; much of which exists at the professional level. Further exploration of shot location and shot types would allow for a deeper understanding of key performance variables and the ability to communicate the way in which they shape game outcomes.

CHAPTER III: SUMMARY AND CONCLUSIONS

The primary purpose of the current study was to investigate the effect of court location and shot types used on made field goals in NCAA Men's DI basketball during the 2017-2018 season. A secondary purpose of the current study was to further expand the analysis based on two additional factors: player position (guard, forward, or center) and team ranking at the end of the regular season. The results of the current study demonstrate that the probability of shot success in NCAA Men's DI basketball was significantly influenced by shot type, court location (distance and angle), position, and team rank. Although these variables were statistically significant, they may not be practically significant to implement in a real-world team sports environment, as the magnitude of the effect observed in some variables (i.e. team rank and angle) were minimal.

Overall, the results of the current study indicated that guards were the most successful shooters from distance, most often utilizing a jump shot, which had the highest probability of shot success from distance when compared to all other shot types. When looking at both two and three point shot attempts, the results of the current study indicated that the probability of shot success decreased with distance for all shot types when compared to a jump shot. This emphasizes the increased difficulty of shot making as players move farther away from the basket.

When looking only at three point shots, results of this model indicated that for every one-meter increase in distance behind the three point line, a player is 12.34% less likely to make that three point shot. Given that the NCAA three point line will be moving back from 6.32 meters to 6.75 meters in the upcoming season, the previous result suggests that a decrease in shot success from behind the three point line may be seen in the upcoming season. The results of this model also indicate that, when considering position, forwards may struggle in the upcoming year behind the arc more than guards. From a practical standpoint, this information is valuable as it may suggest that shooting a high percentage from behind the three point line in the coming years may require players to regularly practice this shot to get used to shooting from farther distances.

The scope of the current study provides meaningful knowledge to coaches, enabling them to gain a better understanding of which shot types and locations are utilized successfully in competition. As a result, this allows them to optimize the basketball training process by focusing on selected techniques in their limited time at practice. Future research investigating factors that influence the probability of shot success should incorporate temporal and tactical aspects of the game such as time left on the shot clock and defensive ball pressure, which may have the potential to further enhance the understanding of successful shooting at the collegiate level.

REFERENCES

Akers, Michael D., Shaheen Wolff, and Thomas E. Buttross. 1992. "An Empirical Examination of the Factors Affecting the Success of NCAA Division I College Basketball Teams." *Journal of Business and Economic Studies*.

Alcock, Alison. 2010. "Analysis of Direct Free Kicks in the Women's Football World Cup 2007." *European Journal of Sport Science* 10 (4): 279–284.

Anderson, Theodore Wilbur. 1962. "An Introduction to Multivariate Statistical Analysis." Wiley New York.

Ángel Gómez, Miguel, Alberto Lorenzo, Jaime Sampaio, Sergio José Ibáñez, and Enrique Ortega. 2008. "Game-Related Statistics That Discriminated Winning and Losing Teams from the Spanish Men's Professional Basketball Teams." *Collegium Antropologicum* 32 (2): 451–456.

Annis, David H. 2006. "Optimal End-Game Strategy in Basketball." *Journal of Quantitative Analysis in Sports* 2 (2). https://doi.org/10.2202/1559-0410.1030.

Bailey, Trevor C., and Anthony C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Vol. 413. Longman Scientific & Technical Essex.

Bryant, Fred B., and Paul R. Yarnold. 1995. "Principal-Components Analysis and Exploratory and Confirmatory Factor Analysis."

Bunker, Rory P., and Fadi Thabtah. 2017. "A Machine Learning Framework for Sport Result Prediction." *Applied Computing and Informatics*.

Buursma, D. 2011. "Predicting Sports Events from Past Results Towards Effective Betting on Football Matches." In *Conference Paper, Presented at 14th Twente Student Conference on IT, Twente, Holland*. Vol. 21.

Cao, Chenjie. 2012. "Sports Data Mining Technology Used in Basketball Outcome Prediction."

Carling, Christopher, Thomas Reilly, and A. Mark Williams. 2008. *Performance Assessment for Field Sports*. Routledge.

Conte, Daniele, Antonio Tessitore, Aaron Gjullin, Dominik Mackinnon, Corrado Lupo, and Terence Favero. 2018. "Investigating the Game-Related Statistics and Tactical Profile in NCAA Division I Men's Basketball Games." *Biology of Sport* 35 (2): 137–43. https://doi.org/10.5114/biolsport.2018.71602.

Csapo, Peter, and Markus Raab. 2014. "'Hand down, Man down.' Analysis of Defensive Adjustments in Response to the Hot Hand in Basketball Using Novel Defense Metrics." *PloS One* 9 (12): e114184.

Curtis, K. M. 2010. "Cricket Batting Technique Analyser/Trainer: A Proposed Solution Using Fuzzy Set Theory to Aid West Indies Cricket." In *Proceedings of the 9th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, 71–76. World Scientific and Engineering Academy and Society (WSEAS).

Curtis, K. M., M. Kelly, and M. P. Craven. 2009. "Cricket Batting Technique Analyser/Trainer Using Fuzzy Logic." In *2009 16th International Conference on Digital Signal Processing*, 1–6. Santorini, Greece: IEEE. https://doi.org/10.1109/ICDSP.2009.5201206.

Dežman, Brane, Slavko Trninić, and Dražan Dizdar. 2001. "Expert Model of Decision-Making System for Efficient Orientation of Basketball Players to Positions and Roles in the Game–Empirical Verification." *Collegium Antropologicum* 25 (1): 141–152.

Ding, Chris, and Xiaofeng He. 2004. “K-Means Clustering via Principal Component Analysis.” In *Proceedings of the Twenty-First International Conference on Machine Learning*, 29. ACM.

Douda, Helen T., Argyris G. Toubekis, Alexandra A. Avloniti, and Savvas P. Tokmakidis. 2008. “Physiological and Anthropometric Determinants of Rhythmic Gymnastics Performance.” *International Journal of Sports Physiology and Performance* 3 (1): 41–54.

Erčulj, Frane, Mateja Blas, Milan Čoh, and Mitja Bračič. 2009. “Differences in motor abilities of various types of european young elite female basketball players.” *Kinesiology* 41 (2).

Erčulj, Frane, and Erik Štrumbelj. 2015. “Basketball Shot Types and Shot Success in Different Levels of Competitive Basketball.” *PLOS ONE* 10 (6): e0128885. https://doi.org/10.1371/journal.pone.0128885.

Eskofier, B., Sandra Tuexen, P. Kugler, Ulf Jensen, and Ian Wright. 2011. “Development of Pattern Recognition Methods for Golf Swing Motion Analysis.” In *Proceedings of 8th International Symposium of the International Association of Computer Science in Sports*, 21–24.

Faria, Brigida Mónica, Luis Paulo Reis, Nuno Lau, and Gladys Castillo. 2010. “Machine Learning Algorithms Applied to the Classification of Robotic Soccer Formations and Opponent Teams.” In *Cybernetics and Intelligent Systems (CIS), 2010 IEEE Conference On*, 344–349. IEEE.

Fonseca, Sofia, João Milho, Bruno Travassos, and Duarte Araújo. 2012. “Spatial Dynamics of Team Sports Exposed by Voronoi Diagrams.” *Human Movement Science* 31 (6): 1652–1659.

Franks, Ian M., and Mike D. Hughes. 2004. *Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport*. Psychology Press.

Freitag, Dayne. 2000. “Machine Learning for Information Extraction in Informal Domains.” *Machine Learning* 39 (2–3): 169–202.

Friedl, M. A., and C. E. Brodley. 1997. “Decision Tree Classification of Land Cover from Remotely Sensed Data.” *Remote Sensing of Environment* 61 (3): 399–409. https://doi.org/10.1016/S0034-4257(97)00049-7.

Fry, Michael J., and Jeffrey W. Ohlmann. 2012. *Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications*. INFORMS.

García, Javier, Sergio J. Ibáñez, Raúl Martinez De Santos, Nuno Leite, and Jaime Sampaio. 2013. “Identifying Basketball Performance Indicators in Regular Season and Playoff Games.” *Journal of Human Kinetics* 36 (1): 161–168.

Gerrard, Bill. 2016. “Analytics, Technology and High Performance Sport.” *Critical Issues in Global Sport Management*, 205.

Goldsberry, Kirk. 2012. “Courtvision: New Visual and Spatial Analytics for the NBA.” In *2012 MIT Sloan Sports Analytics Conference*, 9:12–15.

Gómez, Miguel A., Alberto Lorenzo, Rubén Barakat, Enrique Ortega, and Palao José M. 2008. “Differences in Game-Related Statistics of Basketball Performance by Game Location for Men’s Winning and Losing Teams.” *Perceptual and Motor Skills* 106 (1): 43–50.

Gomez, Miguel Angel, Lorenzo Gasperi, and Corrado Lupo. 2016. “Performance Analysis of Game Dynamics during the 4th Game Quarter of NBA Close Games.” *International Journal of Performance Analysis in Sport* 16 (1): 249–263.

Guyon, Isabelle, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. “Gene Selection for Cancer Classification Using Support Vector Machines.” *Machine Learning* 46 (1): 389–422. https://doi.org/10.1023/A:1012487302797.

Haghighat, Maral, Hamid Rastegari, and Nasim Nourafza. 2013. “A Review of Data Mining Techniques for Result Prediction in Sports.” *Advances in Computer Science: An International Journal* 2 (5): 7–12.

Harmon, Mark, Patrick Lucey, and Diego Klabjan. 2016. “Predicting Shot Making in Basketball Using Convolutional Neural Networks Learnt from Adversarial Multiagent Trajectories.” *Stat* 1050: 15.

Hearst, M. A., S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. 1998. “Support Vector Machines.” *IEEE Intelligent Systems and Their Applications* 13 (4): 18–28. https://doi.org/10.1109/5254.708428.

Hizan, Hazuan, Peter Whipp, and Machar Reid. 2015. “Gender Differences in the Spatial Distributions of the Tennis Serve.” *International Journal of Sports Science & Coaching* 10 (1): 87–96.

Horger, Marc Thomas. 2001. “Play by the Rules: The Creation of Basketball and the Progressive Era, 1891-1917.” PhD Thesis, The Ohio State University.

Hucaljuk, Josip, and Alen Rakipović. 2011. “Predicting Football Scores Using Machine Learning Techniques.” In *MIPRO, 2011 Proceedings of the 34th International Convention*, 1623–1627. IEEE.

Hughes, and Roger M. Bartlett. 2002. “The Use of Performance Indicators in Performance Analysis.” *Journal of Sports Sciences* 20 (10): 739–754.

Ivanković, Z., M. Racković, Branko Markoski, D. Radosav, and M. Ivković. 2010. “Analysis of Basketball Games Using Neural Networks.” In *2010 11th International Symposium on Computational Intelligence and Informatics (CINTI)*, 251–256. IEEE.

Johnson, G. 2019. “Men’s College Basketball 3-Point Line Extended to International Distance.” Ncaa.org. June 5, 2019. https://www.ncaa.com/news/basketball-men/article/2019-06-05/mens-college-basketball-3-point-line-extended-international.

Kahn, Joshua. 2003. “Neural Network Prediction of NFL Football Games.” *World Wide Web Electronic Publication*, 9–15.

Kantardzic, Mehmed. 2011. *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons.

Kugler, P., Dominik Schuldhaus, Ulf Jensen, and B. Eskofier. 2011. “Mobile Recording System for Sport Applications.” In *Proceedings of the 8th International Symposium on Computer Science in Sport (IACSS 2011), Liverpool*, 67–70.

Leicht, Anthony S., Miguel A., and Carl T. Woods. 2017. “Team Performance Indicators Explain Outcome during Women’s Basketball Matches at the Olympic Games.” *Sports* 5 (4): 96.

Leung, Carson K., and Kyle W. Joseph. 2014. “Sports Data Mining: Predicting Results for the College Football Games.” *Procedia Computer Science* 35: 710–719.

Lewis, Michael. 2004. *Moneyball: The Art of Winning an Unfair Game*. WW Norton & Company.

Liang, Qilian, and J. M. Mendel. 2000. “Interval Type-2 Fuzzy Logic Systems: Theory and Design.” *IEEE Transactions on Fuzzy Systems* 8 (5): 535–50. https://doi.org/10.1109/91.873577.

Lindsey, George R. 1963. “An Investigation of Strategies in Baseball.” *Operations Research* 11 (4): 477–501.
Lizotte, Daniel James. 2008. *Practical Bayesian Optimization*. University of Alberta.
Madarame, Haruhiko. 2018a. “Basketball Game-Related Statistics That Discriminate among Continental Championships for Under-18 Women.” *Sports* 6 (4): 114.
Madarame, Haruhiko. 2018b. “Regional Differences in Women’s Basketball: A Comparison among Continental Championships.” *Sports* 6 (3): 65.
Mariscal, Gonzalo, Óscar Marbán, and Covadonga Fernández. 2010. “A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies.” *The Knowledge Engineering Review* 25 (2): 137–66. https://doi.org/10.1017/S0269888910000032.
McCullagh, John. 2010. “Data Mining in Sport: A Neural Network Approach.” *International Journal of Sports Science and Engineering* 4 (3): 131–138.
Mikolajec, Kazimierz, Adam Maszczyk, and Tomasz Zając. 2013. “Game Indicators Determining Sports Performance in the NBA.” *Journal of Human Kinetics* 37 (1): 145–151.
Miljković, Dragan, Ljubiša Gajić, Aleksandar Kovačević, and Zora Konjović. 2010. “The Use of Data Mining for Basketball Matches Outcomes Prediction.” In *IEEE 8th International Symposium on Intelligent Systems and Informatics*, 309–312. IEEE.
Miller, Stuart, and Roger Bartlett. 1996. “The Relationship between Basketball Shooting Kinematics, Distance and Playing Position.” *Journal of Sports Sciences* 14 (3): 243–253.
Min, Byungho, Jinhyuck Kim, Chongyoun Choe, Hyeonsang Eom, and RI Bob McKay. 2008. “A Compound Framework for Sports Results Prediction: A Football Case Study.” *Knowledge-Based Systems* 21 (7): 551–562.
Mohammad, Rami M., Fadi Thabtah, and Lee McCluskey. 2014. “Predicting Phishing Websites Based on Self-Structuring Neural Network.” *Neural Computing and Applications* 25 (2): 443–458.
Mondello, Michael, and Christopher Kamke. 2014. “The Introduction and Application of Sports Analytics in Professional Sport Organizations.” *Journal of Applied Sport Management* 6 (2).
Morgan, Stuart, Morgan David Williams, and Chris Barnes. 2013. “Applying Decision Tree Induction for Identification of Important Attributes in One-versus-One Player Interactions: A Hockey Exemplar.” *Journal of Sports Sciences* 31 (10): 1031–37. https://doi.org/10.1080/02640414.2013.770906.
Morris, Thomas. 2000. “Psychological Characteristics and Talent Identification in Soccer.” *Journal of Sports Sciences* 18 (9): 715–726.
Naismith, James. 1996. *Basketball: Its Origin and Development*. U of Nebraska Press.
Novatchkov, Hristo, and Arnold Baca. 2013. “Artificial Intelligence in Sports on the Example of Weight Training.” *Journal of Sports Science & Medicine* 12 (1): 27.
O’Donoghue, Peter. 2009. *Research Methods for Sports Performance Analysis*. Routledge.
Ortega, Enrique, Diego Villarejo, and José M. Palao. 2009. “Differences in Game Statistics between Winning and Losing Rugby Teams in the Six Nations Tournament.” *Journal of Sports Science & Medicine* 8 (4): 523.
Padhy, Neelamadhab, Dr Mishra, and Rasmita Panigrahi. 2012. “The Survey of Data Mining Applications and Feature Scope.” *ArXiv Preprint ArXiv:1211.5723*.

Pai, Ping-Feng, Lan-Hung ChangLiao, and Kuo-Ping Lin. 2017. "Analyzing Basketball Games by a Support Vector Machines with Decision Tree Model." *Neural Computing and Applications* 28 (12): 4159–4167.

Pal, Mahesh, and Paul M Mather. 2003. "An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification." *Remote Sensing of Environment* 86 (4): 554–65. https://doi.org/10.1016/S0034-4257(03)00132-9.

Parejo, Isabel, Álvaro García, Antonio Antúnez, and Sergio Ibáñez. 2013. "Differences in Performance Indicators among Winners and Losers of Group a of the Spanish Basketball Amateur League (EBA)." *Revista de Psicología Del Deporte* 22 (1): 257–261.

Pena, José M., Jose Antonio Lozano, and Pedro Larranaga. 1999. "An Empirical Comparison of Four Initialization Methods for the K-Means Algorithm." *Pattern Recognition Letters* 20 (10): 1027–1040.

Platt, John C. 1999. "12 Fast Training of Support Vector Machines Using Sequential Minimal Optimization." In *Advances in Kernel Methods*, 185–208. MIT press.

Purucker, Michael C. 1996. "Neural Network Quarterbacking." *IEEE Potentials* 15 (3): 9–15.

Rains, Rob. 2011. *James Naismith: The Man Who Invented Basketball*. Temple University Press.

Ray, Siddheswar, and Rose H. Turi. 1999. "Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation." In *Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques*, 137–143. Calcutta, India.

Rees, Loren Paul, Terry R. Rakes, and Jason K. Deane. 2015. "Using Analytics To Challenge Conventional Baseball Wisdom." *Journal of Service Science (Online)* 8 (1): 11.

Rokach, Lior, and Oded Maimon. 2005. "Top-down Induction of Decision Trees Classifiers-a Survey." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 35 (4): 476–487.

Sampaio, Jaime, Manuel Janeira, Sergio Ibáñez, and Alberto Lorenzo. 2006. "Discriminant Analysis of Game-Related Statistics between Basketball Guards, Forwards and Centres in Three Professional Leagues." *European Journal of Sport Science* 6 (3): 173–178.

Sampaio, Jaime, Carlos Lago, and Eric J. Drinkwater. 2010. "Explanations for the United States of America's Dominance in Basketball at the Beijing Olympic Games (2008)." *Journal of Sports Sciences* 28 (2): 147–152.

Singh, Gursharan, Nitin Bhatia, and Sawtantar Singh. 2011. "Fuzzy Logic Based Cricket Player Performance Evaluator." *IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches & Practical Applications* 1 (3): 11–16.

Skinner, Brian. 2012. "The Problem of Shot Selection in Basketball." *PloS One* 7 (1): e30776.

Smith, Lloyd, Bret Lipscomb, and Adam Simkins. 2007. "Data Mining in Sports: Predicting Cy Young Award Winners." *Journal of Computing Sciences in Colleges* 22 (4): 115–121.

Srivastava, Anurag, Eui-Hong Han, Vipin Kumar, and Vineet Singh. 2002. "Parallel Formulations of Decision-Tree Classification Algorithms." In *High Performance Data Mining: Scaling Algorithms, Applications and Systems*, edited by Yike Guo and Robert Grossman, 237–61. Boston, MA: Springer US. https://doi.org/10.1007/0-306-47011-X_2.

Steyerberg, Ewout W., Marinus JC Eijkemans, Frank E. Harrell, and J. Dik F. Habbema. 2000. "Prognostic Modelling with Logistic Regression Analysis: A Comparison of Selection and Estimation Methods in Small Data Sets." *Statistics in Medicine* 19 (8): 1059–1079.

Tan, Sze Huey, and Say Beng Tan. 2010. "The Correct Interpretation of Confidence Intervals." *Proceedings of Singapore Healthcare* 19 (3): 276–278.

Tavana, Madjid, Farshad Azizi, Farzad Azizi, and Majid Behzadian. 2013. "A Fuzzy Inference System with Application to Player Selection and Team Formation in Multi-Player Sports." *Sport Management Review* 16 (1): 97–110. https://doi.org/10.1016/j.smr.2012.06.002.

Teramoto, Masaru, and Chad L. Cross. 2010. "Relative Importance of Performance Factors in Winning NBA Games in Regular Season versus Playoffs." *Journal of Quantitative Analysis in Sports* 6 (3).

Trawinski, Krzysztof. 2010. "A Fuzzy Classification System for Prediction of the Results of the Basketball Games." In *International Conference on Fuzzy Systems*, 1–7. IEEE.

Van Houwelingen, J. C., and S. Le Cessie. 1990. "Predictive Value of Statistical Models." *Statistics in Medicine* 9 (11): 1303–1325.

Verma, J. P. 2016. *Sports Research with Analytical Solution Using SPSS*. John Wiley & Sons.

Vilar, Luís, Duarte Araújo, Keith Davids, Vanda Correia, and Pedro Tiago Esteves. 2013. "Spatial-Temporal Constraints on Decision-Making during Shooting Performance in the Team Sport of Futsal." *Journal of Sports Sciences* 31 (8): 840–846.

Ye, N. 2003. "The Handbook of Data Mining. New Jersey: Lawrence Erlbaum Associates."

Zimmermann, Albrecht. 2016. "Basketball Predictions in the NCAAB and NBA: Similarities and Differences." *Statistical Analysis and Data Mining: The ASA Data Science Journal* 9 (5): 350–364.

Zimmermann, Albrecht, Sruthi Moorthy, and Zifan Shi. 2013. "Predicting College Basketball Match Outcomes Using Machine Learning Techniques: Some Results and Lessons Learned." *ArXiv Preprint ArXiv:1310.3607*.

APPENDICES

APPENDIX A

Notice of Administrative Institutional Review Board Approval.

NORTHERN MICHIGAN UNIVERSITY

OFFICE OF GRADUATE EDUCATION AND RESEARCH
1401 Presque Isle Avenue
Marquette, MI 49855-5301
906-227-2300
906-227-2315
www.nmu.edu/graduatestudies

Memorandum

TO:	Olivia Perrin School of Health and Human Performance
CC:	Randy Jensen School of Health and Human Performance
FROM:	Dr. Lisa Schade Eckert Interim Dean of Graduate Education and Research
DATE:	May 16, 2019
SUBJECT:	IRB Proposal HS19-1044 "Successful shot locations and shot types used in NCAA men's Division I basketball" **IRB Approval Dates: 5/16/19 – 5/15/20** Proposed Project Dates: 5/16/19 – 5/1/20

Your proposal "Successful shot locations and shot types used in NCAA men's Division I basketball" has been approved via the administrative review process. Please include your proposal number (HS19-1044) on all research materials and on any correspondence regarding this project.

Any changes or revisions to your approved research plan must be approved by the Institutional Review Board (IRB) prior to implementation.

If you do not complete your project within 12 months from the date of your approval notification, you must submit a Project Renewal Form for Research Involving Human Subjects. You may apply for a one-year project renewal up to four times.

All forms can be found at the NMU Grants and Research website:
http://www.nmu.edu/grantsandresearch/node/102