

Performance prediction of basketball players using automated personality mining with twitter data

Dominik Siemon

LUT University, Lappeenranta, Finland, and

Jörn Wessels

Technische Universität Braunschweig, Braunschweig, Germany

Abstract

Purpose – The purpose of this paper is to use Twitter data to mine personality traits of basketball players to predict their performance in the National Basketball Association (NBA).

Design/methodology/approach – Automated personality mining and robotic process automation were used to gather data (player statistics and big five personality traits) of $n = 185$ professional basketball players. Correlation analysis and multiple linear regressions were computed to predict the performance of their NBA careers based on previous college performance and personality traits.

Findings – Automated personality mining of Tweets can be used to gather additional information about basketball players. Extraversion, agreeableness and conscientiousness correlate with basketball performance and can be used, in combination with previous game statistics, to predict future performance.

Originality/value – The study presents a novel approach to use automated personality mining of Twitter data as a predictor for future basketball performance. The contribution advances the understanding of the importance of personality for sports performance and the use of cognitive systems (automated personality mining) and the social media data for predictions. Scouts can use our findings to enhance their recruiting criteria in a multi-million dollar business, such as the NBA.

Keywords Basketball, Automated personality mining, Social media, Twitter, Cognitive systems

Paper type Research paper

1. Introduction

Each June, teams of the National Basketball Association (NBA) search for a new superstar during the annual draft among young hopefuls who have played at least one year of college basketball in the United States (Arel and Tomas, 2012). During the draft, every NBA team has the chance to improve its roster for the upcoming season. However, even in the multimillion-dollar business of the NBA, there is an amount of uncertainty in evaluating the right talent for a team based on gathered information (Berri *et al.*, 2011; Groothuis *et al.*, 2007; Sawant *et al.*, 2019). Despite many evaluation options, such as scouting reports, workouts and performance tests, no team official can guarantee that the selected player will be a suitable fit for the NBA or the team roster (Arel and Tomas, 2012).

Various analyses have been done on unsuccessful draft picks, or players drafted in late rounds who turned out to be fortunate picks with successful careers (Berger and Daumann, 2021; Teramoto *et al.*, 2018). While media often focuses on who will be picked, experts have



their own opinions on available players and see different strengths and weaknesses in them (Beene, 2019). Every one of them is convinced to have the perfect formula for knowing what player should be drafted at which position by any team. Nevertheless, even their assessments often contain errors due to missing information, which shows that even the greatest experts are not omniscient (Beene, 2019; Sailofsky, 2018). This raises the question of whether the selection process can be improved with further information not yet known to talent scouts.

Players are not scouted solely on their previous game statistics and their physical talent. Other factors include their work ethic, maturity, mental toughness, trainability and personality (Beene, 2019; Sailofsky, 2018). Especially the personality of players is considered highly relevant, as studies have shown that personality has an impact on the behavioral performance of athletes and predictive power (Berri *et al.*, 2011; Craighead *et al.*, 1986; Johnson, 1972; Maddi and Hess, 1992; Schurr *et al.*, 1988). Due to the Covid-19 pandemic, classic information-gathering of the teams prior to the draft through pre-draft camps, the NBA combine, medical testing, interviews and scouting was largely prevented (Quinn, 2020). While many teams already use digital support to evaluate talent, even this comes with restrictions (Atlas and Zhang, 2004; Beene, 2019). For example, psychological components of the evaluation process are particularly missing that would normally be covered by a large number of face-to-face conversations among talents and teams' decision makers (Quinn, 2020; Teramoto *et al.*, 2018). Until now, questionnaires are often used to predict prospects' personalities prior to the draft, including the NBA wonderlic test (Matthews and Lassiter, 2007) or the athletic intelligence quotient (Bowman *et al.*, 2021); these self-report questionnaires can admittedly provide rich information about people's conscious self-concepts. However, the reliability of self-reported tests to truly portray participants' personalities is often doubted (Boyd and Pennebaker, 2017; McCrae and Costa, 1982). Indeed, various researchers have found that even common and well-validated self-assessment instruments are inadequate for accurately capturing even fundamental human patterns such as expressions of happiness, physical activity, on-the-job behavior or other emotional conditions (Boyd and Pennebaker, 2017).

Therefore, a growing number of studies have posited the way people use words as an alternative to assess personality (Boyd and Pennebaker, 2017; Tausczik and Pennebaker, 2010). It is reliable over time, internally consistent, predictive of a wide range of behaviors, even biological activity and varies considerably between individuals (Boyd and Pennebaker, 2017; Faliagka *et al.*, 2012). An advantage is that individuals do not have to undergo a special test, such as a self-report questionnaire to provide useful personality data (Boyd and Pennebaker, 2017). Considering that everyone uses words uniquely and language used in everyday life in part reflects a person's psychological state, everyday communication such as that used in social media can be analyzed (Pennebaker, 2013). One place where people express themselves, interact and communicate with each other is social media, such as Facebook or Twitter, which is therefore very well suited as a data basis for making various predictions (Ballouli and Sanderson, 2012; Filo *et al.*, 2015; Quercia *et al.*, 2011).

The goal of our research is to improve the prediction of NBA performance based on past basketball performance and personality traits automatically detected by Twitter data. Consequently, we try to answer the following research question:

RQ. Can personality traits obtained from social media data be used to predict the performance of basketball players beyond the extent to which player statistics do?

To evaluate the prospects' personalities, automated personality mining using social media data is deployed. The overarching goal is to provide additional information to be leveraged by NBA teams to examine players' talent prior to the NBA draft. In addition to this practical contribution, we aim to make theoretical contributions that help in the understanding of personality and high-performance sport. We show which personality traits are particularly relevant for the sport of basketball and how these traits can determine certain players'

careers. Further, we contribute to the understanding of how social media data can be used to produce valuable individual information of potential athletes.

The remainder of our paper proceeds as follows: First, we introduce the topic and the research purpose and justification. From this, the second section presents the theoretical background of the NBA, college basketball and player statistics as well as the theoretical basics of personality measuring models and the procedure of automated personality mining. The current state of research on the topics of personality and sports performance, as well as the examination of interrelationships among basketball statistics is included in [section 3](#). Based on the theoretical backgrounds and literature review, research hypotheses are derived in [section 4](#). [Section 5](#) provides the methods used for data collection, description and analysis, followed by the results in [section 6](#). Our findings are discussed in [section 7](#), after which [section 8](#) summarizes the main conclusions and suggests recommendations for further research.

2. Theoretical background

2.1 Basketball and player performance statistics

In 1946, the NBA was founded and became a multimillion-dollar business with an average franchise value of \$2.2 million in 2020, in which 30 teams and over 500 players play 82 games each season ([Badenhausen, 2020](#); [Berri, 2017](#)). In the United States, college basketball has a similar status to professional basketball, even if the business does not bring nearly such enormous economic success as the NBA ([Kian et al., 2008](#)). The status of being an amateur college athlete enables players to receive higher education in an institution and allows athletes to showcase their skills on a larger, televised platform ([Zizzi et al., 2003](#)).

Scouts tend to agree that video tapes of players do not always tell the whole truth about their skills ([Beene, 2019](#); [Sailofsky, 2018](#); [Teramoto et al., 2018](#)). Observations such as how a player's teammates respond to his failure or success, the body language and response to coaching during time-outs and in the game, his mood on the bench now provide a more complete picture about a future star ([Quinn, 2020](#); [Reiter, 2020](#)). Additionally, a player's personality is normally evaluated primarily with personal interviews, such as those used by a company for evaluating potential candidates to join their organization ([Craighead et al., 1986](#); [Matthews and Lassiter, 2007](#)). For questions of personality and behavioral assessment, most of the NBA teams rely on a standard personality questionnaire given to the players prior to the personal interview ([Craighead et al., 1986](#)).

Game statistics are recorded at every organized basketball game and reveal to the public who won, who lost and which player performed which way ([Berri, 2017](#)). They also play a decisive role in the signing of new players for the next season ([Beene, 2019](#); [Schwamborn, 2014](#)). Statisticians use score sheets to follow passes, dribbles, shots, rebounds, steals, fouls and everything else that happens during the game ([Oliver, 2004](#)). In a NBA stat sheet, there are 22 corresponding statistics; on an official NCAA college basketball stat sheet, there are 15 ([Oliver, 2004](#)).

Basic statistics contain values captured on the stat sheets of basketball officials without a mathematical formula, like points per game (PPG), rebounds per game (RPG), assists per game (APG) and steals per game (SPG) of a player. Correspondingly, the field-goal percentage (FG%), free-throw percentage (FT%) and three-point percentage (3P%) are recorded, which resulting in the division of shot attempts and successful makes ([Oliver, 2004](#)). However, basketball is much harder to predict with standardized statistics, especially since it is much more of a team sport, which is why specific performance measurements have been constructed ([Berri, 2017](#); [Oliver, 2004](#); [Schwamborn, 2014](#)).

The first performance measurement, known as player efficiency rating (PER), accumulates all positive performance contributions and deducts all negative performance contributions of a player to create a pace-adjusted, per-minute rating of the player's performance ([Kubatko et al., 2007](#); [Oliver, 2004](#)). This rating allows for comparing players, even if their playing time is

different. The second performance measurement, called win shares (WS), aims to spread the team's performance among individual team members. It is computed based on player, team and league-wide statistics and includes both offensive statistics and defensive statistics. With WS being an accumulated statistic, win shares per 48 min (WS48) provides a clearer idea of how effective a player is per game over the course of a player's career (Kubatko *et al.*, 2007; Oliver, 2004). The last performance measure relevant to our research is Beech expected performance rating (BEPR), which uses the career statistics of points, rebounds, and assists per game and generates a rating value by totaling those numbers, to categorize players into a total of five categories. With a rating higher than 20 being the highest (called "star") and a rating between 5 and 9.99 being the lowest (Beech, 2009).

Our research also covers two college-related performance statistics. Strength of schedule (SOS) is defined "as the number of games a team on the borderline of the annual national tournament would expect to win if they played that schedule. This gives a direct way of quantifying how well different teams have done relative to the schedules they have played" (Fearhead and Taylor, 2010). The NCAA calculates the SOS of teams with the all-time won-lost percentage against other teams they played in their schedule and averages those percentages (Wright, 2008). The other value is WS per 40 min (WS40) that corresponds to the described WS48. However, the playing time in college games lasts only 40 min instead of 48, which leads to the difference (Berri *et al.*, 2011).

2.2 *The big five model of personality*

Personality can be loosely defined as a construct that makes a person's behavior, thoughts and feelings reasonably consistent, but at the same time differentiates individuals (Allport, 1961; McCrae and Costa, 1997). The foundations for the big five model are psycho-lexical studies of Allport and Odbert (1936), based on the view that all significant facets for the characterization of important behaviors for the co-existence of individuals are represented by natural language (Allport and Odbert, 1936). This tenet implies that personality-descriptive terms available in everyday language, such as adjectives and nouns, are suitable for depicting all important differences between human beings (Pennebaker, 2013; Tausczik and Pennebaker, 2010). This has caused Allport and Odbert (1936) to analyze over 4,500 terms attributable to personality traits (Allport and Odbert, 1936).

The fundamental idea of the big five originated in studies conducted by Tupes and Christal in 1958. In their surveys, different samples were compared with each other. The result of the analysis provided five matching factors (Tupes and Christal, 1992). The first use of the term "big five" was by Goldberg in 1981, with the universality of the model evidenced by a high degree of agreement among findings from different studies in multiple languages and cultures (Saucier and Goldberg, 1998). Later, the five-dimensional personality model, consisting of the dimensions neuroticism, extraversion, openness to experience, agreeableness and conscientiousness, was introduced (McCrae and Costa, 1997; McCrae and John, 1992). This theory is based on the starting point of stable and consistent personality differences among individuals, which are traced back to a substantial degree to genetic differences. This accordingly reflects a result of human adaptation to environmental conditions (McCrae and John, 1992; Tupes and Christal, 1992).

The dimension of neuroticism encompasses emotional robustness and emotional sensitivity (McCrae and John, 1992). The dimension of extraversion is a measure of the quantity and intensity of relationships with the environment. Openness to experience captures the interest and extent of engagement with new experiences, sensations and impressions. The named desire possesses a differing degree of expression in most individuals. This involves, among other things, the need for variety, intellectual interest in others and the independent formation of opinions. The agreeableness scale measures attitudes and habitual behaviors in social relationships, which

are explained by social desirability and value beliefs of individuals. Conscientiousness as a dimension explores the determination and purposefulness with which a person approaches challenges and the way in which he or she executes imposed tasks. Aspects of competence, sense of responsibility, prudence, striving for achievement, sense of arrangement and self-discipline also belong to this description (McCrae and Costa, 1997; McCrae and John, 1992).

2.3 Automated personality mining of twitter data

Automated personality mining is a way to classify the personality profile of social media users, such as of Facebook, Twitter or Instagram using a personality model namely the big five (Adi *et al.*, 2018; Buettner, 2017; Faliagka *et al.*, 2012). Several studies have been conducted on predicting the personality from written texts on social networks (dos Santos *et al.*, 2017; Ramezani *et al.*, 2022; Sun *et al.*, 2018). For example, in the field of Twitter data, Golbeck *et al.* (2011) used automated personality mining to predict the personality traits by analyzing the usage frequency of particular categories of words and the variation in word usage. The study was able to predict personality with a deviation of 10% from the personality test used as the benchmark (Golbeck *et al.*, 2011). Another study based on Twitter and Facebook data, in which users use both social media services, found that user personality can be easily and effectively predicted from public social media data (Quercia *et al.*, 2011). Further, Arnoux *et al.* (2017) studied the accuracy of prior work on big five personality and the dependence on the size of the input text, and introduced a method using word embedding and Gaussian process. The study shows that the method used outperformed the compared state-of-the-art methods for personality prediction and was able to predict the big five personality traits of users based on their social media texts in a real-world context (Arnoux *et al.*, 2017).

To analyze the personality characteristics of the college basketball players considered, IBM Watson Personality Insights, developed by IBM, is used (Ferrucci, 2012; Pennington *et al.*, 2014). The service used to evaluate texts is based on an open-source word-embedding algorithm of global vectors for word representation, or GloVe, which creates a vector representation of the provided text (Pennington *et al.*, 2014). As training data, the service uses survey results from thousands of users and the texts of these users' Twitter feeds (IBM Watson, 2020). IBM Watson Personality Insights creates big five personality profiles based on text input annotations, combining the psychology of language with various data analysis algorithms. The different text must contain a minimum number of 100 words for a linguistic analysis and must comply with a particular quality. As the number of words increases, the accuracy of the analysis improves, with the maximum accuracy of the service being provided by 3,000 words (Ferrucci, 2012; IBM Watson, 2020).

The real-time Internet service Twitter is primarily used to share text messages limited to 140 characters called tweets (Li *et al.*, 2019). It is used for its ability to disseminate information quickly and unfiltered (Ballouli and Sanderson, 2012; Filo *et al.*, 2015; Stavros *et al.*, 2014). In his study on suitability of Twitter as a data basis for scientific work, Pfaffenberger (2016) argue that Twitter is an interesting and noteworthy source of data (Pfaffenberger, 2016), which makes it an important tool for many kinds of data analysis (Stieglitz *et al.*, 2018). Its unrestricted and rapid communication allows detailed insights into the interests, opinions and moods of users, important factors influencing personality (Pfaffenberger, 2016). According to Twitter, the NBA was the most tweeted-about sports league in 2018. With Twitter as the common meeting spot for the team's fans, it has grown to a community including players, executives, fans and journalists (Li *et al.*, 2019; Nisar *et al.*, 2018). NBA-Twitter has become a platform where the participants can easily jump into, follow or start a conversation with and about the game's players and personalities. Unlike many athletes from other sports, NBA stars are open to share parts of their private lives, which helps create a deeper connection with the fans (Maese, 2018), which is one reason they possess a large number of followers.

3. Related work

In sports psychology, the relationship between the athletic performance and personality is of high interest (Craighead *et al.*, 1986; Johnson, 1972; Maddi and Hess, 1992; Mirzaei *et al.*, 2013; Schurr *et al.*, 1988). Especially through cognitive skills, emotion regulation, motivation enhancement and interventions in the athlete's social environment, the influence of personality on athletic performance has been scientifically proven (Alfermann and Stoll, 2017). Alfermann and Stoll (2017) compared athletes and individuals who do not participate in sports, regarding their personality. The high self-confidence and low anxiety of professional athletes led to low scores of neuroticism and their performance orientation and competitiveness led to higher conscientiousness scores (Alfermann and Stoll, 2017). In addition, several studies have identified persistence as an essential trait for athletes (Piedmont *et al.*, 1999). Furthermore, in a study of the big five model of NBA players, Siemon *et al.* (2018) focused on analyzing the difference in the values of the personality features among NBA All-Stars and other players. The results showed that the traits of conscientiousness and agreeableness have the biggest positive difference (Siemon *et al.*, 2018).

In a meta-analysis of 42 different sport groups regarding the relationship between personality traits and athletic performance, contradictory results were initially found for the dimension of extraversion (Hardman, 1973). After an in-depth analysis, it was concluded that certain sports represent advantages or disadvantages for extraverts. It was found that they need and seek a higher cortical excitation. This is more likely to be provided by high-performance sports than by recreational sports. Extraverted athletes are also better at tolerating pain and, therefore, are better at contact sports (Sohrabi *et al.*, 2011). Another factor that emerged in the study is that individuals with high scores on this dimension prefer team sports as they respond to social stimuli. Introverts have advantages in sports that require fine motor skills or concentration on a few key stimuli, such as sport shooting (Hardman, 1973).

The dimension of neuroticism has also been seen as an influencing factor for athletic performance in various studies. Emotional robustness is to be expected at low levels among high-level athletes in the field of neuroticism (Morgan, 1980). Another research source has concluded that athletes must be able to perform optimally at specific times, such as during competitions. In this context, psychological stress is often increased by adverse conditions. Only if athletes can deal with those kinds of situations adequately are they able to reach the top levels. This background is seen as the basis for the idea that top athletes must not be neurotic in order to be successful (Gaudreau and Blondin, 2004; Sarkar and Fletcher, 2014).

In the field of conscientiousness, Ostendorf and Angleitner (2004) have described the dimension as a cognitive control of impulses. This includes the ability to control, plan and execute actions and is associated with success in one's profession as well as extraordinary achievements in the areas of music and sports. According to this, conscientiousness is a requirement for high-performance sports, represented by the will to achieve results (Mirzaei *et al.*, 2013). However, this assertion is not corroborated by a study, which leads to questioning this statement (Ostendorf and Angleitner, 2004). Nonetheless, in another study, athletes with high scores in conscientiousness and low scores in neuroticism were found to perform better over the course of a competitive season (Allen and Laborde, 2014).

Research has revealed that high scores among athletes in the dimension of agreeableness lead them to have more beneficial relationships with their teammates and coaches (Piedmont *et al.*, 1999). The three dimensions of extraversion, conscientiousness and agreeableness can also predict unhealthy exercise behavior among older adults in the context of strength and mobility (Allen and Laborde, 2014).

Finally, openness to experience is often associated with sports involving higher risk, such as mountain climbing or motorcycling (Tok, 2011). Similarly, in a study of free divers with

high risk tolerance, it was found that high openness and extraversion scores, combined with low emotional scores, resulted in a positive effect on their performance (Baretta *et al.*, 2017).

4. Hypothesis development

The first objective of this paper is to identify a relationship between selected college basketball statistics, big five trait personality traits and NBA players' performance. The values of PER, WS, WS48 and BEPR are used to measure the player productivity. For our hypothesis development, we primarily rely on research from the field of sports and specifically team sports, which requires similar physical activities as basketball. While there is already a lot of research on which personality traits are responsible for certain behaviors in the area of job performance, desk activities or art and creativity (Amabile, 1983; Buettner, 2017; Hogan and Holland, 2003; Judge *et al.*, 2002), there is only little research on correlations and relations between personality traits and sports performance. Since studies of high-performance athletes have indicated that people with a high level of extraversion are better at enduring pain and prefer team sports (Hardman, 1973), which are often associated with a successful athlete, it is assumed that extraversion and the performance of players have a positive relationship. Therefore, the following hypothesis is made:

H1. A positive correlation exists between extraversion and NBA performance.

Studies on neuroticism have shown the importance of the body and mind working together in an optimum way. Therefore, it is important that athletes are emotionally robust and have low scores in this trait (Morgan, 1980). They must be able to manage competition stress and disadvantageous conditions as well (Gaudreau and Blondin, 2004). Since basketball players are often faced with stressful situations, such as last-second game decision-making as well as perceived disadvantageous refereeing decisions (Craighead *et al.*, 1986; Zizzi *et al.*, 2003), the neuroticism domain and the NBA performance are expected to be negatively related. The following hypothesis is thus proposed:

H2. A negative correlation exists between neuroticism and NBA performance.

The characteristic of conscientiousness includes, according to studies, the ability to control, plan and execute actions, which must be especially present among high-performance athletes to achieve success (Ostendorf and Angleitner, 2004). Moreover, characteristics such as performance orientation and competitiveness are expected to lead to a high value in this dimension (Piedmont *et al.*, 1999). The features in strong expressions, such as responsible, achievement-oriented, reliable and hardworking, also match in this regard (Bipp, 2006). All these characteristics mark a successful and high-performing athlete. Hence, it can be concluded that the values in the range of conscientiousness have a positive relationship with the performance of basketball players, creating the following hypothesis:

H3. A positive correlation exists between conscientiousness and NBA performance.

Based on study findings suggesting that high scores in agreeableness lead to better relationships with teammates and coaches (Piedmont *et al.*, 1999), it can be assumed that the score in this category has a positive relationship with the athletes' performance in team sports, leading to the following statement:

H4. A positive correlation exists between agreeableness and NBA performance.

Given that, thus far, only studies related to extreme sports could present effects of openness to experience (Baretta *et al.*, 2017; Tok, 2011), it is assumed that the trait has no significant relationship with performance in team sports, this study posits:

H5. No correlation exists between openness to experience and NBA performance.

Relationships and dependencies between college statistics and NBA performance have been demonstrated by several studies, as mentioned previously. In addition, the size of the conference to which the college belongs to has been identified as relevant in draft position and college statistics. The size of the conference can also be related to the SOS value (Coates *et al.*, 2010). Therefore, the following hypothesis is put forward:

H6. A positive correlation exists between the college statistics of FG%, FT%, 3P%, PPG, RPG, SPG, APG, and SOS and NBA performance.

After an examination of the hypotheses, the possibility to predict prospects' NBA performance potential based on college statistics and the personality is analyzed. To specify the research, the following exploratory question is addressed:

EQ. Can the combination of two or more values of the college players statistics and the personality trait proficiencies predict performance statistics of NBA players?

5. Methodology

A database, including basketball performance statistics and personality traits mined via Twitter, was created for the testing of the hypotheses and answering of the exploratory question in accordance with the data analysis process outlined below. In the following, the steps of data procurement are specified. The collected data is presented and, the data analysis to provide the results is described.

5.1 Data collection and data analysis

Collecting data for the data analysis required three major steps. First step involves the collection of all related draft classes including relevant undrafted players and their statistic values from college and the NBA. Draft classes and the decisive NBA statistics for all players are taken from the website Basketball References (www.basketball-references.com), and statistics for college basketball were extracted from the website Sport Reference (www.sport-reference.com). To retrieve the statistics from the websites, the robotic process automatization tool UiPath was applied (Tripathi, 2018; van der Aalst *et al.*, 2018).

In the second step we implemented a Java-based tool using the Twitter API to extract text from Twitter timelines of players who were eligible for analysis. When querying timelines of relevant players, it was ensured that retweets (i.e. tweets that were not written by the player) were not included. The third step was the analysis of the received Twitter data to get the personality traits of the players. This step was performed using the IBM Watson Personality Insights Service (Ferrucci, 2012) using our Java-based tool.

In our data collection we consider draft classes between 2007 and 2011, because players drafted in these years have either proven themselves in the league by now or were not resigned by their respective teams due to lack of performance. Due to the restriction that only players who have participated in college basketball, and therefore have at least one season of statistics at their alma maters, were evaluated, those who either played abroad prior to their NBA career or did not attend college were excluded. The second restriction includes a minimum number of 20 NBA games and a minimum of one year of service (YOS), because with this number of games, a quarter of a complete NBA season was played, and therefore the statistic contains an acceptable minimum.

Additionally, the use of IBM Watson Personality Insights also limited the sample, since a personality profile could only be created by retrieving enough text (minimum of 1,600 words).

Each dimension's calculation results in a value between 0 and 1 with 0.500.500 presenting the neutral boundary. While a result close to 0 represents a low expression of the personality trait, a result close to 1 reflects the opposite.

After filtering by the restrictions, the initial sample size was brought down from 400 players to 185. Data collection of the Twitter data was performed in January 2021, including all Tweets of our selected players until this time. An overview of the retrieved dataset can be found in [Table 1](#).

We used SPSS version 26 for the descriptive data analysis, as well as for the Pearson's product-moment correlation and the multiple linear regressions. To test our hypotheses, a Pearson's product-moment correlation test with a 95% confidence interval was performed. For the analysis of the strength and direction of the correlations, the correlation coefficient r is used ([Cohen, 2013](#)). The exploratory question was examined using multiple linear regressions. The results of the correlation analysis and previous research, as stated in the related work section, served as the basis for the calculation. The aim was to select college statistics and the personality characteristics that had a highly significant correlation effect with each of the NBA performance measures. Overall, four multiple linear regressions were performed, one respectively for each of the selected performance statistics PER, WS, WS48 and BEPR. To assess variance explanation (goodness of fit), guidelines according to ([Cohen, 2013](#)) were used in the analysis of the regressions.

6. Results

6.1 Descriptive data summary

Data of the basketball players' personality presented low mean scores regarding the characteristics of agreeableness ($M = 0.260$) and conscientiousness ($M = 0.310$). In terms of extraversion ($M = 0.660$) and openness to experience ($M = 0.725$), more average means with the tendency toward positive were displayed. The trait of neuroticism featured a high mean score of $M = 0.860$.

In terms of the descriptive summary of the basketball statistics, mean value of the career PER of the players in the data set is 13.403, slightly below the value of 15 expected as a benchmark for an average starting-five player. All NBA performance statistics showed a high difference between the minimum and maximum but displayed rather moderate standard deviations. In terms of college values, SOS featured a high mean value ($M = 6.152$). Comparing the means of WS per minutes between NBA (WS48) and college (WS40) results indicated that the college value has a higher expression at 0.195 to 0.082, respectively. The standard deviation of both values showed similarities. [Table 2](#) presents the descriptive statistics.

6.2 Person's product-moment correlation

A highly significant ($p < 0.01$) light positive correlation was found between all performance values and the trait extraversion, which confirms [H1](#) for these values. In the case of [H2](#), the

	Quantity	%
Total number of players	400	100.00
Not played in college	58	14.50
No twitter profile available	89	22.25
Not enough NBA games	21	5.25
Not enough tweets	47	11.75
Total dataset	185	46.25

Note(s): Quantity of drafted players per year: 60

Table 1.
Player dataset
numbers and
percentages

	Min	Max	M	SD
Agreeableness	0.063	0.653	0.260	0.096
Conscientiousness	0.068	0.785	0.310	0.135
Extraversion	0.429	0.944	0.660	0.103
Openness	0.216	0.933	0.725	0.075
Neuroticism	0.569	0.958	0.860	0.060
NBA PER	1.400	25.200	13.403	4.306
NBA WS	-1.300	144.400	22.740	26.681
NBA WS48	-0.141	0.700	0.082	0.079
NBA BEPR	1.500	38.800	13.699	8.048
College FG%	0.371	0.646	0.490	0.060
College 3P%	0.000	0.511	0.283	0.142
College FT%	0.437	0.950	0.754	0.539
College PPG	1.000	26.600	13.195	4.122
College RPG	1.900	12.400	5.627	2.237
College APG	0.300	6.500	2.276	1.478
College SPG	0.200	2.600	1.080	0.474
College SOS	-10.190	10.220	6.152	3.423
College WS40	0.088	0.790	0.195	0.073

Performance prediction of basketball players

Table 2. Descriptive statistics for NBA and college values

Note(s): $N = 185$

null hypothesis for the neuroticism trait could not be rejected because, although negative correlations were found with the performance scores, none of the scores possessed significance. Conscientiousness had a high significant ($p < 0.01$) light positive correlation with all NBA performance measures, which confirms the hypothesis H3. For the correlation of NBA performance measures and the trait agreeableness only the values PER ($r = 0.196, p < 0.01$) and BEPR ($r = 0.202, p < 0.01$) revealed highly significant but low positive correlations. WS ($r = 0.179, p < 0.05$) showed a significant correlation with a low positive effect, while WS48 showed no significant correlations with the trait. Therefore, the null hypothesis is rejected for the values PER, BEPR and WS, while it is supported for WS48. As determined in hypothesis H5, openness to experience showed no significant correlations with PER, WS, WS48 or BEPR. The results are presented in Table 3.

Correlations between college statistics and NBA performance scores are provided in Table 4. Except for SOS and FT%, both rejecting the formulated hypothesis, all of the players' average statistics during their college careers showed significant ($p < 0.05$) or highly significant ($p < 0.01$) correlations. For the PER attribute, a highly significant correlation was shown by FG% ($r = 0.365, p < 0.01$) and RPG ($r = 0.305, p < 0.01$) with a moderate positive correlation together with PPG ($r = 0.217, p < 0.01$) and WS40 ($r = 0.287, p < 0.01$), yet only to a slight positive extent. All these attributes confirm hypothesis H6 for the PER metric. The values that rejected H6 were 3P%

	NBA PER	NBA WS	NBA WS48	NBA BEPR
Extraversion	0.239**	0.273**	0.208**	0.232**
Neuroticism	-0.075	-0.098	-0.014	-0.090
Conscientiousness	0.223**	0.242**	0.240**	0.194**
Agreeableness	0.196**	0.179*	0.109	0.202**
Openness	-0.083	-0.051	-0.142	-0.035

Table 3. Correlation between big five traits and NBA performance

Note(s): Pearson's product-moment correlation. $N = 185$. * $p < 0.05$, ** $p < 0.01$

($r = -0.166, p < 0.05$), which indicated a significant slightly negative effect, and APG, showing a nonsignificant negative correlation. SPG presented a nonsignificant correlation as well.

With respect to WS, FT% showed a negative but not significant correlation, hence contradicting hypothesis H6. The same applied to the values 3P% and APG, which both showed no significance. All other values evaluated confirmed the hypothesis, with significant positive correlations exhibited. Among them, FG% ($r = 0.213, p < 0.01$), PPG ($r = 0.235, p < 0.01$), SPG ($r = 0.226, p < 0.01$) and WS40 ($r = 0.209, p < 0.01$) indicated highly significant low positive relationships. In addition, RPG ($r = 0.170, p < 0.05$) had a significant slight positive correlation.

The measurement of WS48 and FG% showed the highest positive highly significant correlation among all interrelations ($r = 0.427 p < 0.01$), although the effect can be rated as moderate. Other highly significant correlations were detected with RPG ($r = 0.277, p < 0.01$) along with WS40 ($r = 0.227, p < 0.01$), the effect can only be rated as small. The mentioned values reject the null hypothesis. In contrast, this was confirmed by the negative correlations of APG ($r = -0.241, p < 0.01$) and 3P% ($r = -0.162, p < 0.05$), which are highly significant or significant, respectively. The same applies to the nonsignificant values of correlations with PPG and SPG.

BEPR showed a highly significant but moderate relationship with the PPG value ($r = 0.319, p < 0.01$). Furthermore, there remained a highly significant but slight correlation effect with SPG ($r = 0.289, p < 0.01$) and a significant value with the same effect size for RPG ($r = 0.179, p < 0.05$), APG ($r = 0.153, p < 0.05$) and WS40 ($r = 0.188, p < 0.05$). These five values confirm hypothesis H6 in relation to BEPR. FG% and 3P% showed no significant relationships and thus dismissed the alternative hypothesis.

6.3 Multiple linear regression

Following the Pearson's product-moment correlation for testing the stated hypotheses, the explorative question (EQ) was examined based on the results obtained. For this purpose, four multiple linear regressions with a 95% confidence interval were performed.

For the multiple linear regression of the dependent variable of PER, the college statistics FG% and RPG as well as the personality trait of extraversion were employed as independent variables. A significant regression equation was found ($F(3, 181) = 15.918, p < 0.001$) with an R^2 of 0.209. They also showed significant B values, as presented in Table 5.

For the variable of WS, a multiple linear regression was modeled with the independent variables PPG, SPG and extraversion. The model was found significant ($p < 0.001$) with $F(3, 181) = 9.561$ and an R^2 of 0.137. The value PPG showed no significant B scores for this regression different from SPG and extraversion (Table 6).

	NBA PER	NBA WS	NBA WS48	NBA BEPR
College FG%	0.365**	0.213**	0.427**	0.100
College 3P%	-0.166*	0.027	-0.162*	0.110
College FT%	-0.015	-0.023	-0.064	0.082
College PPG	0.217**	0.235**	0.058	0.319**
College RPG	0.305**	0.170*	0.277**	0.179*
College APG	-0.091	0.055	-0.241**	0.153*
College SPG	0.082	0.226**	-0.118	0.289**
College SOS	0.020	0.070	0.063	0.087
College WS40	0.287**	0.209**	0.227**	0.188*

Note(s): Pearson's product-moment correlation. $N = 185$. * $p < 0.05$, ** $p < 0.01$

Table 4.
Correlation between
college statistics and
NBA performance

A significant overall regression ($F(3, 181) = 17.836, p < 0.001$) was found for WS48 along with the independent variables of FG%, RPG and conscientiousness with $R^2 = 0.228$. The personality trait and the value for FG% had highly significant, while RPG had nonsignificant scores, which is shown in Table 7.

In the last multiple linear regression, the dependent variable BEPR was tested with the independent variables of PPG and SPG for the college statistics as well as extraversion as the personality value. The model is statistically significant ($p < 0.001$) with $F(3, 181) = 12.316$. $R^2 = 0.170$ showed a moderate goodness of fit, and all B scores of the independent values are found to be significant (PPG) or highly significant (SPG, extraversion) see Table 8.

7. Discussion

7.1 Findings based on the formulated hypotheses

Considering the results of H1, it is determined that all selected performance values have a positive relationship with the personality trait extraversion which confirms the research

	<i>B</i>	<i>Std. Error</i>	<i>Standardized B</i>
(constant)	-4.269	2.924	
FG%	18.549**	5.539	0.255
RPG	0.374*	0.147	0.194
extraversion	9.811**	2.781	0.235

Note(s): Dependent variable: PER. * $p < 0.05$, ** $p < 0.01$

Table 5.
Coefficients of multiple linear regression for PER

	<i>B</i>	<i>Std. Error</i>	<i>Standardized B</i>
(constant)	-41.866**	12.993	
PPG	0.947	0.500	0.146
SPG	8.372*	4.342	0.149
extraversion	9.811**	2.781	0.235

Note(s): Dependent variable: WS. * $p < 0.05$, ** $p < 0.01$

Table 6.
Coefficients of multiple linear regression for WS

	<i>B</i>	<i>Std. Error</i>	<i>Standardized B</i>
(constant)	-0.203**	0.044	
FG%	0.462**	0.101	0.346
RPG	0.004	0.003	0.116
conscientiousness	0.118**	0.039	0.202

Note(s): Dependent variable: WS48. * $p < 0.05$, ** $p < 0.01$

Table 7.
Coefficients of multiple linear regression for WS48

	<i>B</i>	<i>Std. Error</i>	<i>Standardized B</i>
(constant)	-5.758	3.844	
PPG	0.431*	0.148	0.221
SPG	3.061**	1.285	0.180
extraversion	15.841**	5.306	0.203

Note(s): Dependent variable: BEPR. * $p < 0.05$, ** $p < 0.01$

Table 8.
Coefficients of multiple linear regression for BEPR

conducted by [Hardman \(1973\)](#). Because of the team sports background, but also especially the theory that players with pronounced extraversion can endure more pain fits the conditions under which NBA athletes compete and perform in 82 regular-season games and potential playoff games in less than seven months.

For neuroticism, the formulated [H2](#) could not be confirmed. However, it is worth discussing that all players in this dataset tend to have a high value of neuroticism. Through the findings of [Morgan \(1985\)](#) and [Sarkar and Fletcher \(2014\)](#) that athletes must be emotionally robust or able to cope with competitive stress and negative conditions to be successful, a negative relationship between performance and neuroticism was implied. Arguing against this in terms of NBA players, particularly in the last year, the influence of the Covid-19 pandemic must be considered. Players have been isolated throughout the year as never before and due to the abrupt end of the season and the continuation of the season in the so-called bubble, the feelings of isolation and related emotional and mental breakdowns can have resulted in increased emotional expressions on social media ([Reiter, 2020](#)). This assumption is supported by the publicizing of mental health issues by various NBA players ([Medina, 2020](#)), as many players deal with anxiety ([Burke et al., 2000](#); [O'Hallarn et al., 2019](#); [Thomas, 2019](#)). Thus, in all accounts, the theories from previous studies do not necessarily apply to NBA athletes, which is reflected in the results.

[H3](#) has been confirmed by the findings, as all four performance measures showed a highly significant positive correlation with conscientiousness. As mentioned by [Piedmont et al. \(1999\)](#), characteristics such as performance orientation and competitiveness (facets of conscientiousness) lead to a high value. Compared to other sports such as football (i.e. soccer), basketball is an high-speed sport with baskets scored frequently and constant one-on-one duels. Therefore, basketball in particular shows that a high degree of competitiveness leads to higher performance as players can influence the game to a great extent with their individual performances ([Lázaro et al., 2014](#)).

The results concerning hypothesis [H4](#) are unable to completely confirm the findings of [Piedmont et al. \(1999\)](#). Unlike other characteristics, where all performance scores showed equal significant values for the respective trait, WS48 showed no correlation, while WS showed only a low significant correlation. Despite this limitation, the tested correlation values of PER and BEPR associated with NBA performance support the results of [Piedmont et al. \(1999\)](#).

In line with [H5](#), the results of the correlation of NBA performance measures and openness to experience showed no significant relationship. As stated, the trait mainly contributes to extreme sports, which are seen mostly as an individual sport ([Baretta et al., 2017](#); [Tok, 2011](#)); thus, this category does not include basketball, as a recreational team sport.

In the case of FG%, the correlation with PER, WS and WS48 support [H6](#). Only BEPR showed no correlation, which, however, is related to the structure of the measurement variable. The results containing 3P%, FT%, SPG and APG mainly contradict the stated hypothesis, which can be explained by the different positions of basketball players. As players serve different purposes on the court depending on their position, it can be noted that centers typically have worse free-throw and three-point percentages as well as low scores in assists and steals compared to guards and forwards. Contrary to the hypothesized association, no relationship between SOS and NBA performance was observed in the results. It was assumed that a high SOS value indicates that the respective player belongs to a large college conference and thus has an influence on NBA statistics since an influential difference in the size of the conference was determined by the study done by [Coates et al. \(2010\)](#). Since drafting a player from small conferences with a low SOS tends to be the exception, the low number of such examples can suggest that no significant connections could be discovered.

7.2 Findings based on the formulated exploratory question

The results of the multiple linear regressions show that all performed regressions with the NBA performance measures as the dependent value showed statistically significant relationships with at least one college statistic value in combination with a personality trait. When comparing the regression models in terms of their variance explanations, it must be noted that all models showed a value that can be described as moderate.

The regression models with PER, WS and BEPR as the dependent value include the personality trait extraversion, which had the highest significant correlation coefficient among the big five traits of the previously calculated correlations with these values and was therefore selected for the regression model. For all regressed models, the personality trait in combination with the college statistics also showed significance and thus functions as a predictor. Only for WS48 did conscientiousness achieve a higher significant correlation coefficient than extraversion and was therefore preferred in the model construction. Here, the attribute as an independent variable also showed significance within the model in combination with the college variables and is therefore a predictor of NBA performance. These results also support the findings of [Hardman \(1973\)](#) and [Piedmont *et al.* \(1999\)](#) regarding the associations between athlete performance and personality.

Furthermore, FG% and RPG are significant predictors for PER in combination with extraversion. This result can be related to the condition that higher FG%s are achieved by players who usually seek their finishes near the basket. These shots are considered easier but typically involve an increased physical component (including the battle for the rebound). As stated by [Hardman \(1973\)](#), people with increased extraversion are better able to withstand pain.

For WS, and similarly for BEPR only SPG and extraversion are significant predictors. The linkage between SPG and extraversion again can be explained by the findings made by [Hardman \(1973\)](#) as a steal is typically paired with a physical defense.

Examining the WS48 regression model, the results presented by [Ostendorf and Angleitner \(2004\)](#) can be used to connect the significant predictors of value. Focus herein is placed on the ability to plan, control, and execute actions, which is important for the success of high-performance athletes and is attributed to conscientiousness. Players who execute controlled and thoughtful actions in basketball tend to achieve higher FG%s because they make better shooting decisions.

Comparing the number of significant variables within the two models involving PER and WS48 as the dependent variables, PER shows more significant values. This can be explained by PER's all-encompassing concern with a player's offensive and defensive contributions, while WS48 is concerned with a player's contributions to winning and thus less concerned with the actual statistics captured.

7.3 Limitations

In terms of generalizability of the results, it must be stated that only players from draft years who played in college prior to the draft were examined, which limits the results. Others who played outside college leagues (e.g. in Europe, the G-League or Australia) and were drafted in the NBA draft are not included.

Furthermore, only players who have a valid Twitter account and have actively texted could be considered. It was also assumed that each player's Twitter account is verified, implying that players create their own messages and independently decide the content of their posts. With reference to Twitter, it must be noted that social media in particular can include impulsive texts that arise due to certain events or states of mind. Even though Twitter is considered one of the most studied social media systems and is known to predict the personality of the authors very accurately ([Adi *et al.*, 2018](#); [Ahmad and Siddique, 2017](#);

Golbeck *et al.*, 2011; Obschonka *et al.*, 2017; Quercia *et al.*, 2011), minimal variations in personality from the real personality cannot be ruled out. Various factors play a role, such as whether tweets are filtered by managers or staff, or whether the deletion of certain tweets is recommended by any third-party person, or whether players may adapt their communication unnaturally. Nevertheless, it can be said that the large number of tweets and the longer period of time are the main reasons why extreme personality traits are evident. In general, however, social media and especially Twitter should still be regarded as a reliable source for portraying one's own personality.

Another limitation of our study is the use of the commercial service IBM Personality Insights, which is often considered as a "black-box" as the functionality cannot be fully understood (the algorithms are not open source). Although the functionality has been demonstrated many times, alternative methods such as GloVe could be used to extract personality traits from tweets. Pennebaker and Francis (1996) laid the foundation for these methods with Linguistic Inquiry and Word Count (LIWC) (Chung and Pennebaker, 2012; Pennebaker *et al.*, 2015). Further improvements have been made with so-called word embeddings such as word2vec from Google or GloVe from Stanford (Pennington *et al.*, 2014). The advantage of such models is that semantic similarities between words are determined unsupervised, whereas LIWC relies on human assessment and psychologists to determine the meaning of words (Arnoux *et al.*, 2017; Rice and Zorn, 2021). IBM developed Watson Personality Insights in contrast, is a commercial software as a service for ready-to-use personality predictions based on GloVe (Arnoux *et al.*, 2017). The automated personality mining system can therefore be replaced at any time and could consequently influence the results, even if only minimally.

In addition, four individual NBA performance measures were evaluated and selected. However, the performance of a basketball player can be expressed in several other statistical values, based on the standpoint of the observer. Therefore, the results are limited to the statements on the selected metrics.

8. Conclusion and future research

The conducted research aimed to determine the ability to predict future performance of college basketball players in the NBA using college player statistics and personality profiles obtained from automated personality mining.

Based on our results, it can be stated that extraversion and conscientiousness are positively associated with all the performance values examined. The same relationship can be established for the performance measures PER, WS48, and BEPR and agreeableness. Openness to experience and neuroticism showed no significant correlation with future NBA performance. Therefore, players with higher scores in the range of extraversion, conscientiousness and agreeableness can achieve better performance in the NBA. Despite the correlations found, the results also show that a generalization of certain statements of previous studies cannot be fully adapted to the peculiarities of the NBA. This is evident in the case of neuroticism. The results in relation to college statistical values and NBA performance showed that higher college values in FG%, PPG, RPG and WS40 lead to stronger NBA performance.

The results of all regression models showed significant relationships between the NBA performance with at least one college statistic and one personality trait. In this context, extraversion and the college variables of FG% and RPG were found to be related to the NBA performance variable PER, while the performance value of WS48 can be related by the variables of conscientiousness and FG%. The factor of WS has proven to be related to the personality trait extraversion and the college statistic SPG. Similarly, BEPR reveals dependent effects on the variables of PPG, SPG and extraversion.

Accordingly, data from the present study show a relationship between college statistics in combination with measured personality traits obtained from automated personality mining and NBA performance. In summary, it can be concluded that future performance potential can be predicted from these measures. However, further research is needed to fully understand the prediction mechanism and to provide more specific information to NBA executives. Specifically, further studies should look at the differences in player positions, as these result in different performance statistics. In addition, our research only represents a simplification of human personality (i.e. view each trait separated) as we have not looked at how combinations of personality traits affect specific basketball performance measures. Future research should address this and use techniques such as machine learning, for example, to discuss more complex combinations and make predictions on players performance. Furthermore, the precise influences of the personality traits should be investigated and what direct influence they have on the player statistics in comparison with the previous performances from college. Here, structural equation modeling approaches can also be used to include various control variables that may exert influence. The constructed models must also be verified by a validation data set to detect possible weaknesses of the model. Furthermore, expert interviews with mental coaches, scouts and coaches of different teams should be conducted to validate the correlations found and to adjust the models accordingly.

Despite many evaluation options, such as scouting reports, workouts and performance tests, no team official can guarantee that the selected player will be a suitable fit for the NBA or the team roster. However, with our research, we were able to show how additional information using automated personality mining can contribute to the complex understanding of a basketball players success in the NBA.

References

- Adi, G.Y.N., Tandio, M.H., Ong, V. and Suhartono, D. (2018), "Optimization for automatic personality recognition on twitter in Bahasa Indonesia", *Procedia Computer Science*, Vol. 135, pp. 473-480.
- Ahmad, N. and Siddique, J. (2017), "Personality assessment using twitter tweets", *Procedia Computer Science*, Vol. 112, pp. 1964-1973, doi: [10.1016/j.procs.2017.08.067](https://doi.org/10.1016/j.procs.2017.08.067).
- Alfermann, D. and Stoll, O. (2017), *Sportpsychologie*, Meyer & Meyer Verlag, Aachen.
- Allen, M.S. and Laborde, S. (2014), "The role of personality in sport and physical activity", *Current Directions in Psychological Science*, Vol. 23 No. 6, pp. 460-465.
- Allport, G.W. (1961), *Pattern and Growth in Personality*, Holt, Reinhart & Winston, New York.
- Allport, G.W. and Odbert, H.S. (1936), "Trait-names: a psycho-lexical study", *Psychological Monographs*, Vol. 47 No. 1, p. i.
- Amabile, T.M. (1983), "The social psychology of creativity: a componential conceptualization", *Journal of Personality and Social Psychology*, Vol. 45 No. 2, pp. 357-376, doi: [10.1037/0022-3514.45.2.357](https://doi.org/10.1037/0022-3514.45.2.357).
- Arel, B. and Tomas, M.J. III (2012), "The NBA draft: a put option analogy", *Journal of Sports Economics*, Vol. 13 No. 3, pp. 223-249.
- Arnoux, P.H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R. and Sinha, V. (2017), "25 tweets to know you: a new model to predict personality with social media", *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11 No. 1, pp. 1-4.
- Atlas, M. and Zhang, Y.-Q. (2004), "Fuzzy neural agents for online NBA scouting", *IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)*, pp. 58-63.
- Badenhausen, K. (2020), "NBA Draft 2020: Projected Contracts for Edwards, Wiseman, LaMelo and Other First-Round Picks", *Forbes*, available at: <https://www.forbes.com/sites/kurtbadenhausen/2020/11/19/nba-draft-2020-projected-contracts-for-edwards-wiseman-lamelo-and-other-first-round-picks/>.

- Ballouli, K. and Sanderson, J. (2012), "It's a whole new ballgame: how social media is changing sports", *Sport Management Review*, Vol. 15 No. 3, pp. 381-382, doi: [10.1016/j.smr.2012.02.008](https://doi.org/10.1016/j.smr.2012.02.008).
- Baretta, D., Greco, A. and Steca, P. (2017), "Understanding performance in risky sport: the role of self-efficacy beliefs and sensation seeking in competitive freediving", *Personality and Individual Differences*, Vol. 117, pp. 161-165.
- Beech, R. (2009), "NBA draft analysis: expected value of a pick", available at: <https://www.82games.com/nbadraftpicks.htm>.
- Beene, A. (2019), "NBA draft decision-making using play-by-play data", *13th Annual Sports Analytics Conference*, Boston.
- Berger, T. and Daumann, F. (2021), "Anchoring bias in the evaluation of basketball players: a closer look at NBA draft decision-making", *Managerial and Decision Economics*, Vol. 42 No. 5, pp. 1248-1262.
- Berri, D.J. (2017), "National basketball association", in *Handbook of Sports Economics Research*, pp. 21-48.
- Berri, D.J., Brook, S.L. and Fenn, A.J. (2011), "From college to the pros: predicting the NBA amateur player draft", *Journal of Productivity Analysis*, Vol. 35 No. 1, pp. 25-35.
- Bipp, T. (2006), *Persönlichkeit, Ziele, Leistung: Der Einfluss der Big Five Persönlichkeitseigenschaften auf das zielbezogene Leistungshandeln*, Universität Dortmund, Frankfurt am Main.
- Bowman, J.K., Boone, R.T., Goldman, S. and Auerbach, A. (2021), "The athletic intelligence quotient and performance outcomes in professional baseball", *Frontiers in Psychology*, Vol. 12, p. 2489.
- Boyd, R.L. and Pennebaker, J.W. (2017), "Language-based personality: a new approach to personality in a digital world", *Current Opinion in Behavioral Sciences*, Vol. 18, pp. 63-68.
- Buettner, R. (2017), "Predicting user behavior in electronic markets based on personality-mining in large online social networks", *Electronic Markets*, Vol. 27 No. 3, pp. 247-265.
- Burke, K.L., Joyner, A.B., Pim, A. and Czech, D.R. (2000), "An exploratory investigation of the perceptions of anxiety among basketball officials before, during, and after the contest", *Journal of Sport Behavior*, Vol. 23 No. 1, pp. 11-19.
- Chung, C.K. and Pennebaker, J.W. (2012), "Linguistic inquiry and word count (LIWC): pronounced 'Luke,'... and other useful facts", in *Applied Natural Language Processing: Identification, Investigation and Resolution*, IGI Global, pp. 206-229.
- Coates, D., Oguntimein, B. and others (2010), "The length and success of NBA careers: does college production predict professional outcomes", *International Journal of Sport Finance*, Vol. 5 No. 1, pp. 4-26.
- Cohen, J. (2013), *Statistical Power Analysis for the Behavioral Sciences*, Routledge, New York.
- Craighead, D.J., Privette, G., Vallianos, F. and Byrkit, D. (1986), "Personality characteristics of basketball players, starters and non-starters", *International Journal of Sport Psychology*, Vol. 17 No. 2, pp. 110-119.
- dos Santos, V.G., Paraboni, I. and Silva, B.B.C. (2017), "Big five personality recognition from multiple text genres", in Ekštejn, K. and Matoušek, V. (Eds), *Text, Speech, and Dialogue. TSD 2017. Lecture Notes in Computer Science*, Vol. 10415, Springer, Cham, doi: [10.1007/978-3-319-64206-2_4](https://doi.org/10.1007/978-3-319-64206-2_4).
- Faliagka, E., Tsakalidis, A. and Tzimas, G. (2012), "An integrated e-recruitment system for automated personality mining and applicant ranking", *Internet Research*, Vol. 22 No. 5, pp. 551-568.
- Fearnhead, P. and Taylor, B.M. (2010), "Calculating strength of schedule, and choosing teams for March Madness", *The American Statistician*, Vol. 64 No. 2, pp. 108-115, doi: [10.1198/tast.2010.09161](https://doi.org/10.1198/tast.2010.09161).
- Ferrucci, D.A. (2012), "Introduction to 'this is watson'", *IBM Journal of Research and Development*, Vol. 56 No. 34, p. 1.
- Filo, K., Lock, D. and Karg, A. (2015), "Sport and social media research: a review", *Sport Management Review*, Vol. 18 No. 2, pp. 166-181, doi: [10.1016/j.smr.2014.11.001](https://doi.org/10.1016/j.smr.2014.11.001).

- Gaudreau, P. and Blondin, J.P. (2004), "Differential associations of dispositional optimism and pessimism with coping, goal attainment, and emotional adjustment during sport competition", *International Journal of Stress Management*, Vol. 11 No. 3, p. 245.
- Golbeck, J., Robles, C., Edmondson, M. and Turner, K. (2011), "Predicting personality from twitter", *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 2011 IEEE Third International Conference on, pp. 149-156.
- Groothuis, P.A., Hill, J.R. and Perri, T.J. (2007), "Early entry in the NBA draft: the influence of unraveling, human capital, and option value", *Journal of Sports Economics*, Vol. 8 No. 3, pp. 223-243.
- Hardman, K. (1973), "A dual approach to the study of personality and performance in sport", in *Personality and Performance in Physical Education and Sport*, Kimpton, London.
- Hogan, J. and Holland, B. (2003), "Using theory to evaluate personality and job-performance relations: a socioanalytic perspective", *Journal of Applied Psychology*, Vol. 88 No. 1, pp. 100-112.
- IBM Watson (2020), "Personality insights. Personality models", IBM Watson Developer Cloud, available at: <https://www.ibm.com/watson/developercloud/doc/personality-insights/models.html#outputBigFive>.
- Johnson, P.A. (1972), "A comparison of personality traits of superior skilled women athletes in basketball, bowling, field Hockey, and Golf", *Research Quarterly. American Association for Health, Physical Education and Recreation*, Vol. 43 No. 4, pp. 409-415, doi: 10.1080/10671188.1972.10615153.
- Judge, T.A., Heller, D. and Mount, M.K. (2002), "Five-factor model of personality and job satisfaction: a meta-analysis", *Journal of Applied Psychology*, American Psychological Association, Vol. 87 No. 3, pp. 530-541.
- Kian, E.M., Vincent, J. and Mondello, M. (2008), "Masculine hegemonic hoops: an analysis of media coverage of March Madness", *Sociology of Sport Journal*, Vol. 25 No. 2, pp. 223-242.
- Kubatko, J., Oliver, D., Pelton, K. and Rosenbaum, D.T. (2007), "A starting point for analyzing basketball statistics", *Journal of Quantitative Analysis in Sports*, Vol. 3 No. 3, pp. 1-22.
- Lázaro, M.B., Trillas, F. and Escuer, M.A.E. (2014), "Competitive balance in the NBA: comparative analysis of eastern and western conferences", available at: <https://www.semanticscholar.org/paper/COMPETITIVE-BALANCE-IN-THE-NBA-%3A-Comparative-of-and-L%C3%A1zaro-Trillas/c09ecfa357983b9f43dd4496a98138d1bc424243>.
- Li, B., Dittmore, S.W., Scott, O.K.M., Lo, W. and Stokowski, S. (2019), "Why we follow: examining motivational differences in following sport organizations on Twitter and Weibo", *Sport Management Review*, Vol. 22 No. 3, pp. 335-347, doi: 10.1016/j.smr.2018.04.006.
- Maddi, S.R. and Hess, M.J. (1992), "Personality hardiness and success in basketball", *International Journal of Sport Psychology*, Vol. 23 No. 4, pp. 360-368.
- Maese, R. (2018), How the NBA used twitter to dominate sports social media—the Washington post, Washington Post, available at: <https://www.washingtonpost.com/news/sports/wp/2018/05/31/nba-twitter-a-sports-bar-that-doesnt-close-where-the-stars-pull-up-a-seat-next-to-you/>.
- Matthews, T.D. and Lassiter, K.S. (2007), "What does the wonderlic personnel test measure?", *Psychological Reports*, Vol. 100 No. 3, pp. 707-712.
- McCrae, R.R. and Costa, P.T. (1982), "Self-concept and the stability of personality: cross-sectional comparisons of self-reports and ratings", *Journal of Personality and Social Psychology*, Vol. 43 No. 6, p. 1282.
- McCrae, R.R. and Costa, P.T. (1997), "Personality trait structure as a human universal", *American Psychologist*, Vol. 52 No. 5, p. 509.
- McCrae, R.R. and John, O.P. (1992), "An introduction to the five-factor model and its applications", *Journal of Personality*, Vol. 60 No. 2, pp. 175-215.
- Medina, M. (2020), "How NBA players are handling mental health issue during coronavirus crisis", USA TODAY, available at: <https://www.usatoday.com/story/sports/nba/2020/03/26/coronavirus-how-nba-players-handling-mental-health-during-hiatus/5076589002/>.
- Mirzaei, A., Nikbakhsh, R. and Shariffar, F. (2013), "The relationship between personality traits and sport performance", *European Journal of Experimental Biology*, Vol. 3 No. 3, pp. 439-442.

- Morgan, W.P. (1980), "The trait psychology controversy", *Research Quarterly for Exercise and Sport*, Vol. 51 No. 1, pp. 50-76.
- Morgan, W.P. (1985), "Selected psychological factors Limiting performance-a mental health model", *Limits of Human Performance*, pp. 70-80.
- Nisar, T.M., Prabhakar, G. and Patil, P.P. (2018), "Sports clubs' use of social media to increase spectator interest", *International Journal of Information Management*, Vol. 43, pp. 188-195, doi: [10.1016/j.ijinfomgt.2018.08.003](https://doi.org/10.1016/j.ijinfomgt.2018.08.003).
- Obschonka, M., Fisch, C. and Boyd, R. (2017), "Using digital footprints in entrepreneurship research: a twitter-based personality analysis of superstar entrepreneurs and managers", *Journal of Business Venturing Insights*, Vol. 8, pp. 13-23, doi: [10.1016/j.jbvi.2017.05.005](https://doi.org/10.1016/j.jbvi.2017.05.005).
- Oliver, D. (2004), *Basketball on Paper: Rules and Tools for Performance Analysis*, Potomac Books, Washington.
- Ostendorf, F. and Angleitner, A. (2004), *Neo-Persönlichkeitsinventar nach Costa und McCrae: Neo-PI-R: manual*, Hogrefe, Göttingen.
- O'Hallarn, B., Shapiro, S.L., Wittkower, D.E., Ridinger, L. and Hambrick, M.E. (2019), "A model for the generation of public sphere-like activity in sport-themed twitter hashtags", *Sport Management Review*, Vol. 22 No. 3, pp. 407-418, doi: [10.1016/j.smr.2018.06.001](https://doi.org/10.1016/j.smr.2018.06.001).
- Pennebaker, J.W. (2013), *The Secret Life of Pronouns: What Our Words Say about Us (Reprint Edition)*, Bloomsbury Press, London.
- Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K. (2015), The development and psychometric properties of LIWC2015.
- Pennebaker, J.W. and Francis, M.E. (1996), *Cognitive, emotional, and language processes in disclosure*, *Cognition and Emotion*, Vol. 10 No. 6, pp. 601-626.
- Pennington, J., Socher, R. and Manning, C.D. (2014), "Glove: global vectors for word representation", *EMNLP*, Vol. 14, pp. 1532-1543.
- Pfaffenberger, F. (2016), *Twitter als Basis wissenschaftlicher Studien: Eine Bewertung gängiger Erhebungs-und Analysemethoden der Twitter-Forschung*, Springer Nature, Heidelberg.
- Piedmont, R.L., Hill, D.C. and Blanco, S. (1999), "Predicting athletic performance using the five-factor model of personality", *Personality and Individual Differences*, Vol. 27 No. 4, pp. 769-777.
- Quercia, D., Kosinski, M., Stillwell, D. and Crowcroft, J. (2011), "Our twitter profiles, our selves: predicting personality with twitter", *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pp. 180-185.
- Quinn, S. (2020), "NBA will run entire pre-draft process, distribute information to teams, per report", MSN, available at: <https://www.msn.com/en-us/sports/nba/nba-will-run-entire-pre-draft-process-distribute-information-to-teams-per-report/ar-BB197xPb>.
- Ramezani, M., Feizi-Derakhshi, M.R. and Balafar, M.A. (2022), "Automatic personality prediction: an enhanced method using ensemble modeling", *Neural Computing and Applications*, doi: [10.1007/s00521-022-07444-6](https://doi.org/10.1007/s00521-022-07444-6).
- Reiter, B. (2020), "Why the 2020 NBA draft has put teams in uncharted territory with less scouting information than ever before", CBSSports.Com, available at: <https://www.cbssports.com/nba/news/why-the-2020-nba-draft-has-put-teams-in-uncharted-territory-with-less-scouting-information-than-ever-before/>.
- Rice, D.R. and Zorn, C. (2021), "Corpus-based dictionaries for sentiment analysis of specialized vocabularies", *Political Science Research and Methods*, Vol. 9 No. 1, pp. 20-35.
- Sailofsky, D. (2018), "Drafting errors and decision making theory in the NBA draft", available at: <https://dr.library.brocku.ca/handle/10464/13452>.
- Sarkar, M. and Fletcher, D. (2014), "Psychological resilience in sport performers: a review of stressors and protective factors", *Journal of Sports Sciences*, Vol. 32 No. 15, pp. 1419-1434.

- Saucier, G. and Goldberg, L.R. (1998), "What is beyond the big five?", *Journal of Personality*, Vol. 66, pp. 495-524.
- Sawant, P.V., Upadhyaya, N.S. and Berger, P.D. (2019), "Identifying future brand ambassadors in the national basketball association (NBA) for predicting future NBA superstars for superior marketing", *Journal of Economics and Business*, Vol. 2 No. 1, pp. 127-136.
- Schurr, K.T., Wittig, A.F. and Ruble, V.E. (1988), "Demographic and personality characteristics associated with persistent, occasional, and non-attendance of university male basketball games by college students", *Journal of Sport Behavior*, Vol. 11 No. 1, p. 3.
- Schwaborn, M. (2014), *Statistik im Basketball: Entwicklung, Forschung, Relevanz, Modelle, Aussagekraft und aktuelle Anwendungsmöglichkeiten*, AV Akademikerverlag, Chisinau.
- Siemon, D., Ahmad, R., Huttner, J.P. and Robra-Bissantz, S. (2018), "Predicting the performance of basketball players using automated personality mining", *Americas Conference on Information Systems*.
- Sohrabi, F., Atashak, S. and Aliloo, M. (2011), "Psychological profile of athletes in contact and non-contact sports", *Middle-East Journal of Scientific Research*, Vol. 9 No. 5, pp. 638-644.
- Stavros, C., Meng, M.D., Westberg, K. and Farrelly, F. (2014), "Understanding fan motivation for interacting on social media", *Sport Management Review*, Vol. 17 No. 4, pp. 455-469, doi: [10.1016/j.smr.2013.11.004](https://doi.org/10.1016/j.smr.2013.11.004).
- Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C. (2018), "Social media analytics – challenges in topic discovery, data collection, and data preparation", *International Journal of Information Management*, Vol. 39, pp. 156-168, doi: [10.1016/j.ijinfomgt.2017.12.002](https://doi.org/10.1016/j.ijinfomgt.2017.12.002).
- Sun, X., Liu, B., Cao, J., Luo, J. and Shen, X. (2018), "Who am I? Personality detection based on deep learning for texts", *2018 IEEE International Conference on Communications (ICC)*, pp. 1-6.
- Tausczik, Y.R. and Pennebaker, J.W. (2010), "The psychological meaning of words: LIWC and computerized text analysis methods", *Journal of Language and Social Psychology*, Vol. 29 No. 1, pp. 24-54.
- Teramoto, M., Cross, C.L., Rieger, R.H., Maak, T.G. and Willick, S.E. (2018), "Predictive validity of national basketball association draft combine on future performance", *The Journal of Strength and Conditioning Research*, Vol. 32 No. 2, pp. 396-408.
- Thomas, L. (2019), "The N.B.A.'s age of anxiety", *The New Yorker*, available at: <https://www.newyorker.com/sports/sporting-scene/the-nbas-age-of-anxiety>.
- Tok, S. (2011), "The big five personality traits and risky sport participation", *Social Behavior and Personality: An International Journal*, Vol. 39 No. 8, pp. 1105-1111.
- Tripathi, A.M. (2018), *Learning Robotic Process Automation: Create Software Robots and Automate Business Processes with the Leading RPA Tool-UiPath*, Packt Publishing, Birmingham.
- Tupes, E.C. and Christal, R.E. (1992), "Recurrent personality factors based on trait ratings", *Journal of Personality*, Vol. 60 No. 2, pp. 225-251.
- van der Aalst, W.M.P., Bichler, M. and Heinzl, A. (2018), "Robotic process automation", *Business and Information Systems Engineering*, Vol. 60 No. 4, pp. 269-272, doi: [10.1007/s12599-018-0542-4](https://doi.org/10.1007/s12599-018-0542-4).
- Wright, J. (2008), "Frequently-asked questions about the men's lacrosse rating percentage index", NCAA, available at: https://www.ncaa.org/sites/default/files/FAQ_for_MLAX_RPI.pdf.
- Zizzi, S., Deaner, H. and Hirschhorn, D. (2003), "The relationship between emotional intelligence and performance among college basketball players", *Journal of Applied Sport Psychology*, Vol. 15 No. 3, pp. 262-269.

Corresponding author

Dominik Siemon can be contacted at: dominik.siemon@lut.fi

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com