The New Window to Athletes' Soul – What Social Media Tells Us About Athletes' Performances

Arne Gruettner
University of St.Gallen
arne.gruettner@unisg.ch

Min Vitisvorakarn
University of St.Gallen
min.vitisvorakarn@student.unisg.ch

Thiemo Wambsganss
University of St.Gallen
thiemo.wambsganss@unisg.ch

Roman Rietsche
University of St.Gallen
roman.rietsche@unisg.ch

Andrea Back
University of St.Gallen
andrea.back@unisg.ch

Abstract

Professional sports has evolved from a game to an organization that has been codified, strategized, and commercialized. One factor that is shaping the sports industry is the pervasiveness of social media. On the one hand, social media is used as a powerful medium for distributing and getting news, engaging in topical discussions, and empowering brands. On the other hand, social media has become a crucial mouthpiece for athletes to interact with peers, share opinions, thoughts, and feelings. However, millions of followers, tweets, and likes later, researchers, practitioners, and athletes alike ask whether social media has an impact on an athlete's performance. We conducted a social media usage and a sentiment analysis of 124,341 Twitter tweets extracted from 31 tennis athletes. We linked these data to 8,095 corresponding match day performances. The results show that high social media usage has a significant negative impact on athletes' performance.

1. Introduction

In 1975, pupillometry pioneer and psychologist Eckhard Hess conducted a study to investigate the idea that changes in attitude can be detected by measuring changes in pupil size. Hess found that pupils are sensitive indicators of the mental state. This experiment later became known as the "window to your soul experiment", describing the fact that the best way to read someone's mind is to look into their eyes [1]. Today, personal contact to analyze peoples' behaviors and mental states is not necessary anymore, since people are sharing their opinions, thoughts, and feelings on social media platforms, such as Facebook, Instagram, and Twitter. For instance, about 2.1 million Snaps per minute are recorded by Snapchat users and 300 million tweets are tweeted on Twitter each day [2, 3]. As a result, what used to be the exclusive domains of friends and family members can now be accessed and analyzed by the public. 3.02 billion people, or about 40% of the world's population, will use social media platforms in 2021 [4].

One industry that has been strongly shaped by the use of social media in recent years is the sports industry [5, 6, 7]. On the one hand, clubs, fans, and sponsors are using social media as a powerful medium for distributing and getting the latest sports news, engaging in topical discussions and empowering brands. For example, the German sporting goods manufacturer Adidas already uses 90 percent of its marketing budget for digital campaigns and social media [8]. On the other hand, social media has also become a crucial mouthpiece for professional athletes to interact with stakeholders and peers (e.g., fans, sponsors, and celebrities) and, in turn, to build up a personal brand beyond the actual field of play. Consequently, almost every athlete has a social media profile in these days. It is even deemed as a necessity for athletes to maintain a social media presence [7, 9]. However, millions of followers, tweets, and likes later, with social media now at the heart of so many supposed sports scandals and athletes subject to vicious attacks that compromise their mental well-being, researchers, practitioners, and even athletes alike ask whether social media has an impact on athletes' match day performance. For instance, National Basketball Association (NBA) star Stephen Curry just recently challenged the impact of social media in an interview conducted by CNN: "*This is a new era in terms of, you know, the spotlight that every NBA athlete and any athlete in general is under. In terms of social media, in terms of the expectations put on us every single day*" [10].

Researchers have demonstrated in various theoretical as well as empirical contexts that social media can negatively impact user behavior (see for a literature review on social media's impact on user

behavior [11]). For instance, pervasive social media usage can result in technology dependency and excessive usage, which in turn can lead to negative outcomes for its users [12, 13]. However, only a few studies (e.g., [14]) have looked at the impact of social media on athletes' performance. Therefore, several researchers and practitioners call for studies that investigate how social media impacts athletes' performance (e.g., [7, 14, 15]). An understanding of the impact of social media on athletes' performance would not only be interesting for coaches, team managers, scouts, and bookmakers within the sports industry but also for researchers that investigate how information technology (IT) usage impacts human behavior. Consequently, this paper addresses the following two research questions (RQ): *RQ1) How does Twitter usage impact athletes' match day performance?* and *RQ2) How does athletes' pre-game mood extracted by Twitter sentiment polarity scores impact match day performance?*

Our research answers these research questions with the help of two evaluations: Firstly, by investigating the impact of low and high Twitter usage on athletes' match day performance and, secondly, by analyzing the Twitter sentiment polarity, which we claim is a proxy for athletes' pre-game mood. Specifically, we evaluate the performance of 31 professional tennis athletes. Our data set comprises 124,341 tweets. These social media data are linked to 8,095 corresponding match day performances. In future research, the proposed methodological approach consisting of our hypotheses, data set, and analytics techniques should be applied and extended to different sports disciplines.

The remainder of this paper is structured as follows: Section 2 introduces the theoretical background and related work. Our hypotheses are stated in Section 3. The methodology is described in Section 4. Section 5 presents our findings. We then discuss our findings and avenues for future research as well as limitations in Section 6. Finally, we conclude our paper in Section 7.

2. Theoretical Background

2.1. The Role of Social Media in Sports

Although social media has received significant attention from researchers of various fields in the last years, including Information Systems (IS), there is no common definition of social media in academia yet [11]. We define social media as being made up of various user-driven platforms (e.g., Facebook, WhatsApp, or Twitter) that facilitate diffusion of compelling content, dialogue, creation, and communication to a broader audience. Social media

provides an environment that is conductive for interactions and networking to occur at different levels (e.g., personal, professional, business, marketing, political, and societal) [11]. Social media offers the possibility to share user-generated content (UGC). UGC describes the various forms of digital content (e.g., posts, pictures, videos, or audio) created by end-users outside a professional context and which is publicly available [16]. A well-known example of UGC created by an athlete – which we refer to as athlete-generated content in this paper – is when the Italian soccer athlete Francesco Totti scored a goal in the so-called Roma derby, as part of the celebration took a selfie with the emotional, flag-waving supporters in the background, and posted the photo on Facebook. However, as the seemingly unbreakable boundary between the physical world and the online world in sports is getting blurry, it can be asked whether athletes' social media activity has an impact on the performance on the field [7, 14, 15]. On this matter, NBA commissioner Adam Silver believes that social media is doing more harm than good. Silver said during the MIT Sloan Sports Analytics Conference: "*I think we live a bit in the age of anxiety. [...] I think part of it is a direct product of social media*" [17].

Analytics and predictive techniques to investigate athletes' performances have commonly been used in the sports industry since the seminal work by Michael Lewis on the legendary story of Oakland Athletics, a Baseball Team that demonstrated how robust technical analysis based on performance data outperformed intuition and old-school wisdom [5]. Latest developments in digital technologies that led to the availability (and abundance) of statistical data have further stimulated interests in this area, which is commonly referred to as sports analytics [7]. However, the impact of athletes' social media activity on an athlete's performance has only received little attention in literature [7, 14, 15]. Although, for instance, with its unique features and its large user base, Twitter offers a rich data pool of athletes' communications, their opinions, their thoughts, and their feelings. Social media usage (e.g., the amount of Twitter tweets) and specific online athlete-generated content have great potential to become an information source for not only coaches, team managers, and scouts but also for bookmakers to discern athletes' mood status and shaky performance before matches. One of the few existing studies which investigated the impact of social media on athletes' performance is published by Xu and Yu [14]. In this study, a sentiment analysis of athlete-generated content on Twitter was used to detect NBA athletes' pre-match moods. The authors found that the mood of athletes has significant impact on driving athletes' on-court performances.

Similarly, Jones et al. demonstrated in their study that late-night tweeting is associated with a decreased next-day match performance for NBA athletes [18]. In our paper, we adopted and adjusted both studies to a professional tennis context. Therefore, we extended both studies by (1) investigating how low and high pre-match Twitter usage 36 hours prior to a match impacts athletes' match day performance and by (2) utilizing the Valence Aware Dictionary and Sentiment Reasoner (Vader) lexicon for our sentiment analysis, which is specifically designed for sentiments expressed in social media [19].

2.2. Linking Social Media and Performance

A major controversial topic of social media, which is currently not only discussed in the sports industry but also in political and entrepreneurial discussions, is its specific impact on users' performances. A topical example is the debate about the potentially negative impact of social media usage on employee productivity. As a consequence of this concern, many companies have banned social media platforms like Facebook and Twitter from the workplace [20]. In this paper, we define performance as the effectiveness with which people (i.e., athletes) perform activities that contribute to predefined goals and outcomes (i.e., sporting success) [21]. Existing research on the impact of social media on users' performances is two-fold: While a couple of studies have revealed the positive side of social media (e.g., [22, 23]) – for instance, the usage of LinkedIn can result in knowledge sharing and dissemination, or YouTube can be used for training and recruitment [11] – social media's negative impact on performance is evident in the majority of studies. These studies commonly report that pervasive social media usage can, for instance, result in technology dependency and excessive usage, which in turn can lead to negative outcomes for users' performances [16, 18, 24]. For example, too much communication via social media can cause information overload and distraction, confuse user's focus, and hamper the ability to make decisions [24].

In a sports context – as a practical example to illustrate how social media already impacts athletes' behaviors – Kliff Kingsbury, the head coach of the National Football League (NFL) club Arizona Cardinals, recently mandated "cell phone breaks" to browse social media during team meetings. It was created as a way to help those athletes who may lose focus during meetings itching to get back on their mobile device: "*You start to see kind of hands twitching and legs shaking and you know they need to get that social media fix, so we will let them hop over there and then get back in the meeting to refocus*" [17].

In this sense, the stress and pressure felt by today's athletes is arguably greater than ever. Contracts are bigger, stakes are higher. More people are watching, and, with social media, athletes are more visible than ever. Thus, not only the mere usage of social media can be a distraction for athletes. In particular, the content of others relating to the athlete may affect athletes' performances once read by the athlete.

2.3. Distraction-Conflict Theory

In order to explain social media's negative impact on users' performances, researchers commonly use distraction-conflict theory [25]. Distraction-conflict theory explains how distraction can create an attentional conflict or a situation in which an individual feels the tendency, desire, or obligation to allocate attention to multiple exclusive inputs [25, 26]. This, in turn, can lead to performance inefficiency. The distraction–conflict model can be broken down into three steps [16, 25]: (1) distraction, (2) distraction leads to attentional conflict, and (3) attentional conflict impacts the performance and motor behavior. Brooks, for example, tested the impact of social media usage on students' task performances by creating a classroom task environment [16]. Brooks found that higher amounts of social media usage led to lower performance on the task, as well as higher levels of technostress and lower happiness. Similarly, Cao and Yu investigated the effects of social media's different usage patterns on employee job performance [27]. Their results also showed that social media usage can generate conflict between technology use and work demand. This, in turn, negatively influences job performance.

Furthermore, distraction-conflict theory assumes that in settings in which there is pressure to perform a task quickly and well, attentional conflict can be created not only by external distractions but also by internal distractions, i.e., cognitive activity that is not directly relevant to the task solution but has an impact on it [25]. In the context of this paper, such internal distracting cognitions can include, for instance, the content of others read by an athlete on social media in advance of the match but which comes to the athlete's mind during the game. For example, a post from a user who claims that the athlete will never win a trophy because he is not cool enough in tight situations.

3. Hypothesis Development

3.1. Development of Hypothesis 1

Several studies have demonstrated in different contexts (e.g., [16, 24]) that increased social media

usage has a negative impact on performance (see Section 2.2.). Although the performance variable in a sports context differs from the examined contexts in these studies, we applied distraction-conflict theory to the context of our paper for the following reasons:

Firstly, studies have proven that social media usage can lead to distraction in both professional and non-professional settings (e.g., [16, 24, 27]). We believe that high Twitter usage before a match leads to distraction for athletes due to two reasons: (1) The focus and time spent on posting a tweet. As a consequence, the necessary focus on training, tactical preparation for the game, or even enough sleep is missing. (2) Athletes' awareness of the Twitter users' responses to their previously posted tweets and, in particular, the content of others relating to the athlete may trigger internal distractions (see, Section 2.3.) and thus affects athletes' match day performances. Secondly, prior work (e.g., [16, 24]) has utilized the distraction-conflict theory to investigate the impact of social media on cognitive tasks. However, distraction-conflict theory claims validity for both cognitive and motor tasks [25]. In a sports context, for instance, tennis strokes (e.g., a first serve) are understood as motor tasks [28]. We therefore see the distraction-conflict theory as an appropriate theoretical outline for the athletic setting.

Consequently, one should expect that high social media usage, expressed in the number of pre-match tweets, is associated with decreasing match day performance. This leads to the following hypothesis:

H1: *High pre-match social media usage has a negative impact on athletes' match day performance.*

3.2. Development of Hypothesis 2

We extended our research by taking the athlete-generated content of the tweets into account. Several studies have shown that tweets can be used to extract the collective [29] or individual mood (e.g., [14, 30]). For instance, Golder and Macy applied sentiment analysis to Twitter data to extract emotional content from 509 million tweets of 2.4 million users in 84 different countries [30]. Using this information, the authors were able to identify daily and seasonal individual mood variations in cultures across the world.

Extrapolating from this, athlete-generated content can be used as a proxy to determine athletes' pre-match mood which, in turn, can impact athletes' performance on the field. In this sense, Totterdell analyzed the relationship between mood and performance for cricket athletes in an offline context [31]. The results indicated a correlation between

individual performance and the collective mood of teammates. Likewise, Xu and Yu found a positive association between NBA athletes' individual mood, which has been extracted from pre-match tweets, and their match performance [14]. As our focus also lies on the individual athlete's performance, we expect similar results in our setting. This results in the following hypothesis:

H2: *A positive pre-match mood has a positive impact on athletes' match day performance.*
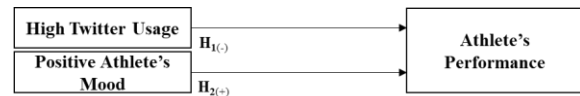
Our research model is depicted in Fig. 1.

Figure 1. Research model

4. Methodology

In this paper, we used social media and performance data of professional tennis athletes to investigate (1) the impact of low and high social media usage on athletes' performance and to analyze (2) whether we can use sentiment polarity scores to detect athletes' pre-match mood, which in turn has an impact on athletes' match day performance. We chose professional tennis athletes for our analysis for two reasons: Firstly, due to the availability of very high-quality performance data recorded at professional sports competitions as well as huge amounts of available social media data and, secondly, due to the nature of tennis as a single sport, which guarantees that no external team factors affect individual performance. To validate our hypotheses, we followed an adjusted knowledge discovery process in databases (KDD) as described by Fayyad et al. [32]. KDD describes the structured extraction of useful information from a large volume of data. For the purpose of our paper, KDD's iterative cycle approach was combined and formed into three steps: (1) data selection, (2) data preprocessing and transformation, and (3) data mining and evaluation, which are explained and applied in the following Subsections.

4.1. Data Selection

The first step (data selection) of our adjusted KDD included the selection of data sources. In this paper, we aimed to acquire Twitter tweets from athletes participating in the Association of Tennis Professionals (ATP) and the Women's Tennis Association (WTA). We chose Twitter for our study as athletes are able to display behavioral traits on

Twitter [14]. We used the python-based library Twint to collect the entire Twitter timeline of a given athlete (see Fig. 2. for the include/exclude criteria for the selection of athletes). To narrow down the selection of athletes, in a first step, we started with the current Top 150 ATP and WTA athletes (as of April 19, 2019). We then included and excluded certain athletes based on chosen criteria, e.g., athletes which posted less than 2,500 tweets, resulting in 31 athletes for further analysis. In total, 124,341 tweets from 31 tennis athletes were collected, with 12 being ATP (i.e., male) athletes and 19 being WTA (i.e., female) athletes. The match performance data for each examined athlete was collected by using open source data sets from tennis-data.co.uk, tennis.wettpoint.com and a GitHub data repository [33, 34, 35]. Our final data set for our analysis is publicly available and can be accessed via the link below.[1]
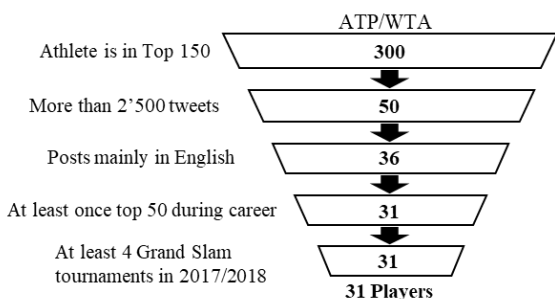

Figure 2. Include and exclude criteria for athlete selection (as of April 19, 2019)

To test the developed hypotheses H1 and H2, we generated two distinct data sets for each athlete – (1) a Twitter data set and (2) a match performance data set. The Twitter data set contained the full tweet text in UTF-8 encoding with emoticons being displayed. Additionally, metadata such as the post time and the username were included. We excluded retweets from our analysis as retweets did not contain useful information for our analysis. The match performance data set included all ATP and WTA tournament singles matches between 2012 and 2018 from the chosen athletes. For an overview of ATP/WTA tournaments, see [35]. In total, the match performance data set counted 9'340 matches before preprocessing and cleaning the data. We included metadata such as the result, the match date and time, the age as well as the rank of the athlete in our data set. As a performance variable, we chose the first serve fault. The first serve fault indicates the percentage of serves missed and consequently leads to a second serve for a given

athlete. We chose the first serve fault as our performance indicator, as this metric is less dependent on the opponent's strength than aces, unforced errors, or break points saved. Thus, the first serve fault displays an athlete's individual primary task performance and can be theoretically linked to the distraction-conflict theory.

4.2. Data Preprocessing and Transformation

After extracting and collecting relevant data, the next step was data preparation. We removed three outliers which posted more than 50 tweets before the match as well as matches with incomplete performance data. In total, we deleted 1,245 matches from our sample. Resulting in a total of 8,095 matches. Then, the preprocessed data got transformed into a structured format that can be accessed and processed for further analysis. We tested H1 by conducting a paired-samples t-test between two groups of social media usage (low and high). To test H2, we first performed a sentiment analysis, then we clustered the tweets into two mood groups (non-positive and positive mood) in accordance to their polarity score as this study focuses only on the impact of a positive mood of athletes on their performance. This step was followed by a paired-samples t-test as well. Thus, the data was preprocessed and transformed to meet the purposes of these two analyses. Preprocessing and data analytics techniques were performed using the programming language Python 2.73, since it is widely known, easy to use, and supports major libraries for Natural Language Processing (NLP) tasks [36].

First, the extracted 124,341 tweets were imported into a Pandas data frame. Then, the text of our tweets was transformed into lower case. Regular expressions were used to filter punctuation (e.g., "?" - except "!") and numbers, since they were not relevant for the sentiment of a tweet. Furthermore, hyperlinks, mentions of other Twitter users and hashtags were removed and substituted with an acronym (e.g., "www.twitter.com/..." was substituted with "URL", or "@UserXY" was substituted with "AT_USER"). However, after applying further libraries we decided to not tokenize and stem our tweets, since the applied sentiment library performed with more accurate results with the full strings.

4.3. Data Mining and Evaluation

4.3.1. Evaluation of Hypothesis 1. The next step of our adjusted KDD included the data mining and

evaluation phase. To test our hypothesis, we linked the Twitter data set to the generated performance data set. We defined a time span of 36 hours prior to a match as the pre-match time span in accordance with [14], who tested the causality relation between different time spans and performance. Therefore, we used the time stamp of the Twitter tweets as well as the concrete start time of the match. Differences due to time shifts were taken into account. Finally, 13,954 tweets out of the totally extracted 124,341 tweets were linked to 8,095 corresponding match day performances. We believe that a higher number of pre-match tweets within this 36-hour time span represents a longer time spent on Twitter. This, in turn, is expected to distract athletes in accordance with the distraction-conflict theory (see, Section 3.1.). If an athlete started her/his career or signed up for Twitter after 2012, the aggregation process started at the earliest date possible. The same methodology was also applied for the sentiment analysis to test H2.

Finally, the impact of high Twitter usage on athletes' performance (H1) was tested by carrying out a two-tailed paired-samples t-test between two groups of social media usage (i.e., low and high). In detail, each tennis athlete's mean performance was compared between matches with low Twitter usage and high Twitter usage. We labeled a data point as low Twitter usage if the athlete posted between zero and four tweets before a match day and as high Twitter usage if the pre-match tweets exceeded four tweets. This classification resulted from calculating the average pre-match Twitter usage and adding one standard deviation to the value in order to define the threshold for high Twitter usage (see, Tab. 2.).

4.3.2. Evaluation of Hypothesis 2. To test H2, we aimed to analyze the Twitter sentiment polarity, which we claim is a proxy for the athletes' pre-game mood. Therefore, we used a dictionary-based method from Natural Language Toolkit (NLTK), that is, the Vader lexicon [19], to extract sentiments from our tweets. Based on this sentiment analysis, polarity scores were retrieved that correspond to athletes' pre-game mood. Scores were labeled between -1 and 1 according to a "positive", a "negative", and a "neutral" mood. A tweet which had a polarity score above zero was classified as "positive". Tweets which showed a score below zero were classified as "negative". All other tweets that had a score that equaled exactly zero were labelled with "neutral". For an example of Vader's polarity score and the resulting pre-game mood, see Tab. 1.

We defined a 36-hour time span prior to a match day as our investigation period to investigate the impact of athletes' mood. The length of this time span ensured that the pre-match mood of a specific match did not conflict with the mood of the precedent match as tennis athletes usually play in a two- to three-day rhythm. In contrast to the evaluation for H1, we deleted all matches in which an athlete did not post a tweet from our data set for further analysis, as these data points did not contain any information on athletes' mood. As a result, the number of matches to test H2 was reduced from 8,095 matches to 4,686 matches. In a last step, we aggregated all polarity scores occurring in the defined time span for the corresponding match day and calculated the mean polarity score. Afterwards, we merged all pre-match modes which were labeled as "negative" or "neutral" according to the results of the sentiment analysis into one group, that is, the "non-positive" pre-match mood group. In a last step, to test whether a positive pre-match mood has a positive impact on athletes' match day performance, we applied a two-tailed paired-samples t-test.

Table 1. Sample tweets showing Vader mood classification

Tweet Text	Tennis Athlete	Polarity Score	Mood/ Label
i don't believe in luck... but if i did mine would be categorized as bad and terrible.	Serena Williams	-.8737	Negative
just woke up and preparing my breakfast now...	Julia Goerges	0	Neutral
happy, happy, happy, happy .. i think i'm happy qualified for the maindraw.. damn feels good thanks for your support	Kirsten Flipkens	.9758	Positive

5. Results

5.1. Descriptive Statistical Results

The final data set to test H1 comprised 13,682 tweets, which were linked to 8,095 corresponding match day performances for the predefined cut-off time of 36 hours. To test H2, these 8,095 match day performances were reduced to 4,686 data points due to matches in which athletes did not post a tweet as described in Section 4. For an overview of the descriptive statistics of our data sample, see Tab. 2. For an overview of the Twitter usage with the corresponding match performance, see Fig. 3. The relation between athletes' mood and performance is depicted in Fig. 4.

Table 2. Descriptive statistics for social media usage, sentiment analysis, and performance

	Social Media Usage		Sentiment Analysis		Performance	
Label	**Low Usage**	**High Usage**	**Non-Positive Mood**	**Positive Mood**	**First Serve Fault**	
Number of Matches	7,363	732	1,332	3,354	8,095	
Total Matches	8,095		4,686		8,095	
Number of Tweets	13,682*		13,954		-	
Demographics	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Total	1.690	3.117	.296	.306	.379	.078
Gender *Female (n=19)*	1.511	2.477	.294	.309	.384	.081
Male (n=12)	1.893	3.702	.298	.302	.373	.075
Age *Experienced*	1.610	3.276	.284	.307	.378	.078
Inexperienced	1.914	2.613	.322	.301	.386	.079
*As described in Section 4.2., for the social media usage analysis the number of tweets is reduced due to the removal of outliers.						

On average, the identified athletes posted 1.69 tweets per match ($\sigma = 3.12$). Male athletes posted 1.89 pre-match tweets per match on average ($\sigma = 3.70$), whereas female athletes posted 1.51 tweets per match on average ($\sigma = 2.48$). The highest number of pre-match tweets posted by a male athlete was 40 (two posts) tweeted by Stanislas Wawrinka. Likewise, the highest number of pre-match tweets posted by a female athlete, that is Serena Williams, was 40. Taking different age groups into account (i.e., athlete's age on match day) and defining experienced athletes as any athlete over 24 years of age, we identified a slight difference between experienced and inexperienced athletes. On average, experienced athletes posted 1.61 pre-match tweets per match ($\sigma = 3.28$), while inexperienced athletes posted 1.91 pre-match tweets per match ($\sigma = 2.61$). Looking at the performance data, the average for the first serve fault for male and female athletes was 0.38 ($\sigma = 0.08$). On average, the first serve fault for male athletes was 0.37 ($\sigma = 0.08$), while female athletes' first serve failed 0.38 on average ($\sigma = 0.08$).
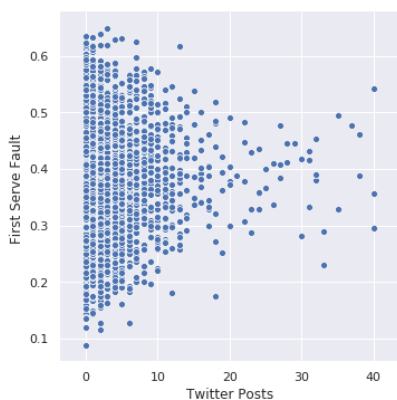

Figure 3. Twitter usage and performance

As described above, the data set to test H2, including our sentiment analysis, comprised 4'686 matches. The average polarity score for the identified athletes counted 0.30 ($\sigma = 0.31$), which indicated that most tweets were associated with a positive athletes' mood. Splitting the data set into male and female athletes, male athletes average a polarity score of 0.30 ($\sigma = 0.30$), whereas female athletes average a polarity score of 0.29 ($\sigma = 0.31$). Comparing experienced and inexperienced athletes in the same way as described above, experienced athletes expressed a polarity score of 0.28 ($\sigma = 0.31$) and inexperienced athletes scored 0.32 on average ($\sigma = 0.30$), indicating that inexperienced athletes tweeted more positively than experienced athletes.
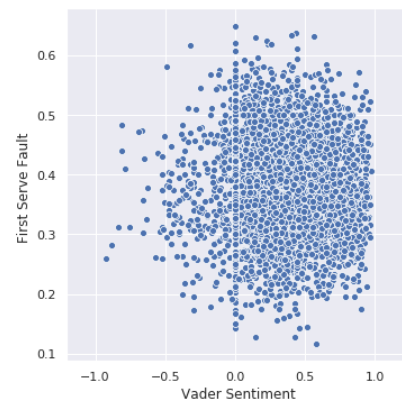

Figure 4. Pre-game mood and performance

5.2. Impact of High Social Media Usage on Performance

H1 states that high social media usage (i.e., an athlete posted five or more tweets before a match day) has a negative impact on athletes' match day performance. The performed two-tailed paired-samples t-test between the low Twitter usage group and high Twitter usage group showed that the latter had a negative impact on athletes' match day performance in contrast to low Twitter usage. This difference was significant at $p < 0.05$. On average, tennis athletes performed 1.011 percentage points worse in terms of delivering the first serve when having high Twitter usage before a match.

Consequently, as a result of our analysis, H1 can be validated. Tab. 3 summarizes the results from the two-tailed paired-samples t-test.

Table 3. The impact of high social media usage on match performance

Variable	Low Twitter Usage	High Twitter Usage	Diff.	t
First Serve Fault	.3821	.3922	-.0101	-2.270*
Significance with * = p < .05 / ** = p < .01 / *** = p < .001				

5.3. Impact of Athletes' Mood on Performance

To test whether a positive pre-match mood has a positive impact on athletes' match day performance (H2), we also applied a two-tailed paired-samples t-test between the non-positive mood group and the positive athlete mood group. The results of our analysis showed no significant difference ($p > 0.05$). We repeated the same analysis for different polarity score thresholds, e.g., classifying polarity scores above 0.25 as positive mood, but no significant results were found. We believe that state-of-the-art sentiment lexica such as Vader still need to be trained and refined in order to deliver strong results for detecting the individual mood of athletes. Therefore, H2 has to be rejected. Tab. 4 indicates the results from the two-tailed paired-samples t-test.

Table 4. The impact of athletes' moods on match performance

Variable	Non-Positive Mood	Positive Mood	Diff.	t
First Serve Fault	.3813	.3849	-.0036	-1.166
Significance with * = p < .05 / ** = p < .01 / *** = p < .001				

6. Discussion, Future Research, and Limitations

To investigate how social media usage (RQ1) and athletes' pre-game mood (RQ2) impact athletes' match day performance, we conducted a social media usage as well as a sentiment analysis of 31 tennis athletes and extracted 124,341 tweets. We linked these data to 8,095 corresponding match day performances. To test our hypothesis that high social media usage has a negative impact on athletes' performance (H1), we counted the number of tweets that were tweeted 36 hours before match day by an individual athlete and formed two groups of Twitter usage (low and high). The results of the two-tailed paired-samples t-test confirmed our hypothesis and demonstrated that high social media usage has a significant ($p < 0.05$) negative impact on athletes' performance. Hence, our results support the concerns and fears of practitioners, researchers, and athletes alike that social media usage can have a negative impact on athletes' performances. By conducting a sentiment analysis, we identified polarity scores for each tweet, which were used as indicators for athletes' pre-match mood to test our hypothesis that a positive athletes' pre-match mood has a positive impact on athletes' performance (H2). Although there are studies which have proven a positive relationship between athletes' pre-match mood and their on-field performance in offline and online contexts (e.g., [14, 31]), our analysis showed no significant results ($p = 0.253$) that an athlete with a positive pre-match mood performed better on the field than an athlete who did not show a positive pre-match mood in her/his tweets. Therefore, we must reject H2.

From an academic perspective, our paper makes several contributions to the existing body of literature. By answering our first research question, we contribute to IS research by addressing the call made by several researchers for the investigation of how social media usage impacts athletes' performance (e.g., [7, 14, 15]). While our results demonstrate that high social media usage can have a significant negative impact on athletes' performances in the context of this study, our research model should be applied to different contexts. It can be expected that depending on, for instance, the type of sports (i.e., individual vs. team sports) results will differ. Likewise, it should also be investigated whether there are individual athletes for whom social media usage has a positive impact. In addition, we contribute to the upcoming literature stream of sports analytics by providing an explanation of how distraction-conflict theory can be applied to explain the impact of social media usage on athletes' performance. Moreover, we contribute to the body of knowledge by applying distraction-conflict theory in a new setting, that is, athletes' performances. Most existing studies on the impact of social media usage on performance investigate how social media usage distracts users' cognitive tasks at the concrete moment the task is actually performed (e.g., [16, 24]). We enrich existing findings by investigating how internal distraction impacts performance, not only at the actual moment of activity but also over a longer time span (i.e., 36-hour time span) for motor tasks. We further contribute to research as well as practice by making our data set publicly available.

To answer our second research question, we conducted a sentiment analysis using the Vader lexicon. We think that the Vader lexicon is inadequate for the analysis of athlete-generated content, although it was developed specifically for the analysis of social media data, since athletes use a very specific language

in their tweets. Social media offers a rich data pool of athletes' communications, their opinions, their thoughts, and their feelings. Hence, we believe it has great potential to become a relevant information source for future research. Therefore, our research contributes to the literature stream of sports analytics by representing one of the early efforts from the IS field to bring together athlete-generated content and analytics techniques to investigate athletes' pre-match mood. Our paper analyzes the impact of social media in a sports context. We argue that the sports industry provides many opportunities for comprehending IT-driven phenomena that might display interesting dynamics due to the uniqueness of the context at a theoretical level [6, 7, 37]. Thus, dismissing the sports industry as just another empirical context will translate to missed opportunities. Therefore, this paper also contributes to the IS literature in general by examining how social media impacts human behavior.

From a practical perspective, our results provide important insights into the impact of social media on athletes' match day performance. Thus, they are especially relevant for the stakeholders of the sports industry. For instance, coaches can identify athletes who use social media excessively. In turn, they can adjust the line-up on match day accordingly. Similarly, team managers and scouts can use our results to determine which athlete they want to hire based on their social media usage and athlete-generated content. Likewise, bookmakers can adjust their calculation of odds based on the social media activity of athletes. Regarding this, the development of an interactive live dashboard that displays various key information to stakeholders could be interesting in the future. Finally, our results should also be seen as a wake-up call for athletes themselves, inspiring them to reduce their social media activities and to focus on their performance on match days.

This paper provides numerous opportunities for future research: For instance, researchers can investigate whether different types of social media activities (e.g., listening, posting, reading, or watching) have a different impact on athletes' performances. Similarly, it would be interesting to explore how various communications via social media (e.g., communications with peers vs. fans) impact athletes' match day performances. To extend our research, an analytical model can be built to predict athletes' performances. Furthermore, our research model can be extended to various other social media platforms such as Instagram using, for instance, image recognition. Future research can also focus on the development of a more specific and comprehensive lexicon which can be applied to the context of professional sports athletes.

This paper is not free from limitations: Firstly, our performance variable (i.e., first serve fault) is not completely independent from other factors, such as age, current score, match importance, or opponent. Secondly, we believe that the lack of control variables (such as post length) is a limitation. Thirdly, we only investigated the negative impact of social media usage. It should also be investigated whether there are athletes on whom social media has a positive effect. Fourthly, it is reasonable to assume that the distilled moods from athletes' tweets are not free of bias, which could be attributed to a number of factors. For instance, professional athletes often use agencies to maintain their social media profiles in these days. Lastly, our data set consists of 31 athletes only. Thus, an extension of this study's scope can become relevant.

7. Conclusion

Based on a social media usage and a sentiment analysis, we investigated the impact of (1) low and high social media usage and of (2) athletes' pre-match mood on professional athletes' match day performance. For research, the conducted analysis provides theoretical evidence that high social media usage has a significant negative impact on athletes' performance. Future research should focus on different types of social media activities (e.g., posting or reading) or on various types of communications (e.g., communications with peers vs. fans) and their specific impact on athletes' performance. Implications for practitioners are insights into the impact of social media on athletes' performances. Therefore, supporting them in specific decision-making processes (e.g., line-up adjustments on match days).

References

[1] Hess, E.H., "The Role of Pupil Size in Communication", *Scientific American 233*(5), 1975, pp. 110–119.

[2] Leetaru, K., "A Fading Twitter Changes Its User Metrics Once Again", *Forbes*, 2019. https://www.forbes.com/sites/kalevleetaru/2019/04/23/a-fading-twitter-changes-its-user-metrics-once-again

[3] Statista, "User-generated internet content per minute 2019", 2019. https://www.statista.com/statistics/195140/new-user-generated-content-uploaded-by-users-per-minute/

[4] Statista, "Number of social media users worldwide 2010-2021", 2019. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

[5] Davenport, T.H., "What Businesses Can Learn From Sports Analytics", *MIT Sloan Management Review 55*(4), 2014, pp. 10–13.

[6] Gruettner, A., "What We Know and What We Do Not Know About Digital Technologies in the Sports Industry", *AMCIS 2019 Proceedings*, (2019).

[7] Xiao, X., J. Hedman, F.T.C. Tan, et al., "Sports Digitalization: An Overview and A Research Agenda", *ICIS 2017 Proceedings*, (2017).

[8] Klingelhoefer, C., "Adidas Goes After Nike with Digital Strategy", *ISPO DIGITIZE*, 2018. https://www.ispo.com/en/companies/adidas-goes-after-nike-digital-strategy

[9] Tan, F.T.C., J. Hedman, and X. Xiao, "Beyond 'Moneyball' to Analytics Leadership in Sports: An Ecological Analysis of FC Bayern Munich's Digital Transformation", *AMCIS 2017 Proceedings*, (2017).

[10] CNN, "Is social media harming sports stars?", 2019. https://edition.cnn.com/videos/sports/2019/04/23/nba-nfl-social-media-tiffany-jones-adam-silver-steph-curry-spt-intl.cnn

[11] Kapoor, K.K., K. Tamilmani, N.P. Rana, P. Patil, Y.K. Dwivedi, and S. Nerur, "Advances in Social Media Research: Past, Present and Future", *Information Systems Frontiers 20*(3), 2018, pp. 531–558.

[12] Lowry, P.B., J. Zhang, C. Wang, and M. Siponen, "Why do adults engage in cyberbullying on social media?", *Information Systems Research 27*(4), 2016, pp. 962–986.

[13] Turel, O., and H. Qahri-Saremi, "Problematic Use of Social Networking Sites: Antecedents and Consequence from a Dual-System Theory Perspective", *Journal of Management Information Systems 33*(4), 2016, pp. 1087–1116.

[14] Xu, C., and Y. Yu, "Measuring NBA players' mood by mining athlete-generated content", *HICSS 2015 Proceedings*, (2015).

[15] DiMoro, A., "The Growing Impact of Social Media On Today's Sports Culture", *Forbes*, 2015. https://www.forbes.com/sites/anthonydimoro/2015/07/02/the-growing-impact-of-social-media-on-todays-sports-culture

[16] Brooks, S., "Does personal social media usage affect efficiency and well-being?", *Computers in Human Behavior 46*, 2015, pp. 26–37.

[17] York, J., "Is this generation of sports stars the 'loneliest' ever?", *CNN*, 2019. https://edition.cnn.com/2019/04/23/sport/nba-social-media-tiffany-jones-spt-intl/index.html

[18] Jones, J.J., G.W. Kirschen, S. Kancharla, and L. Hale, "Association between late-night tweeting and next-day game performance among professional basketball players", *Sleep Health 5*(1), 2019, pp. 68–71.

[19] Hutto, C.J., and E.E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text", *ICWSM-14 Proceedings*, (2014).

[20] Bizzi, L., "Should HR managers allow employees to use social media at work?", *The International Journal of Human Resource Management 1*, 2017, pp. 1–28.

[21] Borman, W.C., and S.J. Motowidlo, "Task performance and contextual performance: The meaning for personnel selection research", *Human Performance 10*(2), 1997, pp. 99–109.

[22] Utz, S., and J. Breuer, "The relationship between use of social network sites, online social support, and well-being", *Journal of Media Psychology 29*(3), 2017, pp. 115–125.

[23] Dienlin, T., P.K. Masur, and S. Trepte, "Reinforcement or Displacement? The Reciprocity of FtF, IM, and SNS Communication and Their Effects on Loneliness and Life Satisfaction", *Journal of Computer-Mediated Communication 22*(2), 2017, pp. 71–87.

[24] Mansi, G., and Y. Levy, "Do instant messaging interruptions help or hinder knowledge workers' task performance?", *International Journal of Information Management 33*(3), 2013, pp. 591–596.

[25] Baron, R.S., "Distraction-conflict theory: Progress and problems", *Advances in Experimental Social Psychology 19*, 1986, pp. 1–40.

[26] Baron, R.S., D. Moore, and G.S. Sanders, "Distraction as a source of drive in social facilitation research", *Journal of Personality and Social Psychology 36*(8), 1978, pp. 816–824.

[27] Cao, X., and L. Yu, "Exploring the influence of excessive social media use at work: A three-dimension usage perspective", *International Journal of Information Management 46*, 2019, pp. 83–92.

[28] Wolpert, D.M., and J.R. Flanagan, "Motor learning", *Current Biology 20*(11), 2010, pp. 467–472.

[29] Bollen, J., H. Mao, and X. Zeng, "Twitter mood predicts the stock market", *Journal of Computational Science 2*(1), 2011, pp. 1–8.

[30] Golder, S.A., and M.W. Macy, "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures", *Science 333*(6051), 2011, pp. 1878–1881.

[31] Totterdell, P., "Catching moods and hitting runs: Mood linkage and subjective performance in professional sport teams", *Journal of Applied Psychology 85*(6), 2000, pp. 848–859.

[32] Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth, "Knowledge discovery and data mining:towards a unifying framework", *KDD 1996 Proceedings*, (1996).

[33] Sackmann, J., "Tennis ATP - GitHub", *GitHub repository*, 2019. https://github.com/JeffSackmann/

[34] Wettpoint.com, "Tennis Statistics Wettpoint", 2019. https://tennis.wettpoint.com/en/

[35] Tennis-data.co.uk, "tennis-data.co.uk", 2019. http://www.tennis-data.co.uk/alldata.php

[36] Bird, S., E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly, 2009.

[37] Chiasson, M.W., and E. Davidson, "Taking Industry Seriously in Information Systems Research", *MIS Quarterly 29*(4), 2005, pp. 591–606.