Publication III

From Data to Dollar –
Using the Wisdom of an Online Tipster Community
to Improve Sports Betting Returns

Author(s)

Arne Grüttner

Thiemo Wambsganss

Andrea Back

In press for publication in the European Journal of International Management (EJIM) – Special Issue on: "International Sports Management".

Abstract

With thousands of (online) bookmakers accepting wagers on sporting events, sports betting has become a billion-dollar business worldwide. Therefore, researchers and practitioners have gathered interest in investigating the "wisdom-of-crowds" effect in online tipster communities to predict the outcomes of sports events. We extracted 1,534,041 tips of 3,484 tipsters from Blogabet.com and used this user-generated content to investigate whether there is wisdom in online tipster communities that can be used to improve betting returns. We applied state-of-the-art data mining and natural language processing techniques and tested our hypotheses using quantitative research methods. Our results demonstrate that there is indeed wisdom in such online tipster communities that can improve sports betting returns. Tipsters won 3.29% more tips than the implied win probability set by bookmakers and produced averaged yields of 3.97%. We further identified four characteristics that are significant indicators for smarter sub-crowds within the overall crowd of an online tipster community.

Keywords: data mining; natural language processing; online communities; sports betting; sports innovation; user-generated content; wisdom-of-crowds

Introduction

Sports betting has become a billion-dollar business around the globe. In Germany, for example, each day people place bets worth EUR 25 million. An increase of 21 percent over the previous year (Deutscher Sportwettenverband, 2020). Online tipster communities, such as Betadvisor.com, Blogabet.com, or Oddsportal.com, offer semi-professional sports bettors, so-called tipsters, the opportunity to publish, share and explain their carefully elaborated tips over the internet. Community members, on the other hand, can comment and discuss those publicly available tip recommendations. Online tipster communities can be seen as a new type of sports-based entrepreneurship (see, Ratten (2011)) driven by the emergence and rise of innovative digital technologies in the sports industry that cover a wide range of sports and have a lively community (Gruettner, 2019; Ratten, 2017). Considering the forecasting power (e.g., in terms of diverse knowledge and expertise), recent academic studies published in scholarly journals have shown interest in investigating the underlying dynamics of online tipster communities (e.g., Brown and Reade (2019)). In this vein, the user-generated content of tipsters – which we will refer to as tipster-generated content (TGC) in this study – offers the potential to become a revealing data source to improve sports betting returns.

TGC has proven to be valuable in predicting the outcomes of sports events, as it not only contains concrete predictions of match results but also (often) background information about the tipsters or even detailed textual match analyses. Existing studies have been published on the "wisdom-of-crowds" effect of Surowiecki (2004) (e.g., Brown and Reade (2019), O'Leary (2017), Peeters (2018), or Schumaker et al. (2016)). The wisdom-of-crowds effect operates on the premise that an averaging of forecasts eliminates individual prediction errors and thus leads to greater accuracy. In other words, large groups of individuals are better at making predictions than individuals are. The effect has tremendous practical implications: First, it suggests that decisions made by the majority rule (or by averaging opinions) will outperform decisions made by single experts. Second, it suggests that decisions made by the majority rule will often be accurate in an absolute sense – an implication that partially accounts for the rapidly increasing use of information markets to predict events (Simmons et al., 2011).

TGC can be seen as a valuable source to extract the wisdom of an online tipster community to predict the outcomes of sports events. However, existing studies on TGC and the wisdom-of-crowds effect come with several shortcomings: First, existing studies – except Brown and Reade (2019) – have not tested the wisdom-of-crowds effect in a realistic setting using data extracted from a real-world online tipster community. Second, the data sets with which existing studies have performed their analyses are not very comprehensive. As a consequence, they have not included rich information on the characteristics of the crowd to identify, for instance, smarter sub-crowds, although relevant literature on the wisdom-of-crowds effect have concluded that not only can crowdsourcing outperform experts but additionally the characteristics of the crowd are likely to influence the prediction results (O'Leary, 2017). Third, current evidence from related literature streams, such as financial studies, innovation management and entrepreneurship, suggests that textual user-generated content provides a rich crowd-based pool on information that can be easily collected and analysed to extract the wisdom of crowds (Chen et al., 2014; Otterbacher, 2009; Rhyn & Blohm, 2019). However, existing studies have just taken the tipsters' concrete results predictions as well as a few tipster-related characteristics, such as tipsters' prior tip experience, into account. Therefore, they have not examined the detailed textual match analysis attached to a published tip.

To address the abovementioned shortcomings, we extracted data from Blogabet.com, an online tipster community founded in 2006 that covers a wide range of sports and has a lively community with around 291,156 members.[1] Our final data set compromises 1,534,041 verified tips (which we will refer to as picks in the rest of this study) extracted from 3,484 tipsters. For each of these picks, our data set includes (1) contributor-/(tipster)-related variables such as prior experience, past performance and location, (2) textual-content-/(match-analysis)-related variables such as its length, its readability and its specificity as well as (3) community-related variables that measure the feedback and reaction of the community to specific tipsters in terms of the number of followers. We aim to leverage our collected data set to investigate whether (1) there is wisdom in a crowd of an online tipster community that can be used to improve betting returns and whether (2) we can identify specific characteristics that are indicators for smarter sub-crowds within the overcall crowd of an online tipster community.

To do so, we conducted two evaluations: First, we averaged and compared the implied win probability of the odds set by the bookmakers with the actual win percentage of the picks proposed by tipsters on Blogabet.com. Second, we conducted a mixed-effects logistic regression model (MELR) that identifies the characteristics of significantly smarter sub-crowds. The contributions of our study are as follows: For researchers, we prove that there is indeed wisdom in online tipster communities. Moreover, we propose a set of variables that explain specific characteristics that are indicators for smarter sub-crowds. For practitioners, our results provide important insights into TGC and the wisdom of online tipster communities that are especially relevant for bookmakers as well as tipsters, which either protect them from losses or improve their (betting) returns.

The remainder of this study 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 results as well as the discussion. We then propose the implication, avenues for future research and the limitations in Section 6. Finally, we conclude our study in Section 7.

Using Tipster-Generated Content to Extract the Wisdom of an Online Tipster Community to Predict Sports Outcomes

In recent years, user-generated content has raised the interest of researchers and practitioners alike as it allows to leverage publicly available data that often contains valuable information (e.g., customer needs or opinions) (Krumm et al., 2008). Therefore, many organisations are currently racing towards extracting insights from

[1] The numbers are based on Blogabet.com's website and are as of 13 December 2019.

user-generated content and leveraging them based on new business models (Byrum & Bingham, 2016). In the sports industry, user-generated sports content becomes a revealing data source as it offers (freely) available information such as discussions and experiences of fans or even thoughts and feelings of professional athletes (Gruettner et al., 2020). In the context of this study, TGC can be seen as user-generated tip recommendations that provide a valuable source to extract the wisdom of an online tipster community to predict the outcomes of sports events.

The wisdom-of-crowds effect operates on the premise that the independent judgement of a crowd of individuals (as measured by any form of central tendency) will be relatively accurate, even when most of the individuals in the crowd are ignorant and error-prone (Surowiecki, 2004). The effect has been studied and discussed in many research fields and scholarly articles in recent years, for instance, in financial studies to predict future stock returns. In this vein, Chen et al. (2014) proved that a textual analysis of users' posts on Seekingalpha.com, a popular opinion forum for stock market investors, has predictive power for future stock returns. In (product) innovation management, Hoornaert et al. (2017) as well as Beretta (2018) demonstrated that adding crowd-related information such as feedback (e.g., in the form of online comments) to the idea selection process helps to identify ideas that are more likely to be successful. Similarly, in the entrepreneurship literature, Mollick and Nanda (2016) showed evidence that support on the crowdfunding website Kickstarter.com is a better predictor of the success of theatre productions than evaluations by a designated expert panel. However, although many of the conducted studies support the wisdom-of-crowds effect, there have also been studies which challenge the accuracy and fundamental premises of crowd prediction: For example, Haan et al. (2005) concluded that experts are less sensitive to the emergence of new information than crowds. Thus, experts are likely to act less impulsively. Critics also observed that some crowd members might simply select crowd favourites rather than evaluate the data independently. This reinforcing behaviour could, for instance, lead to over-valuing the crowd favourite and can have a negative impact on the prediction accuracy (Budescu & Chen, 2015; Peeters, 2018). As a consequence of these ongoing discussions, we believe that in a sports outcome prediction context a fair test of the wisdom-of-crowds effect requires an investigation of a crowd in a realistic real-world market setting.

Several research articles based on the wisdom-of-crowds effect have been published in recent years to investigate whether crowd wisdom can be used to predict sports outcomes. One of the most prominent examples includes Twitter, which has been studied as a predictor of soccer games. For instance, Schumaker et al. (2016)

investigated whether the sentiment contained in tweets can serve as a meaningful proxy to predict match outcomes. The authors found that crowdsourced sentiment can be a better predictor of match outcomes than odds. Likewise, Peeters (2018) concluded that information extracted from Transfermarkt.de evaluations – where online users submit transfer valuations of soccer players – could be used to generate sizeable betting returns. In detail, the author showed that forecasts of international soccer results based on the crowd's evaluations are more accurate than those based on standard predictors (e.g., FIFA ranking). O'Leary (2017) compared the performance of a Yahoo crowd to experts in predicting the outcomes of matches in the FIFA World Cup 2014. The analysis found that the crowd was statistically significantly better at predicting outcomes of matches than experts and very similar in performance to established betting odds. However, none of the existing studies dealt with a specific online tipster community context. There is just one study – to the best of the knowledge of the authors – which evaluated the wisdom-of-crowds effect in a realistic real-world online tipster community setting. Brown and Reade (2019) extracted data from Oddsportal.com and investigated the accuracy of crowd forecasts. The authors found that the crowd outperforms bookmakers in specific cases, leading to the conclusion that tip recommendations (i.e., TGC) in online tipster communities contain information that is not in betting prices.

Existing studies come with several weaknesses in the way that they did not meet the four conditions for crowd wisdom set out by Simmons et al. (2011). According to the authors, a crowd is smart when the members in the crowd are (1) knowledgeable, (2) motivated to be accurate, (3) diverse and (4) independent. The majority of the existing studies used either social media platforms such as Twitter (e.g., Schumaker et al. (2016)) or data extracted from online websites such as Yahoo (e.g., O'Leary (2017)) for their analyses. Social media platforms or online websites have not been set up with the aim of eliciting crowd wisdom. Thus, they usually generate collective judgements (Peeters, 2018). Furthermore, they typically do not provide users with any explicit incentives to induce accurate reporting. For example, the incentives on social media to provide accurate information for forecasting may arguably be weak. Unlike online tipster communities, accurate social media forecasts may enhance an individual's reputation but are not directly profitable. Even worse, there are many instances of misinformation on social media (Antretter et al., 2019; Chen et al., 2014). In the same vein, social media platforms and online websites often make little attempt to reach a diverse user population. Finally, they usually allow (and indeed stimulate) communications between users, which may limit the independence of users' opinions (Peeters, 2018).

We believe that our online tipster community setting from Blogabet.com meets the four conditions set up by Simmons et al. (2011) for the following reasons: First, online tipsters are knowledgeable as they publish sports picks regularly, mostly for only a specific selection of sports types or games. Second, we assume that they are motivated not only to publish accurate picks because they (probably) want to place their own money on a particular pick but also to build up a strong track record within the tipster community (e.g., a favourable ranking between all tipsters of the online tipster community) that allows them, for example, to offer access to their picks on a paid basis. Third, tipsters in an online tipster community are embedded in a broad and diverse network of people over the internet. Accordingly, they have different backgrounds and are even interested in different types of sports or sports clubs. Hence, the crowd is sufficiently diverse, decentralised through the reach of the internet, able to be summarised and rapidly independent. Consequently, and fourth, tipsters' evaluations of specific picks (mostly) rely on their own information and are not influenced by other members of the online tipster community.

Extracting TGC and the wisdom of an online tipster community poses several challenges: As described, contributions in online tipster communities are submitted by a diverse network of people with different backgrounds and degrees in expertise. As a result, the quality of the tipsters as well as of the published picks vary drastically from excellent to noise and ambiguity, such as abuse and spam. In addition, besides structured data such as predefined match-related metadata (e.g., the kick-off time), there is a wide array of unstructured data such as TGC. Figure 1 shows an overview of the tipsters ordered by the number of followers of an online tipster community. Figure 2 illustrates an example of a published pick of a tipster. As a consequence of the abovementioned challenges, the process of manually reviewing and filtering the large amount of tipsters and TGC to identify valuable picks and separating them from low-quality contributions that should not be used to extract the wisdom of the crowds is a latent challenge. Text mining and Natural Language Processing (NLP) techniques represent promising solutions to cope with the vast amount of contributions in an online tipster community. Thus, they provide the means to discover patterns and extract useful information from textual data in a fast, automatic, scalable and repeatable way (Rhyn & Blohm, 2019).

Research from prior literature on the wisdom-of-crowds effect as well as from related literature streams applied such techniques to extract the wisdom of vast amounts of contributions. The findings commonly reported that highly valuable contributions are marked by specific characteristics. In this vein, Hoornaert et al. (2017) proposed a model that can be adapted to the context of this study and helps to analyse TGC to extract the wisdom of an online tipster community. The authors identified three sources of information (the “3Cs”) available in online communities: (1) Characteristics about the contributor, (2) the textual content of the contribution and (3) the community's feedback and reaction. The “contributor” category refers to TGC that contains information about the tipster who published a pick. For instance, previous studies have proven evidence that crowds improve their forecasting performance over time as they become more experienced and skilled (e.g., Budescu and Chen (2015), Goldstein et al. (2014), or Lamberson and Page (2012)). Therefore, it is reasonable to assume that experienced and skilled tipsters also form smarter sub-crowds in an online tipster community. The “content” category refers to the content-related textual features of a textual contribution, that is, the detailed textual match analysis attached to a published pick, which is usually expressed in unstructured, written text. Content-related textual features, such as the length (e.g., Riedl et al. (2013) or Wang and Strong (1996)), the readability (e.g., Flesch (1943) or Otterbacher (2009)) or the specificity (e.g., Otterbacher (2009) or Weimer and Gurevych (2007)), can be used to examine how carefully a tipster has elaborated their picks. This, in turn, can be used to assess the quality of a pick and, hence, helps to identify patterns of smarter sub-crowds. The “community” category refers to the feedback and reactions of the community. For instance, studies highlight that positive feedback from the community in the form of, for example, comments or likes can be seen as a proxy for the communities’ satisfaction with a specific tipster (e.g., Antretter et al. (2019), Beretta (2018), and Hoornaert et al. (2017)). Thus, it can be used as a means to identify smarter sub-crowds of tipsters.

In this study, we build upon the principles of the “3Cs” to derive our hypotheses that are explained in the following section.

Figure 1

Overview of the Tipsters Ordered by the Number of Followers of an Online Tipster Community (Here Blogabet.com)

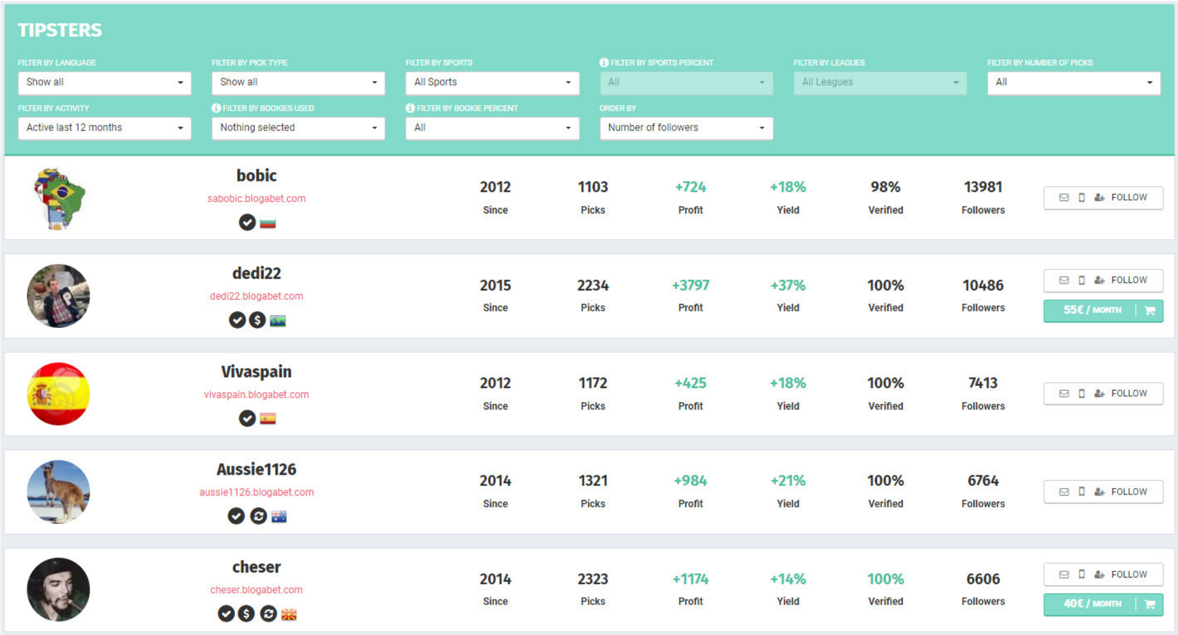
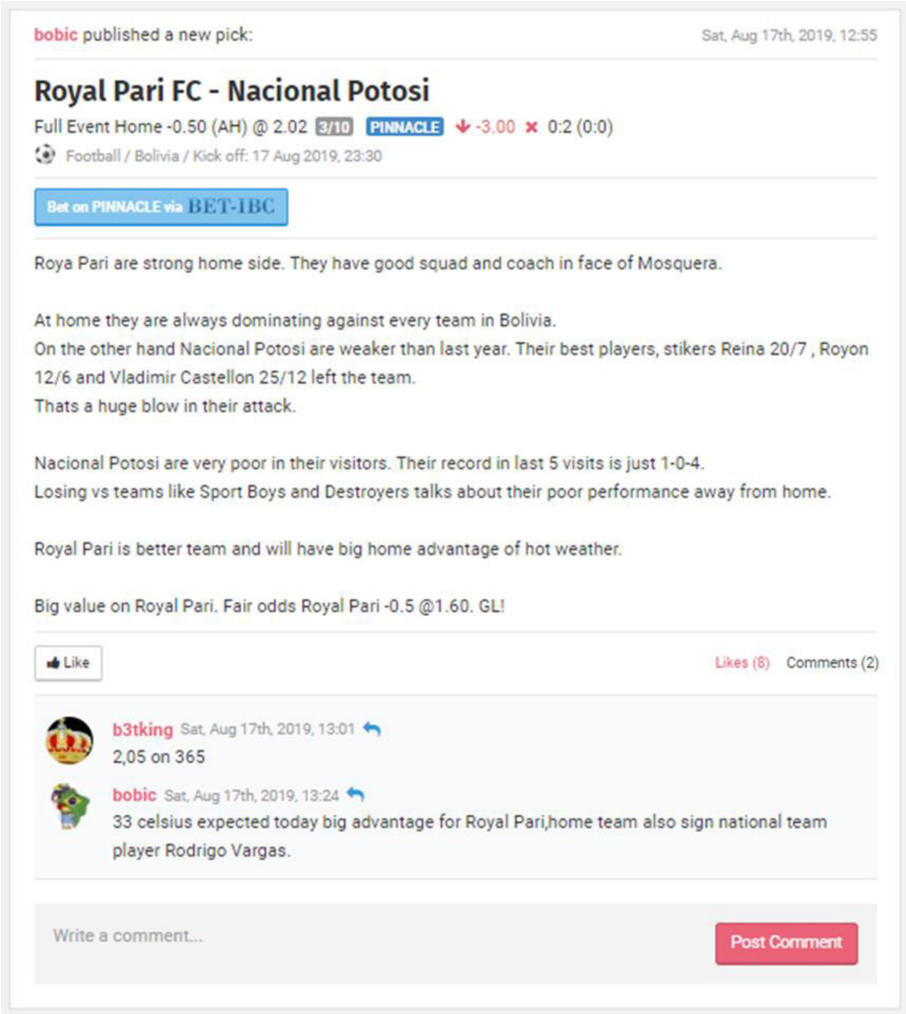


Figure 2

Illustration of a Published Pick of a Tipster in an Online Tipster Community (Here Blogabet.com)

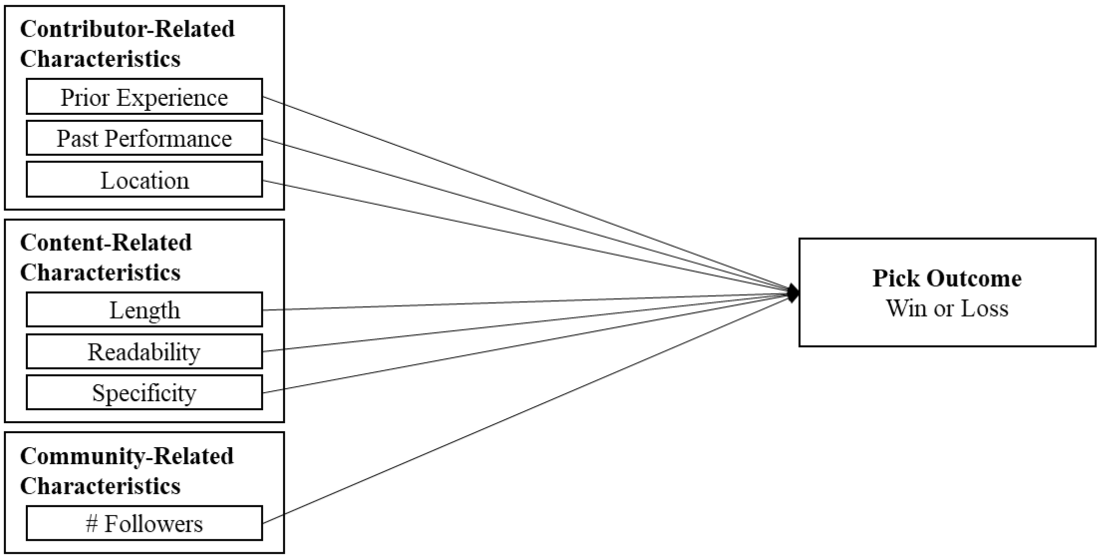


Hypotheses Development

For the development of our research model (see, Figure 3), the following hypotheses draw from prior literature on the wisdom-of-crowds effect as well as from related literature streams (e.g., crowdsourcing, entrepreneurship, financial studies, innovation and idea generation, and data quality and computational filtering approaches) to provide a discussion of what can be measured about each of the “3Cs” proposed by Hoornaert et al. (2017) to extract the wisdom of an online tipster community.

Figure 3

Research Model of this Study



Development of Contributor-Related Crowd Characteristics (Hypothesis 1)

Contributor (i.e., tipster-related) characteristics, in our study, refer to (1) the prior experience, (2) the past performance and (3) the location of the crowd in an online tipster community. Previous studies have noted that the prior experience and the past performance of a crowd are positively related to the accuracy of predictions and can, therefore, outperform larger crowds as well as experts. For instance, Goldstein et al. (2014) identified smaller sub-crowds that beat the wisdom of bigger crowds in a Fantasy Football Player selection. Their findings were that both the prior experience, measured as average years the crowd has played, and the past performance, measured as ranking within the community, have a significant positive impact. These findings are in line with Lamberson and Page (2012), who investigated the optimal group composition for accurate forecasts, as well as with Budescu and Chen (2015), who improved the quality of aggregate forecasts by eliminating poorly performing individuals from the crowd. Extrapolating from these studies, prior tip experience and a history of successful picks in the past may indicate a sub-crowd in an online tipster community that has expertise and knowledge, which in turn suggests that they are indicators of smarter sub-crowds within the overall crowd of an online tipster community.

Evidence for the wisdom-of-crowds effect has been shown with large, diverse samples in many different contexts. However, several previous findings concluded that it matters far less whether the crowd is homogenous or diverse along demographic dimensions, such as age, sex or the location of crowd members (e.g., de Oliveira and Nisbett (2018) or van Dijk et al. (2012)). In contrast to these findings, we assume that demographic characteristics of the crowd could be indicators of smarter sub-crowds in the context of sports betting for the following reasons: Recently published studies commonly reported that the sports betting markets reveal inefficiency and information asymmetries. For example, Elaad et al. (2019) found that individual bookmakers are not efficient. Their own odds do not appear to fully use the information contained in their competitors' odds. Furthermore, Brown and Reade (2019) demonstrated, as described, that online tipster communities contain information that is not in betting prices. We expect that crowds of tipsters that bet on picks played in their home country will have more detailed information and expertise about the pick than foreign tipsters and bookmakers. This is especially true for marginal sports like table tennis or lower country-specific sports leagues in which usually less public information is available (Peurala, 2013). Therefore, the location of a crowd is added to our research model.

Consequently, one should expect that the prior experience, the past performance and the location of a crowd of tipsters are indicators of smarter sub-crowds.

> **H1a:** *The characteristics of the crowd in terms of its prior experience is positively related to correctly predicting the outcome of a pick.*
>
> **H1b:** *The characteristics of the crowd in terms of its past performance is positively related to correctly predicting the outcome of a pick.*
>
> **H1c:** *The characteristics of the crowd in terms of its location is positively related to correctly predicting the outcome of a pick.*

Development of Content-Related Crowd Characteristics (Hypothesis 2)

The analysis of content-related textual features is commonly used in practice to extract, for instance, the creativity, expertise or workforce of a given crowd (Rhyn & Blohm, 2019). Hence, several recent studies provide supportive evidence of the value of this type of crowd wisdom. For instance, Klein and Garcia (2015) presented an approach called the "bag of lemons", which enables crowds to filter ideas based on textual features with accuracy superior to conventional approaches. Chen et al. (2014) showed, as described, that a textual analysis of users' posts on Seekingalpha.com has predictive power for future stock returns. In this study, we focus on three content-related textual features for picks that include a detailed textual match analysis that are likely to be indicators of smarter sub-crowds: That is (1) the length, (2) the readability and (3) the specificity of a published pick.

In an online tipster community, tipsters can use text to explain why they have chosen to bet on a specific pick. Providing sufficient and detailed information about the pick facilitates the evaluation process and demonstrates that the tipster has spent time to elaborate it. In turn, this increases the likelihood that a pick has a positive outcome. The amount of information in a textual contribution (i.e., its length) has frequently been discussed as one of its most important features by related literature (Wang & Strong, 1996). Longer contributions contain more information that could potentially be relevant for the outcome of the pick than shorter ones. On the other hand, researchers emphasised that contributions that are short and less elaborated tend to deliver less information (e.g., Riedl et al. (2013)). In the same vein, the readability of a pick can be used to analyse the syntactic and semantic complexity of a published pick, which we claim is also a proxy of how carefully a tipster has elaborated their pick (Flesch, 1943). Higher readability of a pick often indicates a better-evaluated pick and thus makes it easier to extract relevant cues or information. Past research has shown that a better readability score is likely to enhance the interpretability or clarity of a textual contribution and may enable the acquisition of the embedded information (Otterbacher, 2009). Lastly, related literature emphasises the need to consider the specificity and relevance of the information in a textual contribution. For instance, Weimer and Gurevych (2007) used similarity features to measure the relatedness of an online post to a forum topic. Likewise, Otterbacher (2009) quantified the extent to which a product review contains terms that are statistically important across other reviews as a significant indicator of the helpfulness of a review. In the context of this study, we believe that smarter sub-crowds base their decisions to publish a pick on clear and specific criteria, such as information that is based on quantitative pre-match analyses (e.g., win/loss ratio of home and away games), or other specific information, such as missing or injured players.

Following the argumentations above, we assume that the length, the readability and the specificity of a pick are used by smarter sub-crowds. Consequently, our content-related hypotheses to extract the wisdom of an online tipster community are as follows:

> **H2a:** *The content-related characteristics of the crowd's picks in terms of its length is positively related to correctly predicting the outcome of a pick.*
>
> **H2b:** *The content-related characteristics of the crowd's picks in terms of its readability is positively related to correctly predicting the outcome of a pick.*
>
> **H2c:** *The content-related characteristics of the crowd's picks in terms of its specificity is positively related to correctly predicting the outcome of a pick.*

Development of Community-Related Crowd Characteristics (Hypothesis 3)

While H1 and H2 focus on the characteristics of the crowd in terms of its contributors and content, this study finds that it is also desirable to incorporate the feedback and the reaction of the online tipster community into the research model to identify specific characteristics that are indicators for smarter sub-crowds in an online tipster community. Studies on crowdsourcing (e.g., Hoornaert et al. (2017)) and idea selection (e.g., Antretter et al. (2019)) highlighted that positive feedback from the community indicates the popularity and approval of a member within the online community. Thus, it can represent a way of evaluating a particular crowd of tipsters. For example, Antretter et al. (2019) showed that the number of followers on Twitter is among the most important predictors for new venture survival. Likewise, Beretta (2018) noted that one should not only attract a large number of contributors to extract the wisdom of crowds but also considers types of participants to engage with to enable access to diverse knowledge and expertise. Therefore, we pose the following hypothesis:

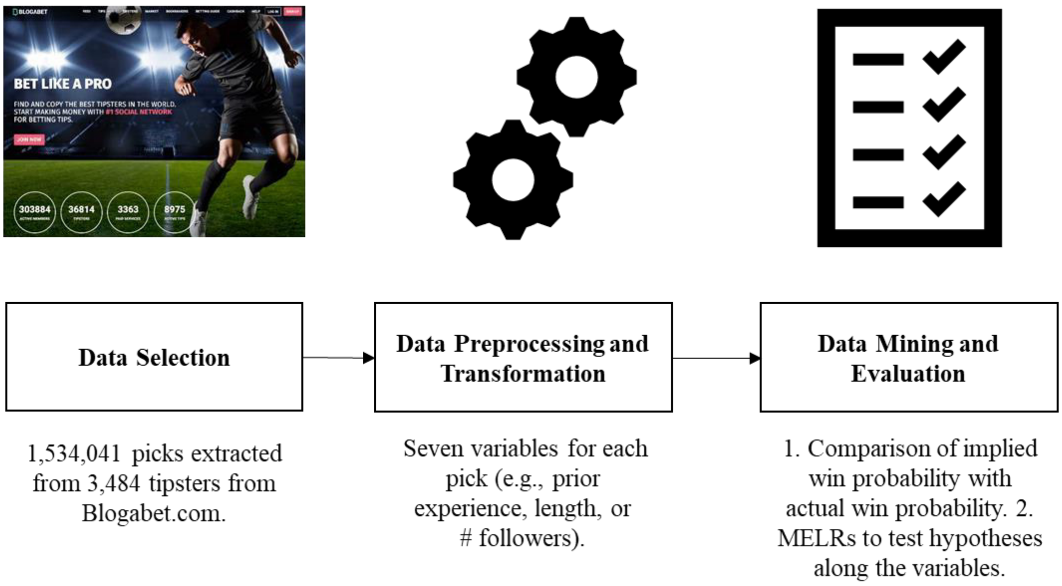
> **H3:** *The feedback and the reaction of the community in terms of the number of followers are positively related to correctly predicting the outcome of a pick.*

Methodology

To validate our hypotheses, we followed an adjusted knowledge discovery process in databases (KDD) as described by Fayyad et al. (1996). KDD describes the structured extraction of useful information from a large volume of data. For the purpose of this study, KDD's iterative cycle approach was combined and formed into three steps: (1) data selection, (2) data pre-processing and transformation and (3) data mining and evaluation (see, Figure 4), which are explained in the following subsections.

Figure 4

Adjusted KDD and the Corresponding Results/Approach



Data Selection

The first step of our adjusted KDD included the selection of data sources. As described, we chose Blogabet.com as a representative of an online tipster community. We chose Blogabet.com for two major reasons: First, all historical data about tipster-related characteristics (i.e., prior experience, past performance and location), content-related characteristics (i.e., the length, readability and specificity of the textual match analyses), community-related characteristics (i.e., number of followers) as well as the pick outcome are publicly available. Second, Blogabet.com has implemented the “verified odds” concept. This concept guarantees that the odds displayed are accurate and available at the time of publishing the pick. To achieve that, Blogabet.com uses multiple direct bookmakers’ feeds where tipsters can verify the right picks for their selections in real time. Furthermore, Blogabet.com has implemented a review system in which tipsters correct and verify themselves. Consequently, a high reliability and accuracy of the picks included in our data set are guaranteed.

We added the entire pick archive of a given tipster (see, Figure 5 for the inclusion/exclusion criteria for the selection of tipsters and picks) to our data set for further analysis: Blogabet.com recorded 11,378 active tipsters (i.e., the tipster published at least one pick in the last twelve months) who had published 4,005,176 picks at the time of our analysis (the numbers are as of 28 September 2019). To narrow down the selection of tipsters, in a first step, we deleted all tipsters that had less than 100 picks or more than 2,000 picks from our data set. It is generally accepted in the sports betting market that tipsters that have less than 100 picks are not reliable, and their past performance may be due to luck or coincidence. Likewise, a closer look at the tipsters that had more than 2,000 picks made the impression that multiple tipsters or commercial ventures maintained these accounts. In the next step, we excluded all tipsters and picks that were not 100% verified as described above, as well as tipsters that were banned, paused or reset their pick history during the data collection phase. We then excluded all picks whose match analysis was not written in English in order to obtain reliable results, since some of our applied data analytics techniques are designed for the English language. In total, 1,534,041 picks from 3,484 tipsters were included in our final data set for further analysis. This data set is publicly available and can be accessed via the following reference: Gruettner (2020).

Figure 5

Inclusion and Exclusion Criteria for Tipsters and Picks

	# Tipsters	# Picks
Overview Blogabet.com as of September 28th, 2019	11,387	4,005,176
Tipsters with at least 100 picks	4,721	3,831,027
Tipster with not more than 2,000 picks	4,307	2,101,116
Tipsters with 100% verified picks	3,544	1,694,022
Unmatched tipster or pick (banned, paused, or reset)	3,484	1,580,211
Textual match analysis in English	**3,484**	**1,534,041**

Data Pre-Processing and Transformation

After extracting and collecting relevant data, the next step was data preparation and transformation. Therefore, we transformed the collected data into a structured format that can be accessed and processed for further analysis. To validate our developed hypotheses (i.e., H1 to H3), we generated several variables that are based on the extracted data. Three example picks of the used variables can be found in Table 1 and are explained as follows: To examine if there is wisdom in a crowd of an online tipster community that can be used to improve betting returns, we averaged and compared the implied win probability of the odds set by bookmakers with the actual win percentage of the picks proposed by tipsters. The implied win probability was calculated as:

Implied Win Probability = 1 / Odds

The actual win percentage was calculated as:

Actual Win Percentage = Number of won picks / Number of all picks

To identify specific characteristics that are indicators for smarter sub-crowds within the overall crowd of an online tipster community, we first pre-processed and transformed all data related to the contributor-related crowd characteristics (H1). The prior experience variable displayed the number of all picks that a tipster had published. The past performance variable contained the so-called yield of a tipster. The yield is commonly used in sports betting to compare the performance of tipsters. To test whether a pick was published in the home country of a tipster (H1c), we created a location dummy variable in which 1 stood for a pick that was placed in the same location and 0 for a pick that was not.

To test the content-related crowd characteristics (H2), different data analytics techniques using the programming language Python 3.7 were applied, since it is widely known, easy to use and supports major libraries for NLP tasks: First, all picks that included a textual match analysis with less than four words were set to 0 because they did not contain any relevant information or included noise for further analysis. Afterwards, we calculated the length of each match analysis by splitting the picks' strings into tokens. We deleted all English stop words based on the Natural Language Toolkit (NLTK) stop word list for the English language (Bird et al., 2009). Finally, we counted the number of filtered tokens of each pick. The token length of a pick was then used to test our hypothesis H2a. Next, we aimed to measure the readability of each match analysis to test if it correlates to the outcome of a pick. To measure the syntactic readability of texts, several measures have been used in research (Khawaja et al., 2010). We selected the Flesch-Reading-Ease (FRE) to capture the readability of the picks since this score combines language complexity measurements, such as the average sentence lengths and the average syllables per word, into one number (Flesch, 1943). The score has been widely used before to determine the readability of a message in computer-mediated communication (e.g., Walther (2007)) or for measuring the readability of posts in online forums (e.g., Wambsganss and Fromm (2019)). We used the following formula:

Flesch Reading Ease = 206.835 – (1.015 * asl) – (84.6 * asw)

asl: average sentence length of a response, asw: average syllables per word

The scores of our answers ranged from 0 to 120.20. The higher the FRE score was, the better the readability of the match analysis. The FRE was used to test our hypothesis H2b. Moreover, we aimed to test the specificity of information in a given pick. Therefore, we retrieved a dictionary of domain-related vocabularies from “word net” (Princeton University, 2019), representing related words for sports injuries. We controlled if the word stem of any token in a pick matches the word stem of any word in our dictionary. For stemming the tokens of our picks and the dictionary entries, we used the English Porter Stemmer provided by NLTK by Bird et al. (2009).

For the validation of H3, we used the number of followers as a variable that measures the community-related crowd characteristics extracted from Blogabet.com.

To measure the success of a published pick, we chose the pick outcome in terms of whether a tipster won or lost the pick. We chose the outcome of the pick as our dependent variable, as it is the performance measure used by tipsters in practice and, therefore, can be applied as a proxy for the quality of the pick and in turn helps to extract the wisdom of the crowd.

Table 1

Three Examples of a Published Pick and their Corresponding Variables

Textual Match Analysis	Pick Outcome	Implied Win Probability	Prior Experience	Past Performance	Location	Length	Readability	Specificity	# Followers
United has got a lot of injuried player, Arsenal have got a high quality attacker players...	Win	47.62%	108	2%	0	8	55.97	1	14,464
I think Leganes wants this game much more. They are strong in their home. Even the games they lost in their home they played pretty well. Leganes has the momentum and the passion to win this game. Possible scores 1-0 / 2-1	Loss	27.78%	435	7%	0	22	98.55	0	133
-	Win	55.56%	1,769	-4%	1	0	0	0	5

Data Mining and Evaluation

The last step of our adjusted KDD included the data mining and evaluation phase. As described, to examine if there is wisdom in a crowd of an online tipster community that can be used to improve betting returns, we averaged and compared the implied win probability of the odds set by the bookmakers with the actual win percentage of the picks proposed by tipsters. The difference between both variables indicates whether there is wisdom in a crowd of an online tipster community that can be used to improve betting returns (see, Brown and Reade (2019)). To identify specific characteristics that are indicators for smarter sub-crowds within the overall crowd of an online tipster community, we measured the impact of the abovementioned variables on the outcome of the pick. In our data set, a particular tipster produces multiple data points (i.e., picks) over time. Therefore, traditional statistical approaches (e.g., linear/logistic regressions or analysis of variance) are of limited use because of restrictive assumptions concerning the variance-covariance structure of the repeated measures in longitudinal data sets (Hedeker & Gibbons, 2006; Laird & Ware, 1982). For instance, (1) error terms that are correlating with each other or (2) variances that lead to different sources of heterogeneity (e.g., between tipsters or within a tipster themselves) (Fitzmaurice et al., 2012). To deal with these challenges and to predict dichotomous outcome variables (i.e., binary outcome variables), researchers commonly use MELRs when observations are correlated. MELRs have shown to be sensitive and statistically powerful while dealing with longitudinal data sets as well as missing values in various theoretical as well as practical studies (e.g., Vermunt (2005)). Therefore, we believe that an MELR guarantees robustness and reliability to test H1 to H3 in our study. For our analysis, we scaled the values of our data set to meet the requirements of the lme4 MELR R package version 1.7 as proposed by Bates et al. (2014).

Results and Discussion

This study set out with the aim of assessing whether (1) there is wisdom in the crowd of an online tipster community that can be used to improve betting returns and whether (2) we can identify specific characteristics that are indicators for smarter sub-crowds of an online tipster community. To validate our hypotheses, as mentioned, a comparison of the implied win probability with the actual win percentage of the picks and an MELR was conducted. The detailed results, including a discussion, are presented in the subsections below.

Descriptive Statistical Results

Our final data set comprised 3,484 tipsters, who published a total of 1,534,041 picks (see, Table 2 for an overview of the descriptive statistics). Overall, the results showed that 797,769 or 52.00% out of the total 1,534,041 picks had a positive pick outcome (i.e., the tipster won the pick). In contrast, the averaged implied win probability set by the bookmakers of all picks was only 48.71%. The difference is 3.29%. Extrapolating from this, we can prove that there is indeed wisdom in the crowd of an online tipster community that can be used to improve betting returns. In this vein, Figure 6 shows a detailed overview of the relationship between the implied win probability, the number of picks and the pick outcome. In 67,024 picks, or to put it differently, in 4.44% of the overall analysed picks, tipsters included a textual match analysis. These picks showed an actual win percentage of 54.90%. In contrast to the averaged implied win probability of 50.54%. The difference, that is 4.36%, indicates that the subset of picks which included a textual match analysis achieved even better results than the overall dataset including all picks. Out of the total 1,534,041 picks, 205,039 picks (or 13.37%) were played in the local home country of the tipster who published the pick. 55.54% of these picks had a positive pick outcome, in contrast to an average implied win probability of 50.55%. The difference is the largest (4.99%), indicating that smarter crowds publish picks in their home country.

Tipsters' prior experience showed a mean of 478 published picks. They were registered on Blogabet.com for a mean of 2.83 years. Furthermore, tipsters had average yields of 3.97%, indicating a positive past performance of the majority of tipsters in our data set. The length of picks that included a textual match analysis showed a mean of 19.95. While the readability score of these picks presented a mean of 82.89, the specificity score showed a mean of 0.07. The community's feedback, in the form of the number of followers for each tipster included in our data set, showed a mean of 66.

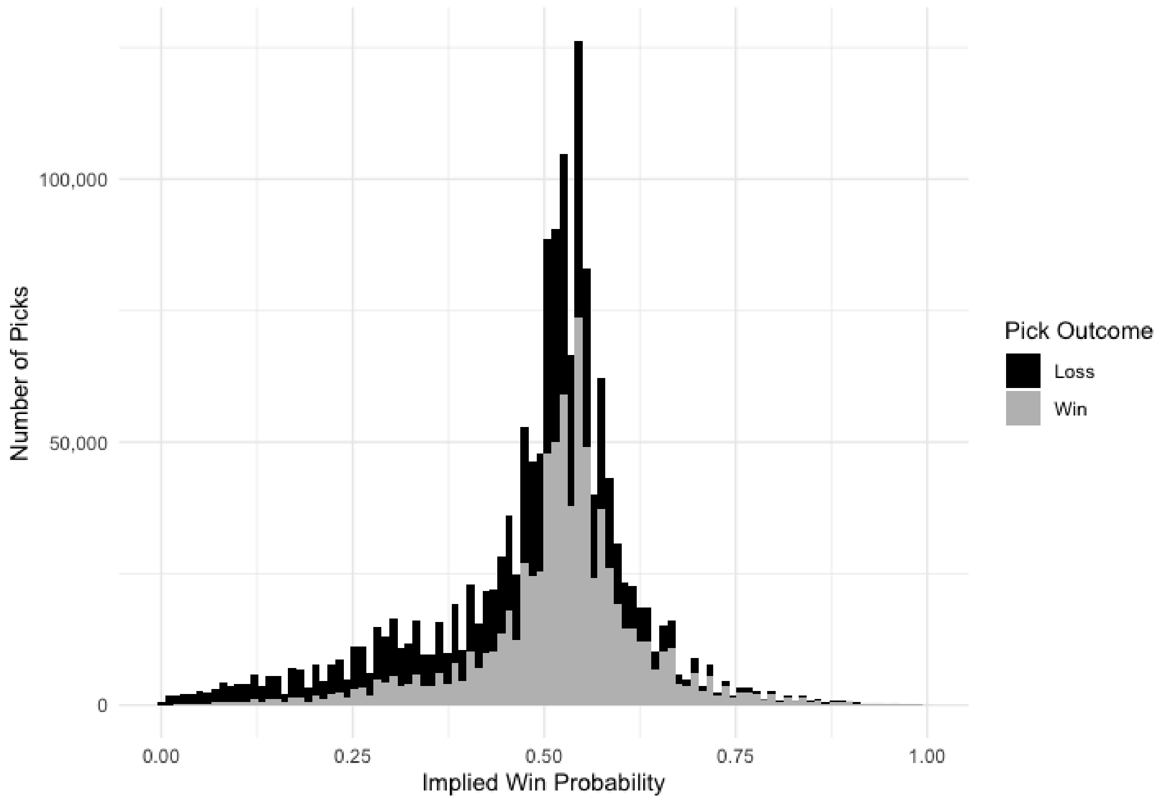
Table 2

Descriptive Statistics of the Final Data Set

Data Set Overview Number of tipsters = 3,484	**Overall**		**Including Textual Match Analysis**	**Location**
Number of Picks	1,534,041		67,024 (4.44%)	205,039 (13.37%)
Actual Loss Percentage	48.00% (736,272)		45.10% (30,225)	44.46% (91,168)
Actual Win Percentage	52.00% (797,769)		54.90% (36,799)	55.54% (113,871)
Av. Implied Win Probability	48.71%		50.54%	50.55%
Difference*	3.29%		4.36%	4.99%
Contributor Characteristics	**Mean**	**SD**	**Min**	**Max**
Prior Experience	477.63	417.44	100	1,997
Past Performance	3.97%	11.03%	-96%	110%
Content Characteristics				
Length	19.95	22.73	0	418
Readability	82.89	16.60	0	120.20
Specificity	0.07	0.31	0	9
Community Characteristics				
Number of followers	65.82	254.99	0	6,288
* The difference is calculated as: Actual Win Percentage – Av. Implied Win Probability SD = Standard deviation				

Figure 6

Relationship Between Implied Win Probability, Number of Picks, and Pick Outcome



Evaluating Contributor-Related Crowd Characteristics

The first hypothesis proposed in this study was that the characteristics of the crowd in terms of its prior experience (H1a), its past performance (H1b) and its location (H1c) are positively related to correctly predicting the pick outcome. As Figure 7 and Table 3 show, the results of our performed MELR for the prior experience variable were significant at the $p < 0.01$ level and for the past performance variable highly significant at the $p < 0.001$ level. These findings confirm that prior experience (H1a) and past performance (H1b) are significant indicators of a smarter sub-crowd in an online tipster community that possesses expertise and knowledge and, thus, is likely to propose additional successful future picks.

These results stand in contrast to the only study by Brown and Reade (2019) that also investigated the wisdom of a crowd effect in a realistic real-world online tipster community setting. In their study, Brown and Reade (2019) concluded that sub-crowds that are evaluated based on prior experience (more past tips) or past performance (higher historical returns on their tips) did not achieve better accuracy than the overall crowd in an online tipster community. Brown and Reade (2019) defined their crowd of experienced tipsters as "those who have previously lodged more tips than the median tipster who lodged a tip on the same event" (Brown & Reade, 2019, p. 3). Likewise, they defined a skilled crowd as "those who have, at the time, a higher hypothetical return on their tips than the median tipster who lodged a tip on the same event" (Brown & Reade, 2019). We believe that classifying tipsters on an event basis involves several risks: First, depending on which specific event is analysed, the threshold for being classified as an experienced/skilled or as an inexperienced/unskilled tipster can vary drastically. Second, therefore, it is likely that the same tipster is classified once in the crowd of experienced/skilled tipsters and once in the crowd of inexperienced/unskilled tipsters, depending on which other tipsters it is compared with. As a consequence, the data set will have dependencies in its observations. Thus, we believe that a Micer-Zarnowitz regression, which was applied in Brown and Reade's study, is not appropriate for such an analysis. To handle the abovementioned risks, we applied an MELR and implemented specific inclusion and exclusion criteria (see, Subsection 4.1) to our final data set. Therefore, we think that our results are reliable as they stand in line with previous findings on the wisdom-of-crowds effects (e.g., Budescu and Chen (2015), Goldstein et al. (2014), or Lamberson and Page (2012)). In this vein, we assume that although sports betting is often associated with gambling tipsters improve their betting performance over time as they become more experienced and skilled (Levitt et al., 2012).

It was also hypothesised that a crowd of tipsters could be smarter based on its location (H1c). In detail, we expected that tipsters that bet on games played in their local home country would have more detailed information and expertise about the game and hence are more likely to propose additional information for successful future picks. The findings of our analysis showed significance at the $p < 0.001$ level (see, Table 3). Thus, H1c can also be confirmed. Prior studies have commonly reported that they did not find any evidence that crowds are smarter based on demographics (e.g., de Oliveira and Nisbett (2018) or van Dijk et al. (2012)). In our opinion, one can assume that the sports betting market differs from the previously conducted studies for two reasons: First, in an online tipster community, many tipsters specialise themselves, for example, on one specific type of sport or a specific sports league within a specific country. In turn, these specialised tipsters have a higher likelihood to have information that is not included in bookmakers' odds and consequently are more likely to propose additional successful future picks. Second, the sports betting market is characterised by information asymmetries (Brown & Reade, 2019; Elaad et al., 2019). These asymmetries are often triggered by insider information as well as game manipulations, which can even be multiplied in specific contexts, such as marginal sports, individual sports types and lower country-specific sports leagues (Peurala, 2013). Our findings, therefore, are consistent with previous studies that have shown that the wisdom-of-crowds effect is even stronger when less public information is available in the market (e.g., Da & Huang, 2019).

Evaluating Content-Related Crowd Characteristics

The second hypothesis in this study hypothesised that the characteristics of the crowd's textual match analyses in terms of its length (H2a), its readability (H2b) and its specificity (H2c) are positively related to correctly predicting the pick outcome. Content-related textual features are commonly discussed in related literature streams to extract the wisdom out of a crowd and have proved evidence of the high value of this type of crowd wisdom (e.g., Chen et al. (2014), Klein and Garcia (2015), or Rhyn and Blohm (2019)). Thus, we expected similar results in our study. However, the findings of our study did not support previous research (see, Table 3). Neither the length nor the readability and specificity variable could achieve any significant results. Therefore, we must reject H2a, H2b and H2c. Several factors could explain this observation: First, on Blogabet.com it is possible to only allow access to picks on a paid basis. When collecting the data set, we noticed that some of the tipsters who offered their picks on a paid basis deleted their match analysis after the game of a particular pick was played. A reason for that could be that they did not want to give insights into their betting strategies to the whole community as historical picks are publicly available to all members. Second, it is reasonable to assume that the picks that did not include a detailed textual match analysis are also elaborated carefully as tipsters in an online tipster community are intrinsically motivated enough to induce accurate picks (see, Section 2). We, therefore, assume that our online tipster community setting differs from other settings investigated so far in that, for example, in contrast to an idea selection context, no detailed textual match analysis needs to be included in a pick. However, although the results of this study were not significant for H2, we are still convinced that tipsters' textual match analyses provide a valuable information source to extract the wisdom of an online tipster community. Therefore, more research on this topic needs to be conducted.

Evaluating Community-Related Crowd Characteristics

The last hypothesis (H3) stated that the feedback and the reaction of the community in terms of the number of followers for a specific tipster is positively related to correctly predicting the pick outcome. We assumed that the number of followers can be seen as a proxy for the communities' satisfaction with a specific tipster. Members of an online tipster community look for tipsters that publish picks that will help them to improve their betting returns. Therefore, they start following certain tipsters to monitor their published picks. This, in turn, leads to even more members being attracted, as other members also trust the tipsters with the most followers. In this sense, the tipsters in our study on Blogabet.com are ordered by default according to the number of followers. The results showed evidence for this hypothesis at the $p < 0.01$ level (see, Figure 7 and Table 3). This finding broadly supports the work of other studies from, for instance, product innovation in which positive feedback measured as the number of likes or the number of positive ratings have shown evidence to be a significant indicator for idea generation (e.g., Hoornaert et al. (2017)). While this finding is supported by previous studies in different contexts, it would be worthwhile to investigate whether the feedback and reaction of the community can also be used to improve pick forecasts. For example, it would be interesting to investigate if comments on a particular pick that are discussed on Blogabet.com contain information that is not included in the previously published pick, which could, in turn, improve the picks' forecasts.

Figure 7

Research Model Including Statistical Results of Conducted MELR

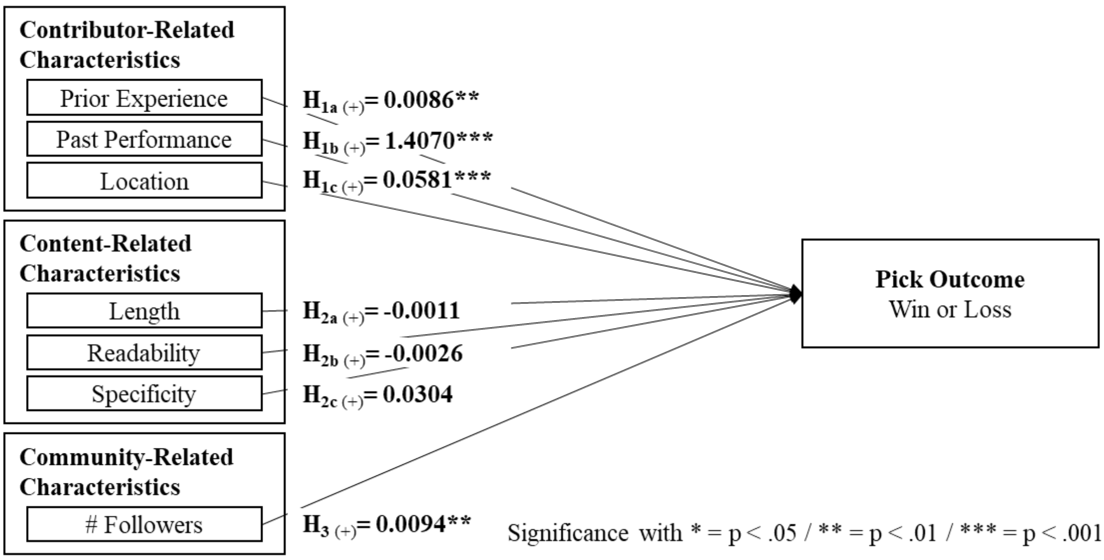


Table 3

Results of the MELRs

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Intercept	-1.999*** (0.0088)	-1.995*** (0.0089)	-2.074*** (0.0081)	-2.005*** (0.0087)	-2.078*** (0.0081)	-1.999*** (0.0087)	-1.999*** (0.0087)	-1.999*** (0.0087)	-1.999*** (0.0088)	-1.995*** (0.0087)	**-2.077*** (0.0081**
Implied Win Probability	4.246*** (0.0163)	4.247*** (0.0162)	4.290*** (0.0155)	4.243*** (0.0162)	4.289*** (0.0153)	4.246*** (0.0162)	4.246*** (0.0162)	4.246*** (0.0162)	4.246*** (0.0163)	4.248*** (0.0161)	**4.289*** (0.0153)**
Prior Experience		0.0085* (0.0041)			0.0105*** (0.0027)						**0.0086** (0.0027)**
Past Performance			1.430*** (0.0255)		1.426*** (0.0252)						**1.4070*** (0.0258)**
Location				0.0553*** (0.0060)	0.0581*** (0.0056)						**0.0581*** (0.0056)**
Length						0.0002 (0.0019)			-0.0009 (0.0024)		**-0.0011 (0.0023)**
Readability							0.0006 (0.0019)		0.0008 (0.0023)		**-0.0026 (0.0022)**
Specificity								0.0259 (0.0260)	0.0281 (0.0273)		**0.0304 (0.0272)**
# Followers										0.0590*** (0.0044)	**0.0094** (0.0031)**
Random Effect Tipster Intercept Variance	0.0257	0.0257	0.0079	0.0251	0.0073	0.0258	0.0258	0.0258	0.0258	0.0236	**0.0073**
AIC	2,008,580	2,008,578	2,006,294	2,008,498	2,006,174	2,008,582	2,008,582	2,008,582	2,008,585	2,008,409	**2,006,169**

Significance with * = p < .05 / ** = p < .01 / *** = p < .001
Logit scale regression coefficients followed by their standard errors in brackets.
An MELR was conducted on various models for this analysis, considering the pick outcome as the binary dependent variable. In models (2)-(4), (6)-(8) & (10) each of our explanatory variables was tested individually, whereas model (5) & (9) show the combined contributor and content variables, respectively. Finally, in model (11) an MELR was conducted with all our independent variables. To explain their relative explanatory power and model fit, we added the variance of the random effect intercept (tipsters) and the Akaike Information Criterion (AIC).

Implications, Future Research and Limitations

To investigate whether (1) there is wisdom in a crowd of an online tipster community that can be used to improve betting returns and whether (2) we can identify specific characteristics that are indicators for smarter sub-crowds within the overcall crowd of an online tipster community, we developed a realistic real-world setting and extracted 1,534,041 picks which stem from 3,484 tipsters on Blogabet.com. The results showed that 797,769 or 52.00% out of the total of 1,534,041 picks had a positive pick outcome (i.e., the tipster won the pick). In contrast, the averaged implied win probability set by the bookmakers of all picks was only 48.71%. The difference is 3.29%. Similarly, tipsters had average yields of 3.97%, indicating a positive past performance of the majority of tipsters in our data set. Extrapolating from both of these findings, we can confirm that there is indeed wisdom in the crowd of an online tipster community that can be used to improve betting returns. To identify specific characteristics that are indicators for smarter sub-crowds of an online tipster community, we developed three hypotheses in accordance with the three sources of information (the "3Cs") available in online communities as proposed by Hoornaert et al. (2017). The results of our MELR confirmed H1 that contributor-related crowd characteristics (i.e., prior experience, past performance and the location) are significantly positively related to correctly predicting the pick outcome. For H2, we assumed that content-related characteristics (i.e., a detailed textual match analysis) of the crowd's picks in terms of its length, its readability and its specificity are positively related to correctly predicting the outcome of a pick. Content-related textual features have proven to be valuable in extracting the wisdom of online crowds in various contexts (e.g., Klein and Garcia (2015) or Ma et al. (2019)). However, we were not able to achieve any significant results for any content-related crowd characteristics in our study. Thus, we must reject H2. The last hypothesis (H3) stated that the feedback and the reaction of the community in terms of the number of followers for certain tipsters is positively related to the correct prediction of the pick outcome. This study showed significant results to confirm this hypothesis.

From an academic perspective, the contribution of this study is twofold: First, this study contributes to the literature stream of technological innovations in sports-based entrepreneurship and especially to online sports communities (see, Ratten (2011)). We proposed the second study, which applied the wisdom-of-crowds effect in a realistic real-world online tipster community setting and demonstrated that there is indeed wisdom in online tipster communities by analysing TGC. In doing so, we identified four characteristics, that is, prior experience, past performance, the location as well as the number of followers, that are significant indicators for smarter sub-crowds within a community of online tipsters. These findings stand in contrast to the only other study conducted by Brown and Reade (2019), which also used data from an online tipster community. In future research, a closer look at the proposed characteristics could become relevant in identifying whether more successful crowds focus on, for instance, specific types of sports such as marginal sport types, lower country-specific sports leagues or individual vs. team sports. Likewise, although we were not able to achieve any significant results for any content-related crowd characteristics, we are still convinced that tipster-generated detailed textual match analyses provide a valuable information source to extract the wisdom of an online tipster community. Therefore, more research on this topic needs to be undertaken. Second, our results provide important insights into the value of user-generated content and the dynamics and the wisdom of online communities in general and, therefore, go beyond the sports literature. In this vein, this study proposes a set of variables to academia that explains specific characteristics that are indicators for smarter sub-crowds in online communities that should be studied in different contexts in future research. Especially the demographic variable, that is, the location of a crowd member, should be used by researchers in future studies since prior studies did not find any evidence for demographic characteristics within smarter sub-crowds (e.g., de Oliveira and Nisbett (2018) and van Dijk et al. (2012)). For example, in a financial market context. We further contribute to research as well as practice by making a novel, comprehensive, reliable and high-quality data set publicly available that provides many possibilities for future research.

From a practical perspective, the contributions are especially relevant for bookmakers and tipsters. On the one hand, we empirically demonstrated how digital technologies such as data analytics solutions can be beneficially implemented to extract the wisdom of online sports communities (Gruettner, 2019; Ratten, 2017). Thus, we provide a concrete example that can be used by bookmakers and tipsters to either protect themselves from losses or to improve their betting returns. In this vein, bookmakers can understand our results as a wake-up call to have a closer look at online tipster communities. For instance, bookmakers can use our findings to identify those crowds of tipsters that are most successful or influential in online tipster communities. In turn, they can adjust their odds directly after specific crowds of tipsters have published a pick. This provides them with an early protection system against losses. To do so, they can either build a dashboard that monitors online tipster communities for specific crowds of tipsters or they can build a prediction model based on machine learning algorithms and our published data set that classifies individual pick outcomes. On the other hand, this study shows that tipsters can improve their betting returns using the wisdom of an online tipster community. Tipsters can, for example, adjust their betting strategies using our identified four characteristics of smarter sub-crowds to identify crowds of tipsters that publish the most valuable picks in online tipster communities. To generalise our study, organisations of any type that are currently racing towards extracting insights from user-generated content should be encouraged to use our findings to leverage user-generated content in their businesses (Byrum & Bingham, 2016).

This study is not free from limitations: First, online tipster communities such as Blogabet.com enable tipsters to observe the picks of others. Tipsters, thus, may be influenced which would lead in some cases to correlated forecast errors and an inferior crowd forecast. Similarly, this presents an issue for researchers as we cannot disentangle individuals' beliefs from the crowd's beliefs. Second, we believe that the lack of further control variables could be a limitation. Third, we used a specific set of Python libraries and pre-trained techniques (such as language detection). The accuracy of the techniques is limited to a certain degree; however, we believe that our results display the overall notion of the data. Last, we only investigated the positive impact of crowd characteristics on the outcome of a pick. It should also be investigated whether there are characteristics of crowds that indicate negative performances in terms of betting returns.

Conclusion

By investigating a real-world online tipster community from Blogabet.com, we analysed whether (1) crowd prediction can be used to improve sports betting returns and whether (2) there are specific characteristics that are indicators for smarter sub-crowds within such communities of sports tipsters. For research and practice, the conducted analyses showed evidence that such communities indeed contain wisdom which can be used to improve betting returns (tipsters won 3.29% more picks than the implied win probability set by bookmakers and produced average yields of 3.97%). We further identified four characteristics, that is, prior experience, past performance, the location as well as the number of followers that are significant positive indicators for correctly predicting the pick outcome and, thus, are characteristics which are typical for smarter sub-crowds. Our results provide important insights into user-generated content and the dynamic and the wisdom of online communities in general and, therefore, go beyond the sports literature. Future research should either try to identify further characteristics of successful sub-crowds or should concentrate on the proposed set of variables of this study and apply it to different research settings. Similarly, future research should dive deeper into whether tipster-generated textual match analyses provide a valuable information source to extract wisdom, which did not show any significant results in this study. Our results are especially relevant to bookmakers and tipsters, who either want to protect themselves from losses or want to improve their betting returns.

References

Antretter, T., Blohm, I., Grichnik, D., & Wincent, J. (2019). Predicting new venture survival: A Twitter-based machine learning approach to measuring online legitimacy. *Journal of Business Venturing Insights*, *11*(e00109). https://doi.org/10.1016/j.jbvi.2018.e00109

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. In *R package version 1.7*.

Beretta, M. (2018). Idea Selection in Web-Enabled Ideation Systems. *Journal of Product Innovation Management*, *36*(1), 5–23. https://doi.org/10.1111/jpim.12439

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc. https://doi.org/10.1097/00004770-200204000-00018

Brown, A., & Reade, J. J. (2019). The wisdom of amateur crowds: Evidence from an online community of sports tipsters. *European Journal of Operational Research*, *272*(3), 1073–1081. https://doi.org/10.1016/j.ejor.2018.07.015

Budescu, D. V., & Chen, E. (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science*, *61*(2), 267–280. https://doi.org/10.1287/mnsc.2014.1909

Byrum, J., & Bingham, A. (2016). Improving Analytics Capabilities Through Crowdsourcing. *MIT Sloan Management Review*, *57*(4), 43–48. https://doi.org/10.7551/mitpress/11633.003.0020

Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *The Review of Financial Studies*, *27*(5), 1367–1403. https://doi.org/10.1093/rfs/hhu001

Da, Z., & Huang, X. (2019). Harnessing the Wisdom of Crowds. *Management Science*, *66*(5), 1783–2290. https://doi.org/10.1287/mnsc.2019.3294

de Oliveira, S., & Nisbett, R. E. (2018). Demographically diverse crowds are typically not much wiser than homogeneous crowds. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9). https://doi.org/10.7302/Z2D50K5H

Deutscher Sportwettenverband. (2020). *Sportwettenmarkt in Deutschland bricht Umsatzrekorde*. https://dswv.de/sportwettenmarkt-in-deutschland-bricht-umsatzrekorde/

Elaad, G., Reade, J. J., & Singleton, C. (2019). Information, prices and efficiency in an online betting market. *Finance Research Letters*. https://doi.org/10.2139/ssrn.3378257

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proceedings of the 2nd International Conference on Knowledge Discovery And Data Mining (KDD) 1996*, 82–88.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied Longitudinal Analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Flesch, R. (1943). Marks of readable style; a study in adult education. *Teachers College Contributions to Education*, *897*, ix + 69.

Goldstein, D. G., McAfee, R. P., & Suri, S. (2014). The Wisdom of Smaller, Smarter Crowds. *Proceedings of the 15th ACM Conference on Economics and Computation (EC) 2014*. https://doi.org/10.1145/2600057.2602886

Gruettner, A. (2019). What We Know and What We Do Not Know About Digital Technologies in the Sports Industry. *Proceedings of the 25th Americas Conference on Information Systems (AMCIS) 2019*.

Gruettner, A. (2020). *Data Set - The Wisdom of an Online Tipster Community*. https://www.alexandria.unisg.ch/259455/

Gruettner, A., Vitisvorakarn, M., Wambsganss, T., Rietsche, R., & Back, A. (2020). The New Window to Athletes' Soul-What Social Media Tells Us About Athletes' Performances. *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS) 2020*.

Haan, M. A., Dijkstra, S. G., & Dijkstra, P. T. (2005). Expert Judgment Versus Public Opinion - Evidence from the Eurovision Song Contest. *Journal of Cultural Economics*, *29*(1), 59–78. https://doi.org/10.1007/s10824-005-6830-0

Hedeker, D. R., & Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Hoornaert, S., Ballings, M., Malthouse, E. C., & Van den Poel, D. (2017). Identifying New Product Ideas: Waiting for the Wisdom of the Crowd or Screening Ideas in Real Time. *Journal of Product Innovation Management*, *34*(5), 580–597. https://doi.org/10.1111/jpim.12396

Khawaja, M. A., Chen, F., & Marcus, N. (2010). Using language complexity to measure cognitive load for adaptive interaction design. *Proceedings of the 15th International Conference on Intelligent User Interfaces, Proceedings (IUI) 2010*. https://doi.org/10.1145/1719970.1720024

Klein, M., & Garcia, A. C. B. (2015). High-Speed Idea Filtering with the Bag of Lemons. *Decision Support Systems*, *78*, 39–50. https://doi.org/10.1016/j.dss.2015.06.005

Krumm, J., Davies, N., & Narayanaswami, C. (2008). User-Generated Content. *IEEE Pervasive Computing*, *7*(4), 10–11. https://doi.org/10.1109/MPRV.2008.85

Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, *38*(4), 963–974. https://doi.org/10.2307/2529876

Lamberson, P. J., & Page, S. E. (2012). Optimal Forecasting Groups. *Management Science*, *58*(4), 805–810. https://doi.org/10.1287/mnsc.1110.1441

Levitt, S. D., Miles, T. J., & Rosenfield, A. M. (2012). Is Texas Hold ’Em a Game of Chance? A Legal and Economic Analysis. *The Georgetown Law Journal*, *101*(581), 581–636.

Ma, J., Lu, Y., & Gupta, S. (2019). User innovation evaluation: Empirical evidence from an online game community. *Decision Support Systems*, *117*, 113–123. https://doi.org/10.1016/j.dss.2018.11.003

Mollick, E., & Nanda, R. (2016). Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. *Management Science*, *62*(6), 1533–1553. https://doi.org/10.1287/mnsc.2015.2207

O’Leary, D. E. (2017). Crowd Performance in Prediction of the World Cup 2014. *European Journal of Operational Research*, *260*(2), 715–724. https://doi.org/10.1016/j.ejor.2016.12.043

Otterbacher, J. (2009). “Helpfulness” in Online Communities: A Measure of Message Quality. *Proceedings of the 27th Conference on Human Factors in Computing Systems (CHI) 2009*. https://doi.org/10.1145/1518701.1518848

Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, *34*(1), 17–29. https://doi.org/10.1016/j.ijforecast.2017.08.002

Peurala, J. (2013). Match-manipulation in football - The challenges faced in Finland. *International Sports Law Journal*, *13*, 268–286. https://doi.org/10.1007/s40318-013-0027-z

Princeton University. (2019). *WordNet*. http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=hurt

Ratten, V. (2017). *Sports Innovation Management* (1st ed.). New York, New York: Routledge.

Ratten, V. (2011). Sport-based entrepreneurship: Towards a new theory of entrepreneurship and sport management. *International Entrepreneurship and Management Journal*, *7*, 57–69. https://doi.org/10.1007/s11365-010-0138-z

Rhyn, M., & Blohm, I. (2019). Patterns of Data-Driven Decision-Making: How Decision-Makers Leverage Crowdsourced Data. *Proceedings of the 40th International Conference on Information Systems (ICIS) 2019*.

Riedl, C., Blohm, I., Leimeister, J. M., & Krcmar, H. (2013). The Effect of Rating Scales on Decision Quality and User Attitudes in Online Innovation Communities. *International Journal of Electronic Commerce*, *17*(3), 7–36. https://doi.org/10.2753/jec1086-4415170301

Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84. https://doi.org/10.1016/j.dss.2016.05.010

Simmons, J. P., Nelson, L. D., Galak, J., & Frederick, S. (2011). Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds. *Journal of Consumer Research*, *38*(1), 1–15. https://doi.org/10.1086/658070

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York, New York: Anchor.

van Dijk, H., van Engen, M. L., & van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, *119*(1), 38–53. https://doi.org/10.1016/j.obhdp.2012.06.003

Vermunt, J. K. (2005). Mixed-Effects Logistic Regression Models for Indirectly Observed Discrete Outcome Variables. *Multivariate Behavioral Research*, *40*(3), 281–301. https://doi.org/10.1207/s15327906mbr4003_1

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, *23*, 2538–2557. https://doi.org/10.1016/j.chb.2006.05.002

Wambsganss, T., & Fromm, H. (2019). Mining User-Generated Repair Instructions from Automotive Web Communities. *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS) 2019*. https://doi.org/10.24251/hicss.2019.144

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–34. https://doi.org/10.1080/07421222.1996.11518099

Weimer, M., & Gurevych, I. (2007). Predicting the Perceived Quality of Web Forum Posts. *Proceedings of the 6th Conference on Recent Advances in Natural Language Processing (RANLP) 2007*.