The Wisdom of the Silent Crowd: Predicting the Match Results of World Cup 2018 through Twitter

Abdullah Talha Kabakus
Duzce University
Faculty of Engineering,
Department of Computer
Engineering

Mehmet Simsek
Duzce University
Faculty of Engineering,
Department of Computer
Engineering

Ibrahim Belenli
Bolu Abant Izzet Baysal University
Faculty of Arts and Science,
Department of Physics

ABSTRACT

Social networks have attracted much attention thanks to their availability, being the source of information in almost everything related to life, and giving the chance for users to discover the things and people related to them. Global and highly popular events such as the FIFA World Cup are always welcomed by social networking sites. Furthermore, these social networking sites customize themselves and provide some specific content for this global events. From this point of view, we have collected 38,371,358 tweets during the FIFA World Cup 2018 which were posted by 7,876,519 unique Twitter users. The aim of this study is proposing a prediction system that evaluates the teams who qualified for the FIFA World Cup 2018 through their squad, and their performances in the early stages of the competition in order to predict the match results of the further stages. For this reason, a number of different types of machine learning algorithms are utilized after building a machine learning model which is based on novel features. According to the experimental results, the best accuracy of the proposed system for the match result prediction is calculated as high as 87.5% which is quite better than the related work. The experimental results demonstrate the effectiveness of social network based features.

Keywords

Machine Learning, Sports Result Prediction, Social Network Analysis, Twitter, FIFA World Cup.

1. INTRODUCTION

Football competitions are always greatly supported and keenly interested by a wide range of people. FIFA (The Fédération Internationale de Football Association) World Cup is regarded as the most prestigious football competition which is arranged once in every four years with the attendance and interest of millions of football fans from all around the world. With the rise of social networks and wide availability, social networks such as Facebook and Twitter has become the best sources to track the trending topics and analyze the opinions of huge crowds [1–3] which use social networking sites to share their opinions about a particular event or topic [4, 5]. The latest FIFA World Cup, FIFA World Cup 2018, is arranged in Russia from 14th of June to 15th of July, 2018. According to the official resources [6], a total of 3,031,768 people have attended to the matches of World Cup 2018 and the attendance is reached almost 100 percent. During the final of FIFA World Cup 2014, more than 618,000 tweets were posted per a minute which was a new record for Twitter [7]. As an overall, 672 million tweets were posted which were related to the FIFA World Cup 2014 [8]. A recent report by Twitter indicates that 115 billion tweets which represented the views of people were posted during the FIFA World Cup 2018 [9]. 32 national football teams which are qualified to

attend World Cup are distributed to 8 groups as each group contains 4 teams. Then in the group stage, each team takes on with the other 3 teams in the group. As a total, 48 matches are played during the group stage where each match has three possible outcomes: (1) Team-1 wins, (2) a draw, and (3) Team-2 wins. Then the best two teams in each group qualify for the next round. The rounds after groups are based on elimination which means there should be a winner of each match. While the loser team of each match is eliminated from the competition, the winner goes to the next round. As overall, 64 matches are played in each World Cup.

Football match prediction (or generally sports match prediction) is generally handled as a classification problem where one possible outcome of each match is predicted. This prediction is based on large numbers of features such as the historical performances of teams, previous matches between the teams, the form the teams and their players, any major absents of each team, and home/away status of teams. For a World Cup, the competition is arranged in a country which is determined before a couple of years from the date the competition starts. The national team of the host country may or may not be in the competition. Both the competition and the performances of teams are widely discussed in social media through the hashtags shared by the official FIFA World Cup Twitter account[1]. Most of the national teams have also official Twitter accounts that share the latest news and updates from their teams. According to Twitter, the peak of 127,000 tweets per minute were posted at the moment England won the quarter-finals thanks to the penalty kick [10]. In this study, a novel prediction system based on machine learning is proposed which predicts the results of matches of FIFA World Cup 2018. The proposed system uses novel features based on social media analysis alongside the statistical features of teams for prediction. Within the scope of this study, 38,371,358 tweets were collected during the FIFA World Cup 2018 which were posted by 7,876,519 unique Twitter users. The rest of the paper is structured as follows: Section 2 presents the related work. Section 3 describes the material and method. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the paper with future directions.

2. RELATED WORK

Purucker [11] proposed a system based on Artificial Neural Network (ANN) to predict match results in the National Football League (NFL). Five features that were used are (1) yards gained, (2) rushing yards gained, (3) turnover margin, (4) time of possession, and (5) betting line odds. The best result was achieved when an ANN with backward-

---

[1] https://twitter.com/FIFAWorldCup

propagation was used which yielded to 61% accuracy. Kahn [12] extended this work and achieved an accuracy of 75%. The features that were used contain (1) total yardage differential, (2) rushing yardage differential, (3) turnover differential, (4) away team indicator, and (5) home team indicator. Godin et al. [13] propose an approach to predict the winner of English Premier League 2013-2014 based on four methods such as: (1) Statistical analysis such as team ranking, total points gathered, total points gathered in last five games, the number of goals scored, and the number of goals conceded, (2) Twitter volume analysis which basically calculates the number of tweets per each team on the basis of the total volume of tweets is a good indicator for the ranking of a team, (3) sentiment analysis, and (4) user prediction analysis based on team mention(s) in the collected tweets.

Tax and Joustra [14] propose a prediction system which predicts the results of football matches of Dutch football competition. They have investigated the effect of the match based features by comparing a model with betting odds and a hybrid model of both betting odds and match based features. They have experimented nine classification algorithms by utilizing the machine learning software WEKA[2] namely Naive Bayes, LogitBoost (with decision stumps), NN with BP, Random Forest, CHIRP, FURIA, DTNB, C4.5, and hyper pipes. They have reported that the highest performing classification algorithms are Naive Bayes (used with a 3-component PCA), and the ANN (used with a 3 or 7-component PCA) which have achieved an accuracy of 54.7%. Herzog and Hertwig [15] have rated the teams and players according to their recognition by people. Troubadour [16] asked participants to rank sets of four countries. Then, the maximum different scaling was used to choose the best and the worst of each set. Schläfli [17] proposed a contest format which consists of over one thousand participants. The participants have predicted the outcomes of both group and knock-out rounds using a pick'em approach.

Radosavljevic [18] proposed a model named Goalr which was based on Tumblr to estimate the strength of each team in the FIFA World Cup 2014 by analyzing the occurrence of both team and player mentions. Goalr's match winner prediction accuracy was calculated as about 48% (23 matches of an overall of 48 matches). Some studies [19, 20] utilized sentiment analysis techniques on social media content to generate predictions for the match results. Goldman Sachs [21] proposed a stochastic model that generated a distribution of outcomes of each match in the World Cup 2014. The prediction for each match is based on a regression analysis. Bloomberg [22] developed a simulation model and generated over 6 million simulations in order to develop a set of predictions at each stage of the World Cup, including outcomes of the matches in group play. Tumasjan et al. [23] used Twitter to predict the outcome of the German Federal elections by counting tweets that mention a candidate or political party. Their reasoning for this approach was tweets mentioning a candidate or a political party also indicates their voting intention. This method has yielded accurate results as the error ratio was calculated as low as 1.65%.

3. MATERIAL AND METHOD
In this section, the data structure of the collected tweets and the way how the data is collected are described alongside the selected features of the proposed machine learning-based prediction system.

3.1 Data Structure and Collection
Tweets related to the FIFA World Cup 2018 are selected through the hashtags shared by the official FIFA World Cup Twitter account. The generic hashtags related to the FIFA World Cup 2018 are "WorldCup", "WC2018", and "Russia2018". Each match is also defined by a hashtag which is a concatenation of both teams three-characters abbreviation codes. For the sports, the Olympic countries abbreviation codes[3] are used while describing the countries instead of ISO (International Organization for Standardization) 3166-1 Alpha-3 codes as the official Twitter account of the FIFA World Cup have shared just before the competition started[4]. Therefore, the Olympic countries abbreviation codes are used to identify the tweets related to the match. The tweets are collected through the Twitter Streaming API (Application Programming Interface) which is provided by Twitter. Since the proposed system is implemented using the Java programming language, an open source Java library for Twitter API named Twitter4J [24] is utilized to fetch the tweets by the pre-defined hashtags in real-time. Within the scope of this study, a dataset named WC2018 is constructed which contains information about the collected 38,371,358 tweets alongside the 7,876,519 unique Twitter users who have posted these tweets. The information about each tweet and each user contains the attributes which are listed in Tables 1 and 2, respectively. Whole information is directly retrieved from the Twitter API.

Table 1. The attributes of the stored tweets

Attribute	Description
_id	The unique id of the tweet which is defined by Twitter
userId	The user id of the user who posted the tweet which is defined by Twitter
username	The username of the user who posted the tweet
message	The message of the tweet
favoriteCount	The favorite count of the tweet
rtCount	The retweet count of the tweet
lang	The language information of the tweet which is identified by Twitter
isRt	The flag that indicates whether the tweet is retweet or not
inReplyToUserId	The userId of the tweet that the current tweet is replied to
inReplyToTweetId	The _id of the tweet that the current tweet is replied to
postdate	The date the tweet is posted in an ISODate format
latitude	The latitude information of the tweet is posted (if available)
longitude	The longitude information of the tweet is posted (if available)
mentionedUsers	The mentioned users in the message of the tweet
mentionedHashtags	The mentioned hashtags in the message of the tweet
rtTweetId	The _id of the tweet which is retweeted from

Table 2. The attributes of the users who posted the stored tweets

Attribute	Description
_id	The unique user id of the user which is defined by Twitter
username	The username of the owner of the tweet
name	The name of the user
tweetCount	The number of tweets which is posted by the user
followersCount	The number of followers of the user
	The number of users which is followed by the user
likeCount	The number of tweets liked by the user
location	The location defined by the user
description	The description of the user which is defined by himself
isVerified	The flag that represents whether the account is verified by Twitter or not
listedCount	The number of lists where the user is listed in
isDefaultProfileImage	The flag that indicates whether the default profile photo[5] is still used by the user or not
hasCustomBackground	The flag that indicates whether the user has a custom background or not
lang	The language of the user
createdDate	The date the account of the user is created in an ISODate format
Website	The website of the user
timezone	The time zone of the user

Alongside the tweets and the users, the footballers who are candidates for the FIFA Ballon d'Or 2017[6], the best player in the world award by FIFA, are stored within the teams in the FIFA World Cup 2018, and the matches of the competition. The information stored for each team contains (1) the abbreviation code of the team, (2) the population of the country that the team represents[7], and (3) the number of candidates footballers for FIFA Ballon d'Or 2017 the team has. The information stored for each match contains (1) the abbreviation codes of each team, (2) the match score, (3) the start time of the match, and (4) the number of tweets which contain the abbreviation code of the team but not the one of the opponent team for both teams. The number of tweets for both teams in each match contains the tweets which are posted during the two hours after the start of the match. The national team of the host country may or may not be in the World Cup and matches are not in the form of home-and-away. Therefore, the home/away status is ignored which is a feature used for match result predictions for the club competitions [12, 25, 26]. All the data within the this study is stored in an open source and the most popular NoSQL (Not Only SQL)

database [27] namely MongoDB[8] which is reported to provide quite better performance in terms of runtime while reading and writing data compared to the relational databases such as Oracle, MySQL, and Microsoft SQL Server [28–31].

The aim of the proposed system is to predict the results of matches in the FIFA World Cup 2018. Therefore, the problem is handled as a classification problem, and a number of different types of machine learning algorithms are utilized within the scope of this study. Each FIFA World Cup starts with 32 teams, and the stages of the competition can be grouped into six. The stages of a FIFA World Cup with the number of teams available and the number of matches played in each stage are listed in Table 3.

Table 3. The stages of a FIFA World Cup

Stage	Number of Teams Available	Number of Matches Played
Group	32	48
Round of 16	16	8
Quarter Finals	8	4
Semi Finals	4	2
Play-off for the Third Place	2	1
Final	2	1

4. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed system is designed to predict the matches results for the two scenarios: (1) Train the system using the matches in the group stage, and predict the later fixtures, and (2) train the system using the matches in the group stage and the round of 16, and predict the later fixtures. The scenarios are listed in Table 4 with the number of matches used for training and the number of matches used for testing. For both scenarios, machine learning algorithms are needed to utilize in order to teach the system the strengths of the team through the training data.

Table 4. The two scenarios the proposed system uses

Scenario	Number of Matches for Training	Number of Matches for Testing
#1	48	16
#2	56	8

The proposed system is based on a model that contains two novel features: (1) The number of Ballon d'Or candidates the national team has (rank), and (2) the ratio of the number of tweets posted for the team to the population of the country which the team represents (*tweetPerFan*). Nine machine learning algorithms namely BayesNet, Naïve Bayes, Naïve Bayes Multinomial, SVM (Support Vector Machines), kNN (k-Nearest Neighbor), Random Forest, Random Tree, Multilayer Perceptron, and Logistic Regression are utilized in order to compare their accuracies of match result predictions, and find out the one which provides the best accuracy. Whole algorithms are utilized through WEKA with their default configurations unless it is explicitly noted. Since the matches after the group stage are based on elimination, the two halves per 15 minutes (which is also known as extra time) are played

---

[5] When a Twitter account is created, a default profile photo (which is an egg) is set by Twitter.
[6] The FIFA Ballon d'Or 2017 was the most recent FIFA Ballon d'Or whose candidates are declared at the time of writing.
[7] https://the18.com/soccer-entertainment/world-cup-countries-populations

[8] https://www.mongodb.com

if the 90 minutes end with a draw. Even the draw still continues at the end of extra time, the penalty kicks start. Finally, the match ends with a winner at the end of penalty kicks. For this reason, the final result is used instead of the score at the end of 90 minutes for the predictions. For the scenario #1, the best accuracy of the proposed system is calculated as high as 81.25% when the Multilayer Perceptron algorithm is used with setting the number of epochs to 30,000 as the experimental result is listed in Table 5. Also, the accuracy is not changed when the number of epochs is increased to 45,000.

Table 5. The experimental result for the scenario #1

Machine Learning Algorithm	Accuracy (%)
Bayes Net	37.5
Naïve Bayes	50
Naïve Bayes Multinomial	50
SVM	75
kNN (k=2)	68.75
Random Forest (1 seed)	56.25
Random Forest (2 seeds)	62.5
Random Tree (0 fold, 1 seed)	50
Random Tree (0 fold, 2 seeds)	43.75
Random Tree (2 folds, 1 seed)	50
Random Tree (2 folds, 2 seeds)	50
Multilayer Perceptron (epochs=15000)	75
Multilayer Perceptron (epochs=30000)	**81.25**
Logistic Regression	43.75

The scenario #2 is specifically designed to include the knock-out round matches in the round of 16 which work in the same way with the predicted matches during the training phase. For the scenario #2, the best accuracy of the proposed system is increased to 87.5% when the SVM algorithm is used as it is expected since the ratio of training is increased. To the best of our knowledge, there is no such work that provides better accuracy in term of predicting match results. The experimental result for the scenario #2 is listed in Table 6. When the attribute based on Twitter (*tweetPerFan*) is deactivated, the best accuracy drops to 75% for both scenarios with utilizing the same machine learning algorithms. This accuracy is calculated when the system is utilized with the Multilayer Perceptron algorithm when the number of epochs is set to 30,000. This experimental result clearly indicates the effectiveness of Twitter at match result prediction

Table 6. The experimental result for the scenario #2

Machine Learning Algorithm	Accuracy (%)
Bayes Net	62.5
Naïve Bayes	62.5
Naïve Bayes Multinomial	25
SVM	**87.5**
kNN (k=2)	50

Random Forest (1 seed)	62.5
Random Forest (2 seeds)	62.5
Random Tree (0 fold, 1 seed)	25
Random Tree (0 fold, 2 seeds)	37.5
Random Tree (2 folds, 1 seed)	50
Random Tree (2 folds, 2 seeds)	50
Multilayer Perceptron (epochs=15000)	62.5
Multilayer Perceptron (epochs=30000)	62.5
Logistic Regression	75

Various works for the different competitions have been proposed as some of them have described in Section 2. These works can be classified through their prediction targets as follows: (1) Match score prediction, (2) match result prediction in terms of a win by team #1, win by team #2, or draw, and (3) the prediction of the winner of the competition. The related work which aims the same in term of the accuracy of predicting match result is listed in Table 7. Since the proposed system focuses on match result prediction, other works which aim to predict match scores or winners of competitions are excluded. Since the Twitter data of other competitions is not available, we are not in a position to directly compare the accuracy of the related work with the proposed system.

Table 7. The two scenarios the proposed system uses

Related Work	Topic of Prediction	Accuracy (%)
Herzog and Hertwig [15]	World Cup 2006	84
Schlafli [32]	World Cup 2014	60
Goldman Sachs [21]	World Cup 2014	37.5
Bloomberg [22]	World Cup 2014	43.75
Goalr [18]	World Cup 2014	48
Chow [33]	World Cup 2014	54.7
Godin et al. [13]	English Premier League 2013-2014	70
The proposed system	**World Cup 2018**	**87.5**

5. CONCLUSION

Social media analysis has opened a new era to mine opinions of community through the status updates they share. The great advancements in both social media and machine learning let researchers do polls online, make predictions about the future through the latest news and events all around the world. Within the scope of this study, the effect of social media to predict World Cup match results is investigated. According to experimental results, the best accuracy of match result prediction is calculated as high as 87.5% which is quite better than the related work in the literature. The main reason behind this great accuracy. The high accuracy calculated through the experimental results proves the effectiveness of social media. The main contributions of this study can be listed as follows:

- *Social media analysis.* The proposed system utilizes social media analysis for the match result predictions.

- *A minimum number of features used with machine learning techniques.* The match result prediction model which is powered by machine learning algorithms does not depend on a large number of features like the related work which are hard to obtain as a result of using effective features.

- *High accuracy.* The proposed system's accuracy is calculated as high as 87.5% which is quite higher than the related work in the literature.

- *Adaptability.* The proposed approach can be easily adapted for the solution of another prediction problem which is quite popular in social networks.

- *Constructed publicly available dataset.* Within the scope of this study, 38,371,358 tweets are collected during the FIFA World Cup 2018 which are posted by the 7,876,519 unique Twitter users. The constructed dataset namely WC2018 is publicly available to download for other researchers from http://bigdata.duzce.edu.tr/#datasets.

As a future work, authors would like to apply the proposed system for the matches in the UEFA European Football Championship 2020 and the FIFA World Cup 2022 to prove its accuracy of predicting match results for other competitions.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Yin Z, Cao L, Han J, et al (2011) Diversified Trajectory Pattern Ranking in Geo-Tagged Social Media. In: Proceedings of the 2011 SIAM International Conference on Data Mining. Mesa, Arizona, USA.

[2] Kurashima T, Iwata T, Irie G, Fujimura K (2010) Travel route recommendation using geotags in photo sharing sites. In: Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). ACM, Toronto, ON, Canada.

[3] Schreck T, Keim D (2013) Visual analysis of social media data. Computer (Long Beach Calif) 20:68–75. doi: 10.1109/MC.2012.430.

[4] Pak A, Paroubek P (2010) Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Lrec 1320–1326. doi: 10.1371/journal.pone.0026624.

[5] Desai M, Mehta MA (2016) Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey. In: 2016 International Conference on Computing, Communication and Automation (ICCCA 2016). Noida, India, pp 149–154.

[6] (2018) Records Broken at World Cup in Russia as Attendance Neared 100% - Deputy Prime Minister. In: FourFourTwo. https://www.fourfourtwo.com.au/feature/records-broken-at-world-cup-in-russia-as-attendance-neared-100---deputy-prime-minister-498625. Accessed 23 Aug 2018.

[7] Dewan P, Kumaraguru P (2017) Facebook Inspector (FbI): Towards automatic real-time detection of malicious content on Facebook. Soc Netw Anal Min. doi: 10.1007/s13278-017-0434-5.

[8] Rogers S (2014) Insights into the #WorldCup conversation on Twitter. In: Twitter. https://blog.twitter.com/official/en_us/a/2014/insights-into-the-worldcup-conversation-on-twitter.html. Accessed 23 Aug 2018.

[9] Bavishi J, Filadelfo E (2018) Insights into the 2018 #WorldCup conversation on Twitter. In: Twitter. https://blog.twitter.com/official/en_us/topics/events/2018/2018-World-Cup-Insights.html. Accessed 23 Aug 2018.

[10] (2018) England´s World Cup penalty win against Colombia draws. In: Dly. Mail Online. http://www.dailymail.co.uk/wires/pa/article-5917361/England-s-World-Cup-penalty-win-against-Colombia-draws-huge-TV-audience.html. Accessed 23 Aug 2018.

[11] Purucker MC (1996) Neural network quarterbacking. IEEE Potentials. doi: 10.1109/45.535226.

[12] Kahn J (2003) Neural Network Prediction of NFL Football Games. http://homepages.cae.wisc.edu/~ece539/project/f03/kahn.pdf. Accessed 28 Aug 2018.

[13] Godin F, Zuallaert J, Vandersmissen B (2015) Beating the Bookmakers: Leveraging Statistics and Twitter Microposts for Predicting Soccer Results. In: Proceedings of the 2014 KDD International Workshop on Large-Scale Sports Analytics.

[14] Tax N, Joustra Y (2015) Predicting The Dutch Football Competition Using Public Data: A Machine Learning Approach. Trans Knowl data Eng. doi: 10.13140/RG.2.1.1383.4729.

[15] Herzog SM, Hertwig R (2011) The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. Judgm Decis Mak.

[16] Gilliam D (2014) Predictions for the #WorldCup2014. In: Troubad. Res. Consult. http://troubadourconsulting.com/initial-predictions-for-the-worldcup2014/. Accessed 23 Aug 2018.

[17] Schläfli S (2014) Predicting World Cup results with the wisdom of the crowd. In: ETH Zurich. https://www.ethz.ch/en/news-and-events/eth-news/news/2014/07/viele-tippen-weiser.html. Accessed 23 Aug 2018.

[18] Radosavljevic V, Grbovic M, Djuric N, Bhamidipati N (2014) Large-scale World Cup 2014 outcome prediction based on Tumblr posts. Sunnyvale, CA, USA.

[19] Bothos E, Apostolou D, Mentzas G (2010) Using Social Media to Predict Future Events with Agent-Based Markets. IEEE Intell Syst. doi: 10.1109/MIS.2010.152.

[20] Schumaker RP, Jarmoszko AT, Labedz CS (2016) Predicting wins and spread in the Premier League using a sentiment analysis of twitter. Decis Support Syst. doi: 10.1016/j.dss.2016.05.010.

[21] Sachs G (2014) The World Cup and Economics 2014.

[22] (2014) World Cup 2014 Predictions. In: Bloomberg. https://www.bloomberg.com/graphics/2014-world-cup/#0,0,-1. Accessed 23 Aug 2018.

[23] Tumasjan A, Sprenger T, Sandner P, Welpe I (2010) Predicting Elections with Twitter: What 140 Characters

Reveal about Political Sentiment. In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. Washington, DC, USA, pp 178–185.

[24] (2018) Twitter4J - A Java library for the Twitter API. http://twitter4j.org/en/. Accessed 23 Aug 2018.

[25] Dixon MJ, Coles SG (1997) Modelling association football scores and inefficiencies in the football betting market. J R Stat Soc Ser C Appl Stat. doi: 10.1111/1467-9876.00065.

[26] Boshnakov G, Kharrat T, McHale IG (2017) A bivariate Weibull count model for forecasting association football scores. Int J Forecast. doi: 10.1016/j.ijforecast.2016.11.006.

[27] (2018) DB-Engines Ranking - popularity ranking of database management systems. In: DB-Engines. https://db-engines.com/en/ranking. Accessed 23 Aug 2018.

[28] Boicea A, Radulescu F, Agapin LI (2012) MongoDB vs Oracle - Database comparison. In: Proceedings of 3rd International Conference on Emerging Intelligent Data and Web Technologies, EIDWT 2012. Bucharest, Romania, pp 330–335.

[29] Parker Z, Poe S, Vrbsky S V. (2013) Comparing NoSQL MongoDB to an SQL DB. In: Proceedings of the 51st ACM Southeast Conference on - ACMSE '13.

[30] Ming Wu C (2015) Comparisons Between MongoDB and MS-SQL Databases on the TWC Website. Am J Softw Eng Appl 4:35–41. doi: 10.11648/j.ajsea.20150402.12.

[31] Nyati SS, Pawar S, Ingle R (2013) Performance evaluation of unstructured NoSQL data over distributed framework. In: Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2013.

[32] Schläfli S (2014) Predicting World Cup results with the wisdom of the crowd. In: ETH Zurich. https://www.ethz.ch/en/news-and-events/eth-news/news/2014/07/viele-tippen-weiser.html. Accessed 24 Aug 2018.

[33] Chow J (2014) World Cup 2014 Data Analysis and Predictions. https://github.com/woobe/wc2014. Accessed 28 Aug 2018.