Analyzing and Predicting Meetup Mobs Outcome Via Statistical Analysis and Deep Learning

Samer Al-khateeb[1](✉), Jack Burright[1], Steven L. Fernandes[1], and Nitin Agarwal[2]

[1] Creighton University, Omaha, NE 68178, USA
{SamerAl-khateeb1,JackBurright,StevenFernandes}@creighton.edu
[2] COSMOS Research Center at UA-Little Rock, Little Rock, AR 72204, USA
nxagarwal@ualr.edu

Abstract. We refer to a "mob" as an event that is organized via social media, email, SMS, or other forms of digital communication technologies in which a group of people (who might have an agenda) get together online or offline to collectively conduct an act and then disperse (quickly or over a long period). To an outsider, such an event may seem arbitrary. However, a sophisticated amount of coordination is involved. Meetup.com is an "Event-Based Social Network" (EBSN) focused on bringing like-minded people together. Meetup hosts a wide range of events, making it crucial and well-suited for studying various events in general and mobs in particular. In this research, we collected data from Meetup and employed statistical analysis to help us better understand the data. Additionally, we utilized a deep neural network-based method to create two classifiers capable of predicting the Meetup mob outcome (success or failure) with great accuracy.

Keywords: Mobs Modeling · Deep Learning · Meetup · Statistical Analysis · Event-Based Social Network

1 Introduction

Meetup.com is one of the most used Event-Based Social Network (EBSN) sites, offering numerous events and extensive data through its API. It allows organizers to plan events related to any topic, ranging from very formal business meetings to casual events, e.g., movie nights [10]. Users can create and join groups based on their interests, and organize events such as mobs [7]. These mobs can be online (in cyberspace) or in person (in physical space). A mob does not have to be deviant (i.e., illegal and involving violence); some mobs are benign and aim to bring fun to a community, e.g., dance flash mobs. This makes Meetup.com a suitable platform to study mobs and mobbers' behaviors. Hence, in this research, we collect data from Meetup.com to help us answer the following research questions: **RQ1**: How can we utilize Descriptive Statistical Analysis (DSA) to comprehend

R. Thomson et al. (Eds.): SBP-BRiMS 2024, LNCS 14972, pp. 164–173, 2024.
https://doi.org/10.1007/978-3-031-72241-7_16

the collected Meetup data, and what insights can we derive from it? **RQ2**: How can we utilize Predictive Statistical Analysis (PSA) via deep learning to create a mob classifier capable of predicting the mob outcome (success or failure)? **RQ3**: How can we leverage and estimate a set of critical success factors (CSFs) and key performance indicators (KPIs) using event data collected from Meetup.com? Furthermore, how does utilizing these factors impact the performance of the deep learning classifier? To address the aforementioned research questions, we utilize DSA and PSA via deep learning.

2 Literature Review

There is a huge body of research related to the topics addressed in this paper; however, we briefly mention a few due to the paper's length requirements. For example, Schneider et al. [11] aimed to assess the potential of Meetup.com as a platform for promoting physical activity and fostering a sense of community. Huang et al. [7] created a deep learning model for predicting the growth and success of Meetup *groups*. Grundke et al. [3] developed a web crawler to collect data from Meetup.com and then utilized this data to conduct experiments aimed at enhancing the events recommendation page known as the "cold-start" page. Weinberg and Williams [13] pursued two research questions during Howard Dean's 2004 Democratic Party presidential nomination campaign when he used Meetup.com to arrange electronic events that would transition into in-person gatherings (referred to as "e2f" - electronic to face-to-face). Ricken et al. [10] conducted a study focusing on Meetup groups related to software development and aimed at addressing several key questions. Finally, Li et al. [9] focused on predicting the popularity of new groups on Meetup.com. This paper focuses on the method we used to answer our research questions. To the best of our knowledge, this work is the first to provide an estimation method for the CSFs and KPIs using Meetup data and to train a deep-learning model capable of predicting the outcome of Meetup mobs.

3 Methodology

In this section, we explain the methodology we followed to collect data from Meetup.com, the data prepossessing and enrichment steps needed to train our deep learning model, the DSA we conducted to have an overall understanding of the collected Meetup data, and the PSA we undertaken to predict the success or failure of mobs based on various inputs.

Data Collection: using Meetup.com's GraphQL API, we gathered data from 27 distinct Meetup groups, all featuring the topic "Flash Mobs". For each group, we collected data from every mob (event) they had organized. This data encompassed details such as the number of attendees, RSVP times, event description, and all comments and replies associated with the mob. This resulted in $3,536$ mobs with over $18,000$ RSVPs. We stored this data in a MySQL database.

Data Preprocessing and Enrichment: to be able to train our deep neural network, we needed to enrich the data, first, by running two dictionary-based tools, namely: Linguistic Inquiry and Word Count (LIWC) and the extended Moral Foundations Dictionary (eMFD) [6] which is based on the Moral Foundation Theory and Biblical Ethics (the six returned scores are split by vice and virtue) on the mob comments and the mob description. LIWC gives a percentage (from 0–100%) of all words in the text that fit a specific linguistic category, e.g., a *pos_emo* score of 4.5 means 4.5% of the words in the document have positive emotion. Also, the eMFD score is a percentage of all words in a text, but it returns scores that range from 0 to 1. For example, a *care.virtue* score of 0.24 would mean 24% of the words in the document are part of the care virtue words in the dictionary. All of these scores were stored in the same MySQL database.

The second data enrichment step involved calculating 13 more data attributes. These data attributes are called "Critical Success Factors" (CSFs) and "Key Performance Indicators" (KPIs) published by [4]. Table 1 shows the 13 CSFs and KPIs we estimated and used to run the second deep learning model which is explained later.

As a preprocessing step, we removed all events from private groups since we couldn't collect their comments. Moreover, events for which we couldn't calculate the number of individuals invited (using Eq. 1) were also excluded. As a result of these filtration steps, 459 mobs (24 cyber mobs and 435 physical mobs) remained for our analytical examination.

$$\#ofInvitedPeople = \#ofEventOrganizers + [\#ofMaxAllowedTickets * (1 + \#ofAllowedGuests)] \quad (1)$$

After the aforementioned preprocessing steps, two individuals in our lab randomly selected 378 mobs and manually labeled each mob as either "successful" or "failed" based on the following three criteria:

1. Presence of a photo depicting people at the event (participants in the mob).
2. Existence of a comment indicating someone's attendance or intention to attend the event.
3. Achievement of the event's target number of attendees.

If none of these three criteria were met, the mob was labeled as a failed mob. By applying these criteria, we manually labeled 211 mobs as successful and 167 as failed. This manually labeled data will serve as the ground truth for validating the two deep learning models explained below.

Statistical Analysis: *Descriptive Statistical Analysis* (DSA) describes the characteristics of the data (i.e., summarize the data) by measuring its central tendency (such as mean, median, and mode), variability (using variance and standard deviation), and frequency distribution (count). DSA aids analysts in gaining insights into the collected data without needing to examine individual data points. For example, a student's grade point average (GPA) offers valuable insight into their academic performance without the need to inspect each class grade [5]. We used DSA to have an understanding of the different mob types (topics), public vs. private mobs, the most popular locations of the mobbers (countries and cities), cyber vs. physical mobs, participation rate of these mobs, average mob *advertisement* time, average mobber *decision* time, and the correlation between participation rate and average mobber *decision* time.

DSA is good at narrating past events and understanding the data attributes through tables, graphs, and textual explanations. However, it cannot be used to make inferences or predictions about future values. To accomplish this, *Predictive Statistical Analysis* (PSA) is required. PSA is used to predict future values of an attribute, trends, or events using historical data. PSA helps in making strategic decisions and has a wide array of applications in finance, entertainment, marketing, and manufacturing, among others. PSA can be performed manually (a slower and more limited approach per application) or through the use of machine learning and deep learning algorithms (a faster and more versatile method) [2]. Therefore, in this paper, we employed an artificial deep neural network to forecast the success or failure of mobs based on various inputs.

Deep Learning Model-1: we used the *Sequential()* model provided by the Keras library from TensorFlow to train an artificial deep neural network with one input layer, two hidden layers, and one output layer (with 128, 64, 32, and 1 neurons, respectively). We added Batch Normalization to enhance the training speed, stability, and performance of the model. We also used LeakyReLU() activation function to take care of the negative values and a dropout of 0.5 after each layer to avoid overfitting. As an input for this model we used our manually labeled mob data that contained *85* attributes (i.e., the input_size) of information about the mobs. We have split the 378 labeled mobs into 70%:10%:20% for training, validation, and testing and used 5-fold cross-validation. This model was trained for 200 epochs where the data was grouped into batches of size 32.

Deep Learning Model-2: for this model, we used the same settings as model-1 above, however, the input_size for this model was 13 (which are the attributes we estimated using the methods highlighted in Table 1). The same training, validation, and testing splits, k-folds, epochs, and batches as model-1 were used here.

All data collection, preprocessing, enhancement, storage, and deep learning models training were done using a *Mac Pro - Tower* with the following specifications: 3.2 GHz 16-core Intel Xeon W processor, Turbo Boost up to 4.4 GHz Processor, 96 GB (6×16 GB) of DDR4 ECC memory. It also has 4TB SSD storage and the Radeon Pro W5500X with 8 GB of GDDR6 memory.

4 Results and Findings

In this section, we present and discuss our findings, grouping the results from the descriptive statistical analysis and the deep learning models based on the research questions addressed in this paper.

Descriptive Statistical Analysis Findings: to address the first research question, we used DSA to better understand the characteristics of the collected

Table 1. shows the CSF and KPIs we estimated and used to train our deep learning model 2. Columns 1 and 2 are borrowed from [4]. Column 3 explain how we estimated these factors using the collected Meetup data.

CSF	KPIs	Estimation Method using Our Meetup Data
Be Unique	Vividness	According to [4] for each user to be unique, the user has to have Vividness and Entertaining Content. We measure the event Vividness by counting the number of users that have attended events of only 1 group
Be Unique	Entertaining Content	We measure the event Entertaining Content by counting the number of URLs in the comments and replies of the event and the number of photos posted on the event
Interactivity	Interaction Rate	The number of replies to the comments of the event
Interactivity	Num Of Postings	The number of comments on the event
Interactivity	Recurring Rate	The number of mobbers that have attended at least 2 mobs from that group (if they just went to one event then they wouldn't have came back)
Increase Customer Happiness	Num of Positive Mentions	The average of *emo_pos* (positive emotion) from all the comments and replies posted on the event
Creative Ways To Address Users	Num Of Attended Events	This value is the number of unique topics hosted by the users that hosted the event divided by the number of organizers of that event
Address Target Group Consistent	Reach Within Target Group	The number of mobbers that have attended at least 2 mobs from that mob *topic*. For example, if the event topic is Flash Mob, how many mobbers attended events with the topic Flash Mob
Be Active	Net-reach	The size of the group (number of members) that hosted the event
Be Active	Num of Postings	The number of events that the group hosted the event have hosted
Unprofessiona lism	Num of Slang Words	This score is calculated using this formula: the *(big word score + clout score - swear score)*. The values used in the formula were calculated using the LIWC of the event description. If this score is positive, it means the event is professional, and if it is negative, it means the event is unprofessional
Building a Reputation	Num of Positive Mentions	The number of comments and replies that have a higher *emo_pos* score than *emo_neg* score
Privacy Protection	–	If the mob is online it gets a 1, if it is in person it gets a 0

data. Initially, we examined the "topics" of the mobs under analysis, which can also be regarded as types of mobs. This analysis helps us understand the nature of the data (the mobs). The 459 mobs we analyzed were organized by 16 different groups, with sizes ranging from 33 members to 9,203 members. These mobs were tagged with 21 different topics such as "Social" (45 mobs), "Outdoors", "Theater", "Fun Times" (each with 29 mobs), etc. Note that more than one topic can be assigned to a single mob. Upon examining these topics, we found that all analyzed mobs were benign (no deviant mobs were included), which is expected considering that the groups that organized these mobs are public.

We also wanted to examine the diversity of the mobs data, so we analyzed the location of the mobbers. We found that mobbers are from different parts of the world, and most of them are located in big cities such as New York, Sydney, and London.

Given that out of the 459 mobs we collected, 345 are physical mobs (in-person) and 24 are cyber mobs (online), we wanted to examine the difference in participation rates to determine whether mobbers participate more in online or in-person mobs. We estimated the participation rate of these mobs using Eq. 2. We found that cyber mobs exhibit a higher average participation rate compared to physical mobs. This trend could be attributed to various factors, including the assumption that cyber mobs entail less risk or that cyber mobs are easier to participate in compared to physical (in-person) mobs.

$$ParticipationRate = \frac{\#ofInvitedPeopleRespondedWithYes}{\#ofInvitedPeople} \tag{2}$$

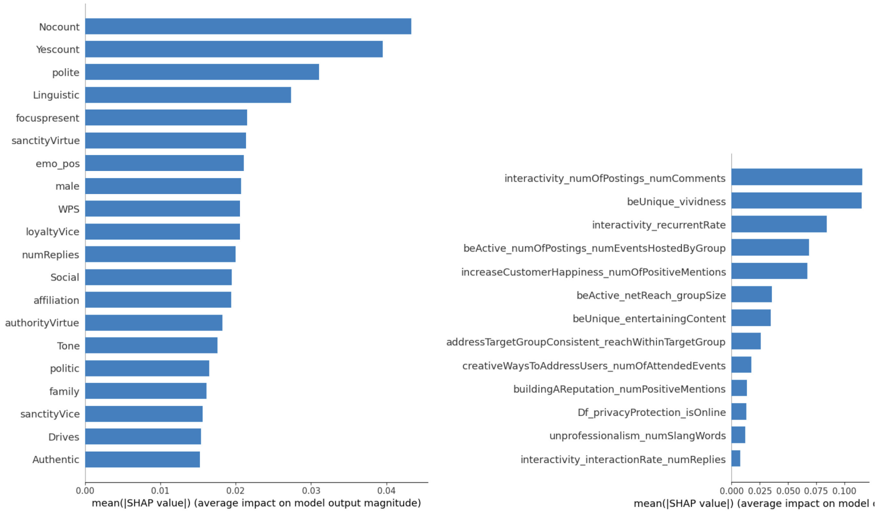
It's also important to understand the lead time mobbers take to advertise their mobs (i.e., the "*recruitment phase*" [1]). Hence, we measured the time difference between the creation of the event on Meetup.com and its scheduled occurrence. By analyzing the time difference we found that mob organizers tend to advertise their mobs well in advance, averaging around 23 days for all the collected mobs. Moreover, we observed that mobs requiring substantial training and effort, such as "Singing Lessons" or "Choir" (ranked top 1 longest time out of 51, averaging $90,230$ minutes which is equivalent to 62.7 days) and "Dance Fitness" (ranked top 6 out of 51), are advertised for significantly longer periods compared to mobs with lower training requirements like "Partying" (46 out of 51), "Music" (50 out of 51), or "Brazilian Culture" which was ranked last with an average advertisement time of $1,634$ minutes, i.e., 1.13 days. It's worth noting that mobs with the "Flash Mobs" topic ranked 15 out of 51, averaging an advertisement time of $45,370$ minutes, i.e., 31.51 days.

Besides considering the advertisement time for mobs, it's also important to examine the duration mobbers take to decide whether they will participate in a mob or not. Therefore, we measured the time difference between the event creation time on Meetup and the moment each mobber responded with either a "Yes" (to attend) or "No". By analyzing the time differences, we found that, on average, individuals invited to participate in a mob take longer to decline (respond with NO) than to accept (respond with YES). The average time taken for a "Yes" response across all mobs is 22,857.16 minutes (equivalent to 15.87 days), whereas the average time taken for a "No" response across all mobs is 33,394.06 minutes (which amounts to 23.2 days).

Finally, we calculated Spearman's correlation coefficient (SCC) to determine the relationship between the 459 Meetup.com mobs participation rate and the average mobbers' time to respond with "Yes". We found a strong, positive monotonic correlation (i.e., a high SCC value) between the mob participation rate (calculated using Eq. 2) and the average mobbers' time to say yes ($SCC = 0.68$, $n = 459$, $p < 0.001$), the average mobbers' time to say no ($SCC = 0.64$, $n = 459$, $p < 0.001$), and the average mobbers' time to respond with either yes or no ($SCC = 0.67$, $n = 459$, $p < 0.001$). Positive monotonic correlation means as the values of one variable increase, the values of the other variable also tend to increase. It doesn't mean that the increase is constant; it only means that higher values of one variable are associated with higher values of the other, even if the relationship is curved or uneven [14].

Predictive Statistical Analysis (via Deep-Learning) Findings: as stated earlier, PSA can be performed manually or through the use of machine learning and deep learning algorithms [2]. Therefore, in this paper, we employed a deep learning algorithm to forecast the success or failure of mobs based on various inputs. Our goal here is to answer *RQ2* and *RQ3*.

Result of Deep Learning Model-1: to address the second research question, we used $Model-1$ described earlier. Using the training, validation, testing data, and 5-fold cross-validation, we found that the mean accuracy for the validation set is 88.64%, and the test set accuracy is 92.11% with a loss of 0.2715. *Accuracy* measures the number of times the model can correctly detect the positive and negative classes. We also calculated the *precision*, which measures "the success probability of making a correct positive class classification" [8], and *recall*, which measures the models ability to minimize false negative scores and were found to be 0.913043 and 0.954545, respectively. This gave a harmonic mean of precision and recall (i.e., F1-score, which "takes into account the type of errors - false positive and false negative - and not just the number of predictions that were incorrect" [12]) of 0.933333. To determine which attribute (out of the 85 attributes used in this model) is most important in determining the success and failure of a mob and to understand the predictions of the machine learning models, we used the SHAP Python library to calculate the SHAP values. SHAP is an additive feature attribution method, meaning the prediction is explained as the sum of the effects of each feature. It is derived from Shapley values, a concept in cooperative game theory developed by Lloyd Shapley. This method provides local and global insights into feature contributions (importance) and model behavior. We found that the number of people responding with "No" or "Yes" and the "Polite" score of the *event description* are the top three most critical attributes in determining success and failure (see Fig. 1-a). Conversely, other linguistic measures calculated from the event description text, such as "Drives", "Authentic", etc., hold less importance but are still in the top 20 most important attributes (out of 85 attributes).



(a) (b)

Fig. 1. Rank of the most important attributes in classifying mobs for both models. (a) shows the average impact of the top 20 attributes of Model-1 while (b) shows the average impact of each of Model-2's attributes.

Result of Deep Learning Model-2: To address the third research question, we estimated the CSFs and KPIs using the estimation methods shown in Table 1. We then used these measures as an input to $Model-2$ described earlier.

Using the training, validation, testing data, and 5-fold cross-validation, we found that the mean accuracy for the validation set is 93.17%, and the test set accuracy is 98.68% with a loss of 0.0417. We also calculated the precision and recall scores, which were 0.973684 and 1.000000. This gave an F1-score of 0.977778. The higher test accuracies for both models compared to the mean validation accuracies in k-fold cross-validation suggests that the models are not over-fitting. Since Model-2 performed very well and better than Model-1, we used it to predict the label of another set of unseen mobs data (containing 80 physical mobs and 1 online mob). Using Receiver Operating Characteristic Curve (ROC curve) best threshold of 0.336709, the model labeled 54 mobs of the unseen data as successful and 27 as failed mobs. Finally, we examined which CSFs and KPIs, out of the 13 listed in Table 1, are most important in determining the success and failure of a mob. We found the CSF "Interactivity" and the KPIs "Num of Postings", estimated via the number of comments the mob received, to be the most important in determining the success or failure of the mob. Many successful mobs had high scores in this measure while many failed mobs had low scores in this measure, indicating mobs with a lot of interaction from the group have a higher chance of succeeding. Also, the CSF "Be Unique" and KPI "Vividness", is ranked as the second most important factor. Also, many mobs had a high score in this measure were successful. Finally, the CSF "Interactivity" and the KPI "Recurrent Rate", estimated via the number of mobbers that have attended at least 2 mobs from that group, to be the third most important factor in determining the success or failure of the mob. Many successful mobs had high scores in this measure, indicating that the participation of committed mobbers will most likely make a mob succeed. Conversely, the CSF "Interactivity" and the KPI "Interaction Rate" measured via the number of replies seems to be the least important factor in determining the success and failure of a mob. See Fig. 1-b for information about the other measures.

5 Conclusion and Future Research Directions

Meetup differs from other social media sites such as Facebook in how members develop their connections. Users take their offline connections on Facebook and then connect with them online. Meetup is the opposite; users can join groups, connect with people online, and then meet them face-to-face. Another example of a platform that has this same interaction would be the dating site Match.com, where users match online and then can go on dates in person [13]. This makes Meetup a crucial platform to study mob creation and mobbers' behaviors because in a mob, a group of like-minded people, who may or may not know each other, get together online or offline to collectively conduct an act and then disperse. In this paper, we collected data from Meetup.com and conducted two types of statistical analysis: descriptive (DSA) and predictive (PSA). For the PSA, we trained two deep-learning models: one using 85 attributes, while the other used 13 attributes and achieved better performance. Additionally, we ranked the importance of all the attributes used in both models. To the best of our knowledge, this work is the first to provide an estimation method for the CSFs and KPIs using Meetup data and to train a deep-learning model capable of predicting the success or failure of Meetup mobs with high accuracy.

Even though we ran multiple experiments that resulted in the aforementioned models' accuracy, our work is limited by the public Meetup data we collected. The models should be able to predict the success or failures of the mobs with very high accuracy, but only for mobs organized by public Meetup groups. We could not test our model on mobs organized by private groups due to a lack of data and privacy issues. Also, our model is trained on Meetup data, so it might not be able to predict the success or failure of mobs organized on other social media sites such as Facebook or X (formerly known as Twitter).

So for future research direction, we plan to leverage the findings of this research to build an agent-based model to simulate mobs. The simulation model will provide a more generic method to study mobs beyond Meetup mobs.

Acknowledgements. This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research

Alliance, and the Jerry L. Maulden/Entergy Endowment at the UA-Little Rock. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

1. Al-khateeb, S., Agarwal, N.: Flash mob: a multidisciplinary review. Soc. Netw. Anal. Min. **11**(1), 97 (2021)
2. Cote, C.: What is predictive analytics? 5 examples — HBS online (2021). https://online.hbs.edu/blog/post/predictive-analytics
3. Grundke, M., Jung, J., Sachse, J.P., Wiese, G.: Meetup miner:measuring event interestingness on meetup (2023). https://www.parnswir.de/blobs/smm_meetup.pdf
4. Hammerl, T., Schwaiger, J.M., Leist, S.: Measuring the success of social media: Matching identified success factors to social media kpis. In: Proceedings of the 52nd Annual Hawaii International Conference on System Sciences, Maui, Hawaii, 8–11 January 2019, pp. 2427–2436. ScholarSpace, Honolulu (2019). https://epub.uni-regensburg.de/41362/
5. Hayes, A.: Descriptive statistics: definition, overview, types, example (2023). https://www.investopedia.com/terms/d/descriptive_statistics.asp
6. Hopp, F.R., Fisher, J.T., Cornell, D., Huskey, R., Weber, R.: The extended moral foundations dictionary (emfd): development and applications of a crowd-sourced approach to extracting moral intuitions from text. Behav. Res. Methods **53**, 232–246 (2021)
7. Huang, T., Liang, Y., Sun, Z.: Predicting group success in meetup with graphs (2020). https://snap.stanford.edu/class/cs224w-2019/project/26422233.pdf
8. Jankay, Y.: The 3 pillars of binary classification: accuracy, precision & recall (2018). https://medium.com/@yashwant140393/the-3-pillars-of-binary-classification-accuracy-precision-recall-d2da3d09f664
9. Li, G., Liu, Y., Ribeiro, B., Ding, H.: On group popularity prediction in event-based social networks. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 12, no. 1 (2018). https://doi.org/10.1609/icwsm.v12i1.15067. https://ojs.aaai.org/index.php/ICWSM/article/view/15067
10. Ricken, S., Barkhuus, L., Jones, Q.: Going online to meet offline: organizational practices of social activities through meetup. In: Proceedings of the 8th International Conference on Communities and Technologies, pp. 139–148 (2017)
11. Schneider, K.L., et al.: An online social network to increase walking in dog owners: a randomized trial. Med. Sci. Sports Exerc. **47**(3), 631 (2015)
12. Sharma, N.: Understanding and applying f1 score: a deep dive with hands-on coding (2023). https://arize.com/blog-course/f1-score/
13. Weinberg, B.D., Williams, C.B.: The 2004 US presidential campaign: impact of hybrid offline and online 'meetup' communities. J. Direct Data Digit. Mark. Pract. **8**, 46–57 (2006)
14. Weir, I.: Spearman's rank correlation- introduction (2023). https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf