Characterising Fake Instagram accounts*

Arend Geerlofs, Alina Baciu, Roger Ruiz Salvat, and Michael Beckers

Vrije Universiteit Amsterdam

Abstract. This paper presents a study on characterizing fake and viral Instagram accounts. To investigate the account descriptions, we performed emoji sentiment analysis. Furthermore, We employ a machine learning technique to analyze various quantitative features of Instagram accounts. This includes the number of followers, the length of usernames and descriptions, and the presence of numeric characters. The study also utilizes a named entity recognition model to detect entities in the text features of the Instagram accounts. The results reveal significant differences between fake and viral accounts in terms of the number of emoji sentiment, following accounts and follower accounts. Additionally, the analysis shows that fake accounts tend to have fewer personal, organizational, and geographical entities in their descriptions and full names compared to viral accounts. The paper concludes that these findings can contribute to the identification and classification of fake Instagram accounts.

Keywords: Fake Instagram accounts · Machine Learning · Feature Importance · Named entity recognition

1 Introduction

As the internet grew into the social web, people started becoming more and more connected. This process, while occurring all over the web, can most notably be seen in the growth of social media platforms over the years. Since the creation of social media platforms like Facebook in 2005, Twitter in 2006 and Instagram in 2010, these social platforms have started amassing huge amount of users and traffic. With Facebook, Twitter and Instagram having 3, 3 and 2 billions active monthly users respectively.

However, with this rise also came the rise of fake or bot accounts. These accounts can and are used to dictate discourse, steer opinions and have even be used to influence elections [1]. To reduce the effect of bots, social media platforms have to try to detect which accounts are fake and remove them from their platforms. However, as can be seen from the fact that about 8.5% of all twitter users are compromised of bots or fake accounts, this is not a trivial task [2]. This means that about $\frac{1}{12}$th of all accounts on X, previously known as Twitter, are non-real accounts.

* Supported by organization VU

In this study, the characteristics of Instagram accounts of both fake and real accounts have been analysed and compared. With the goal being to answer the research question:

What are the most important characteristics in detecting which Instagram? accounts are bots.

This was done by three main methods. Namely: Machine learning Feature Importance, Emoji sentiment analysis and a Named Entity Recognition (NER) model. Using these models the research question was answered.

The paper starts with a small overview into the relevant literature and similar studies which have been done on the detection of fake accounts on social media platforms. After this, a breakdown of the data is given. Thirdly, the methods are explained for the used models along side the data preprocessing. Fourthly, the results are shown. The paper ends with a short summary of the main conclusions, the privacy, ethical considerations and data security and the limitations of the research.

2 Literature Review

A lot of research has been done into the detection of bots on social media platforms. However, due to the fact that real accounts can often show bot-like behaviour, about 9% to 15% of all English twitter accounts do show that behaviour [3]. It is challenging to detect which accounts are actual bots and or fake accounts and which accounts just look like them.

A study on the detection of bots highlights the emergence of 'social bots'. These bots are capable of mimicking human behaviour even better [4] by having daily activity patterns [5] or even having conversations and interacting with other accounts [6]. To still be able to detect these advanced bots, multiple types of detection mechanisms are used. These can be categorised in three main classes [7]:

1 Social network information detection
2 Crowdsourcing and the leverage of human intelligence
3 Feature based detection

The last of these three, is especially important since it can be used in tandem with machine learning algorithms to not only create effective and accurate systems but are also simple to implement in an autonomous manner. This autonomous feature is important to combat the amount of bots that exist which, according to Kerrisa and Utami, is currently more than a million [8].

Numerous versions of machine learning algorithms which try to detect bots exist, with each having different feature they prioritise and use. One of the first machine learning algorithm which detected bots is called 'BotOrNot'[9]. This algorithm made use of more than a 1000 features to try and detect bots. These features were classified into 6 main categories [9]:

1 Network features, which captured information about the spread and collections of data the bots used.
2 User features, which uses meta-data of the account such as location, account creation time and language.
3 Friend features, including amounts of followers, friends, posts and other social contact features.
4 Temporal features, i.e. tweet rates and tweet-time distributions.
5 Content features, also known as linguistic features which consist of natural language models such as part-of-speech tagging and bag of words.
6 Sentiment features, which use models based on sentiment analysis to assign happiness, emoticon and arousal-dominance-valence scores.

Most machine learning algorithms today use somewhere between 4 and 49 features, each of them belonging to one of the before mentioned categories.More specifically, most features were related to 8 main features. These are: number of followers, the number of following accounts, likes, profile pictures, status, posts, and account names. The main upside of these features is that researchers, in most cases, do not need special permission to obtain these features from accounts. [8]

To analyse these features and use them to detect which accounts are bots and which are real, a lot of different machine learning algorithms are used. With each social media platform having different machine learning algorithms being favoured. For Instagram, ANN, Naïve Bayes, Random Forest and SVM are most commonly used [8]. With the Random Forest algorithm being the most accurate with an F-1 score being obtained of around 98 [10] [11].

3 Data

The samples used for the analysis of fake and viral accounts were collected from two different sources. Then, they were combined into a single dataset with both fake and viral account instances. From the combined dataset, two versions with the same instances were generated. One version with text data, and one with numeric data.

3.1 Sources

Collection of fake accounts Fake account records were obtained from a dataset published in `kaggle.com` by the user *reza jafari*. The data is distributed among several JSON files, each containing the information for one account. 693 files were processed, all using the same JSON structure.

Collection of viral accounts The viral accounts were collected from a dataset published in `data.world` by the user *Pio Foco*. Similarly from the fake accounts, the viral accounts were distributed into JSON files, each file corresponding to a viral account. All files were using the same JSON structure, although it was different from the structure used for fake account data files.

Emoji sentiment ranking dataset Additionally, in order to do sentiment analysis on the emojies used in the descriptions of Instagram accounts, a dataset containing mappings of emojies to sentiment scores was used. The dataset [12], was build from 70,000 tweets, labeled by 83 human annotators. Per emoji, it contains a count of positive, neutral, and negative occurrences in tweets. More information can be found at `https://kt.ijs.si/data/Emoji_sentiment_ranking/index.html`.

3.2 Data used for analysis

Numeric dataset With 477 instances and 11 columns including the class (fake/viral). These columns are:

- The number of accounts that the account is following
- The number of followers of the account
- The length (number of characters) of the username
- The length (number of characters) of the full name
- The length (number of characters) of the description
- A boolean value indicating if the username has a numeric character
- A boolean value indicating if the full name has a numeric character
- A boolean value indicating if the description has a numeric character
- The average sentiment score of the emojies used at the description
- The average neutrality of the emojies used at the description
- The class varialble (fake/viral)

Text dataset With 477 instances and 4 columns including the class (fake/viral). These columns are:

- Unique username of the account
- Non-unique full name assigned to the account
- The description of the account profile
- The class variable (fake/viral)

4 Methods

4.1 Data preprocessing

From the original format in which data is collected (3.1), several preprocessing steps were followed in order to obtain datasets with the appropriate format for the analysis.

Firstly, the account data was transformed from the JSON format into a table in which each row belongs to an account. This operation was done using the Python package pandas [13]. Then, the table was split into two datasets, one containing text data, and the other one with numeric data.

For the text version of the dataset, natural language processing (NLP) techniques were applied using the Python package NLTK [14]. These methods include lower casing, lemmatization, removal of stop-words, of punctuation, of URLs, of extra spaces, of numbers, etc. Additionally, some descriptions were translated into English to make the dataset be uniform in language.

For the numeric version of the dataset, Min-Max normalization was applied to scale all feature in the range [0, 1].

4.2 Emoji Sentiment Analysis

In this study, we conducted sentiment analysis on emojis used in Instagram account descriptions, employing two key metrics: sentiment score and neutrality. The sentiment score was calculated by subtracting the negative occurrences from the positive occurrences of each emoji in tweets [15], while neutrality represents the frequency of neutral occurrences. Both metrics are expressed relative to the total amount that each emoji appears in tweets.

4.3 Machine learning Feature Importance

Then, we applied the Random Forest Classifier, a machine learning technique, to investigate the feature importance in distinguishing between fake and viral Instagram accounts. To achieve this, we utilized the scikit-learn library [16] in Python to implement the Random Forest Classifier, which can make accurate predictions by combining multiple decision trees with an ensembling method [17].The dataset was then randomly divided into training and testing sets for model training and evaluation. In order to reduce model overfitting, hyperparameters were fine tuned and was followed by cross-validation. Although a high performing model would be preferable, the goal of this study is to focus on the features themselves and how they contribute to characterizing Instagram accounts. Therefore, we computed feature importance scores based on the mean decrease in impurity. This score, also named "Gini Importance", is also a motivator why RFC was chosen for this project, as it is a built-in mechanishm of the algorithm that identifies the most influential features in our dataset [17]. This approach aimed to unveil the significance of individual features in differentiating between real and fake accounts, providing valuable insights into the characteristics influencing the RFC's decision-making process. Lastly, in order to validate the differences, independent two-tailed t-tests were performed.

4.4 Named Entity Recognition (NER) Model

spaCy is one of the many Python libraries used in Natural Language Processing, to analyse text data. It is employed in various tasks like part of speech tagging, dependency parsing, rule-based matching and named entity recognition. The latter operation involves identifying and categorising different entities (e.g. persons, organisations, locations) that can used in further research activities. spaCY provides a pre-trained model for this task, which was imported in our environment using en-core-web-sm [18] [19].

4.5 Statistical testing

Statistical testing was used to discover significance differences of the sentiment of the emojies used in Instagram account descriptions. As the samples used for analysis are independent samples with similar variances, the chosen method is the student t-test from the scikit-learn Python package [16]. Statistical testing was also applied to assess the significance of the determined text fields patterns that distinguish fake accounts from viral ones. One-tailed independent two-sample t-tests were used from the stats package of the scipy library [20], assuming unequal variances.

5 Results and Discussion

5.1 Emoji sentiment analysis

The results accompained with histograms and descriptive statistics are presented as it follows.

Histograms and descriptive statistics From two independent samples of 128 account descriptions, the sentiment score of emojies of viral and fake accounts is statistically described in table 1.

Sentiment score		
Type	Viral	Fake
Count	128	128
Mean	0.43	0.56
Std	0.24	0.29
Min	-0.5	-0.5
25%	0.26	0.41
50%	0.41	0.54
75%	0.58	0.82
Max	0.95	1

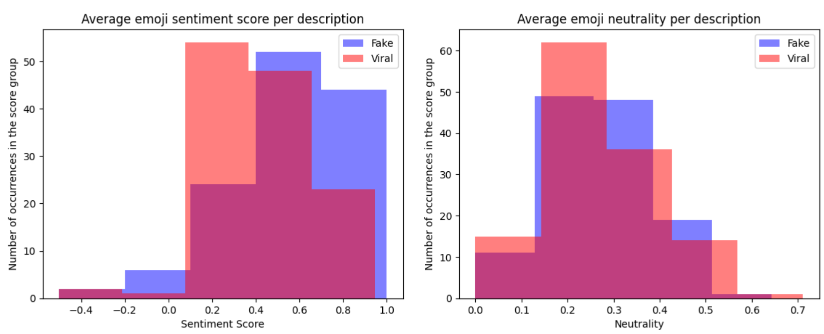
Table 1: Sentiment score descriptive statistics

From the same samples, the neutrality is statistically described in table 2. These tables detail the descriptive statistics, including counts, means, standard deviations, minimum and maximum values, and quartiles, for both sentiment scores and neutrality scores in viral and fake Instagram accounts.

Neutrality		
Type	Viral	Fake
Count	128	128
Mean	0.27	0.27
Std	0.18	0.11
Min	0	0
25%	0.18	0.2
50%	0.24	0.27
75%	0.36	0.33
Max	0.71	0.64

Table 2: Neutrality descriptive statistics

The sample distribution of both the sentiment scores and the neutrality scores are shown in figure 1.



(a) Sentiment score histogram for both fake and viral accounts

(b) Neutrality histogram for both fake and viral accounts

Fig. 1

Table 1 and figure 1a indicate that fake Instagram accounts have a higher mean sentiment score (0.56) compared to viral accounts (0.43). This may suggest that fake accounts use emojis with a more positive sentiment for their account description.

Moreover, table2 and figure1b shows that both viral and fake accounts have similar mean neutrality scores (0.27), suggesting the accounts using neutral emojis in their descriptions. However, the narrower standard deviation for fake accounts (0.11) compared to viral accounts (0.18) may indicate a more consistent pattern in the use of neutral emojis among fake accounts compared to viral ones.

Hypothesis on sentiment score The purpose of this statistical test is to determine whether there is significance difference in the mean of the sentiment score of the emojies found at account descriptions of fake and viral accounts.
With a significance level of $\alpha = 0.05$ the one-tailed hypothesis test is approached as follows:

$$H_0 : \mu_f = \mu_v$$

$$H_a : \mu_f > \mu_v$$

And the obtained p-value:

$$p\_value = 5.58 \times 10^{-5}$$

Therefore, *We reject the null hypothesis.*
There is significant difference in the mean of the sentiment score of the emojies found at account descriptions. These findings suggest that fake accounts tend to include emojies with a higher sentiment score in their descriptions.

Hypothesis on neutrality This statistical test aims to determine whether there is significance difference in the mean of neutrality scores of the emojies found at account descriptions of fake and viral accounts
With a significance level of $\alpha = 0.05$ the two-tailed hypothesis test is approached as follows:

$$H_0 : \mu_f = \mu_v$$

$$H_a : \mu_f \neq \mu_v$$

And the obtained p-value:

$$p\_value = 0.71$$

Therefore, *We fail to reject the null hypothesis.*
There is no significant difference in the mean of the neutrality score of the emojies found at account descriptions. Accordingly, there is no evidence that would indicate that fake and viral accounts include different proportions of neutral emojies in their descriptions.

5.2 RFC Feature Importance

Figure 2 shows the calculated importance score of each numeric feature in the dataset, also including the emoji neutrality and sentiment scores retrieved from the emoji sentiment analysis. As the feature importance is normalized, the sum of all these scores is equal to 1. Figure 2 shows that the "followers" has the highest importance score, followed by "following" and "username has number" respectively. Both emoji scores seems to have a low importance score, while "full name has number" and "description has number" have an importance score of near zero.

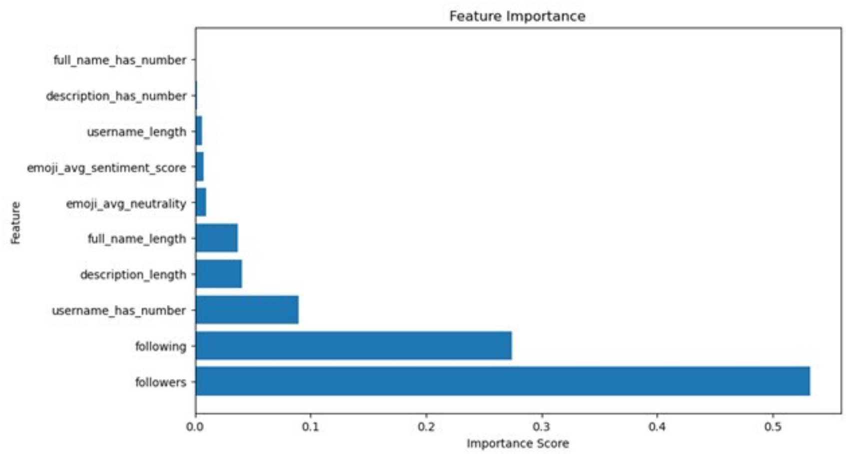


Fig. 2: Feature importance graph

The boxplots of the features with the highest score, that is "following" and "followers", are depicted in figure 3 and figure 4 respectively. These boxplots allows us to further examine the differences between these features between viral and fake accounts. In both figures. 0.0 (blue) are viral accounts and 1.0 (orange) depict the fake accounts. For the "following" boxplot, the fake account boxplot is positioned higher that the one of viral accounts. This indicates that, on average, fake accounts tend to follow more Instagram accounts than viral accounts do. The opposite effect is seen in the followers boxplot in figure 4, where the box of viral accounts is positioned higher indicating that, on average, viral accounts have more followers than fake accounts.

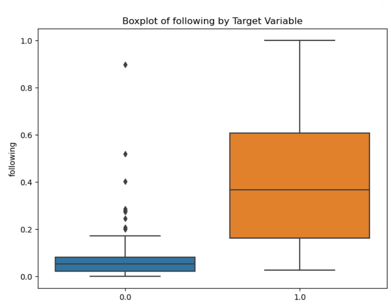


Fig. 3: Following Boxplot

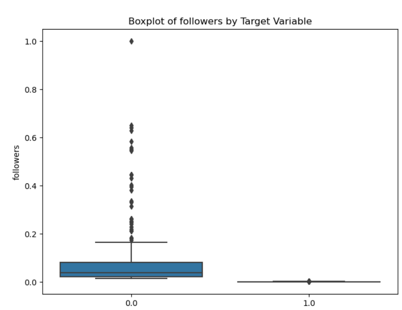


Fig. 4: Followers Boxplot

Lastly, two-tailed t-tests were performed in order to validate if these differences can be considered as significant. With a significance level of $\alpha = 0.05$ both test are approaced as follows:

$$H_0 : \mu_f = \mu_v$$
$$H_a : \mu_f \neq \mu_v$$

The obtained p-value for "following":

$$p\_value = 4.29e - 49$$

The obtained p-value for "followers":

$$p\_value = 5.80e - 16$$

Therefore, *We reject the null hypthesis* for both feature tests.
There is a significant difference in the mean of the "following" and "followers" features.

5.3 Description and Full-names analysis - NER model

The named entity recognition model, presented in section (3), was used to detect entities in all the text features of dataset: description, full names and user names of Instagram accounts. Concerning the statistical tests (3.4), a significance level of 0.05 was used. The findings of the analysis illustrated the statistical significance of the following patterns:

- descriptions - fake accounts include less PERS, ORG, GPE entities than viral accounts
- fullnames - fake accounts include less GPE entities than viral accounts

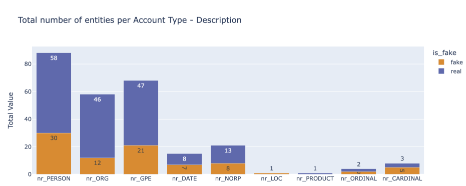


Fig. 5: Caption for the first figure

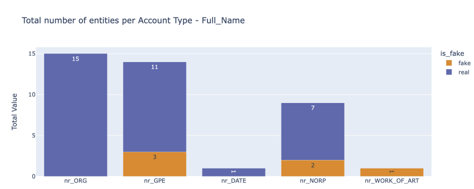


Fig. 6: Caption for the second figure

6 Conclusion

6.1 Summary

So in order to answer the research question of: "What characteristics can be utilised to effectively distinguish between fake and viral Instagram accounts?", we employed three techniques in order to examine different characteristics with purpose of distinguishing viral from fake instagram accounts. We first obtained sentiment and neutrality scores concerning the presence of emojis in instagram account descripitions. It was then concluded that fake accounts tend to include emojies with a highers entiment score than viral accounts.

Secondly, For the quantative numeric dataset, the feature importance scores indicate that the most influential characteristics were the amount of accounts that were followed and the amount of followers of an account. This is in accordance with what was expected based on the literature and other papers in this field of research.

Lastly, concerning qualitative text dataset, namely the description and fullnames variables, our findings indicate the following: persons, organisations, and countries/cities/states entities appear less in fake accounts' descriptions than in viral accounts' descriptions. Also, countries/cities/states entites appear less in fake accounts' fullnames than in viral accounts 'full names. These results are not aligned with the existing literature as the process of identifying entities within the scope of distinguishing fake and viral accounts has not been extensively explored.

6.2 Privacy, Ethical Consideration and Data Security

The distinction between fake and real Instagram accounts topic brings two important issues concerning the privacy of genuine users. These are the consequence of the (in)direct interaction of the victim with the false entity and the ethical surveillance concern that may arise when monitoring potential fake suspects by legal authorities or regular individuals.

To begin with, the harmful effects on genuine users' privacy can vary based on the classification of these false entities, which can be programmed bots or artificially created accounts [21]. Concerning the first category, which typically implies an indirect interaction with users, a negative effect could be the mass theft of personal data [21]. Regarding the second type, which generally implies a direct connection with users, examples of adverse outcomes could be identity theft, harassment and financial scam [22]. Overall, both types of false entities perform invasion of personal spaces.

Secondly, the rise of Big data and the application of appropriate techniques to analyze it has been given researchers the potential to uncover valuable information in various areas, one of them being the social network environments. Many inspection activities of users' behavior are conducted to prevent potential harmful events, thus protecting separate individuals but also the society. As also stated in the Instagram 2022 Data Policy, one of the reasons the platform processes user information is to verify suspicious accounts to promote safety and integrity [23].

Despite this enormous benefit, according to [24], the end-users of social network sites are one of the entities that can be ethically impacted by such studies. In their paper regarding employing fake personas for social network research, [24] recognize several ethical concerns that can be addressed, two of them being relevant for this study: indirect exposure and exposure of human weaknesses. The stated considerations, along with the lack of notification, can be applied when monitoring users' behavior to analyze and detect potential false accounts, for the following reasons.

Firstly, concerning the lack of notification, any genuine public account could be targeted and be part of investigations for an unlimited amount of time without them ever being informed afterwards. If discovered at a later stage, it can significantly impact users' confidence and their trust in the social network platform. The indirect exposure implies that if genuine users have any sort of interaction with the fake entity, they might be unfairly associated with it, since false accounts might operate together to appear credible. Ultimately, the exposure of human weaknesses implies surveillance bias against certain individuals based on their vulnerabilities. For instance, real persons with mental conditions might adopt an unusual communication pattern, a different style than the one of typical users, that might alert entities performing monitoring activities. All these ethical considerations affect the privacy of genuine users.

Data Security represents the essential practice designed and conducted to protect users' rights and to prevent privacy issues from occurring. Performing accurate identification of fake Instagram accounts while considering the previously discussed private and ethical implications, represents the right balance to maintain and ensure Data Security in the social network environment.

6.3 Limitations

From data collection several limitations are raised that affect the outcome of the study. Due to Instagram API restrictions for accessing user data, the data had to be gathered from data sharing platforms. There was no time or geographical information found from the sources where the data is published. This fact did not allow for defining a clear context, and consequently interpret the results within that context. Moreover, with that information present, further aspects of social networks could be researched. Social network analysis which make geographical considerations have proved to provide better context descriptions and the evolution of social networks [25]. On the other hand, time considerations with the application of several network analysis methods, can provide insightful information about social networks dynamics. Examples of this methods are the temporal eccentricity, temporal distance, temporal betweenness and closeness centrality [26].

The are a number of limitations concerning our study. Regarding the use of the Random Forest Classifier, its accuracy was 100 percent of our dataset which could be a sign of overfitting even with the measures (paramater tuning and cross validation) taken to reduce that risk. This could be an indicator that our dataset was not large enough for the RFC model to train. Further research could therefore focus on collecting a larger number of fake Instagram. Furthermore, multiple machine learning methods could be implemented to compare the feature importance scores among them. However, due to time constraints, this was not feasible for this study. Concerning the use of the named entity recognition (NER) model, developed by spaCY, the classifier itself has a main limitation, which is its imperfect accuracy. This implies that some words may be classified wrong and therefore additional assessments would be recommended to ensure the full correctness of the analysis.

References

1. Tobias R Keller and Ulrike Klinger. Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, 36(1):171–189, 2019.
2. Venkatramanan S Subrahmanian, Amos Azaria, Skylar Durst, Vadim Kagan, Aram Galstyan, Kristina Lerman, Linhong Zhu, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. The darpa twitter bot challenge. *Computer*, 49(6):38–46, 2016.
3. Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the international AAAI conference on web and social media*, volume 11, pages 280–289, 2017.
4. Norah Abokhodair, Daisy Yoo, and David W McDonald. Dissecting a social botnet: Growth, content and influence in twitter. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 839–851, 2015.
5. Scott A Golder and Michael W Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.
6. Tim Hwang, Ian Pearce, and Max Nanis. Socialbots: Voices from the fronts. *interactions*, 19(2):38–45, 2012.
7. Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Communications of the ACM*, 59(7):96–104, 2016.
8. Nalia Graciella Kerrysa and Ika Qutsiati Utami. Fake account detection in social media using machine learning methods: literature review. *Bulletin of Electrical Engineering and Informatics*, 12(6):3790–3797, 2023.
9. Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web*, pages 273–274, 2016.
10. Er Pranay Meshram, Rutika Bhambulkar, Puja Pokale, Komal Kharbikar, and Anushree Awachat. Automatic detection of fake profile using machine learning on instagram. *International Journal of Scientific Research in Science and Technology*, 8:117–127, 2021.
11. Subham Das, Sourav Saha, S Vijayalakshmi, and Jitendra Jaiswal. An effecient approach to detect fraud instagram accounts using supervised ml algorithms. In *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pages 760–764. IEEE, 2022.
12. Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Emoji sentiment ranking 1.0, 2015. Slovenian language resource repository CLARIN.SI.
13. The pandas development team. pandas-dev/pandas: Pandas, February 2020.
14. Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

15. Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PloS one*, 10(12):e0144296, 2015.
16. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
17. Bjoern H Menze, B Michael Kelm, Ralf Masuch, Uwe Himmelreich, Peter Bachert, Wolfgang Petrich, and Fred A Hamprecht. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*, 10(1):213, July 2009.
18. Prateek Joshi. spacy tutorial to learn and master natural language processing (nlp). 2023.
19. Named entity recognition. 2022.
20. scipy.stats.ttest$_ind.$2023.
21. Svitlana Chupakhina Yulia Romanyshyn Mykola Pasieka Nadiia Pasieka, Myroslava Kulynych. Harmful effects of fake social media accounts and learning platforms. *CEUR*, pages 258–271, 2021.
22. Bytescare. What is fake profile? 2013.
23. Instagram Meta Platforms. Data policy. 2022.
24. Fire Michael Elovici Yuval Shulman Haya, Herzberg Amir. Ethical considerations when employing fake identities in online social networks for research. *Science and Engineering Ethics*, pages 1027–1043, 2013.
25. Johannes Glückler and Patrick Doreian. Social network analysis and economic geography—positional, evolutionary and multi-level approaches. *Journal of Economic Geography*, 16(6):1123–1134, 2016.
26. Nicola Santoro, Walter Quattrociocchi, Paola Flocchini, Arnaud Casteigts, and Frederic Amblard. Time-varying graphs and social network analysis: Temporal indicators and metrics. *arXiv preprint arXiv:1102.0629*, 2011.