Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://jaida.ikcu.edu.tr

Vol. 1, No. 2, pp. 150-159, 2021
Research Article

Matching Potential Customers and Influencers for Social Media Marketing

Fatih SOYGAZİ [1,*], Muhammet Enes AYDOĞAN [1], Hilmi Can TAŞKIRAN [1], Özgür KAYA [2]

[1] Aydın Adnan Menderes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Türkiye
[2] Fenomio Influencer Marketing, Türkiye

Abstract
Social media platforms are so important for the advertising industry. Companies have a huge amount of budget for advertisement and try to select an influencer as the face of their brand for these advertisements. Each brand is related to a specific segment of customers. When the true influencer is followed by this segment, advertising companies contact him/her. The objective of this work is to facilitate the job of the advertising company by matching the brand and the influencer to use the budget of the advertising company appropriately. Accordingly, our work makes an analysis of real/fake account detection, gender, and age range prediction of the influencer's followers. In this work, it is focused on the real accounts by eliminating the fake ones and the gender, age-range prediction of these real accounts is considered. The detection of fake accounts is transformed into a binary classification problem by observing the features of real and fake accounts. Another binary classification solution is presented for gender detection by checking the pictures of the account owners and their names together. A pre-trained deep learning model for follower age range prediction is provided based on the pictures of these followers. The accuracy of the predictions is evaluated for each of the three situations and the success of our approach is observed for influencer/follower matching.
Keywords: Deep learning; fake account detection; image processing; machine learning; social media analysis.

1. Introduction

Marketing and sales via social media have been so popular recently. It stands out that social media seriously affects users' product habits considering how advertisements on social media sites change the market shares in e-commerce [1], [2], [3]. Companies are now focusing on social media marketing rather than traditional marketing methods in order to increase their product sales [4], [5], [6]. This method, called Influencer Marketing, is the process of making product promotions and the content marketing strategy by establishing relations with the relevant influencer in order to deliver products or services to more users. It is of great importance to choose the correct advertising face, as determining a customer profile on social media and trying to access customers in this profile with a suitable influencer will increase the sales performance. At this point, the job of advertising companies is to find the most appropriate influencer with the help of applications that analyze social media.

Our work, which aims to make influencer/potential customer matching, is an application that analyzes Instagram followers of the potential influencers. In this study, it is determined how many of the followers of an influencer involve fake accounts firstly. After filtering the fake accounts of the influencer, the gender and the demographic data of the followers are considered. Influencer/potential customer matching is done by using image processing, text processing, and classification algorithms. If a corporation requires advertisement via an influencer on Instagram, it can be predicted whether the sales of its products will be parallel with their targets by using our solution.

The contribution of our work is:

- To match the brands and influencers after filtering the fake accounts and obtaining a demographic view of the influencer's followers
- To propose some metrics for understanding the matching prediction of the influencer and the brand more effectively

2. Related Work

The impact of social media has also caused the increase of the influencers publishing diverse area of content to expand their followers. At that point, e-commerce companies are struggling with finding the proper influencer for their brand and it is the biggest challenge for most of the brands from the view of advertisement [7]. The brands and the especially influencers' accounts must be represented to make this matching with a mathematical model. The social account's history of an influencer must be fetched with his/her information and stored in a social information pool to obtain a general knowledge about that influencer to match with brands. Gan et. al. [8] proposed a multi-modal social account representation involving the history pooling of the account and the social

*Corresponding author e-mail address: fatih.soygazi@adu.edu.tr

Received: 02/12/2021; Accepted: 27/12/2021

account embedding to represent the influencer. Then, they proposed a ranking strategy inspired from engagement metric named as competence score to understand the matching rate between influencers and brands. Wang et. al. [9] proposed two adaptive learned metrics, endorsement effect score and micro-influencer influence score, following the approach in [8]. [9] considered more interpretable concept-based parameters for marketing decisions and applied their experiments on a real-world dataset.

The detection of fake accounts called bots that do not express real identity is important in terms of understanding real behavior in social media. Akyon and Kalfaoğlu [10] created two datasets consisting of real and fake accounts. They detected fake accounts at 86% and 96% accuracy by the traditional machine learning methods they tried on these datasets. They obtained trained models based on different features such as the number of people followed by the user, the number of people following the user, the length of the username, the number of numerical values in the username to detect real and fake accounts.

Jeon et al. [11] aimed to determine the gender of users with deep learning methods in order to be used in the field of advertising. Gender detection was made by analyzing photos and daily activities over 33752 Instagram posts. It was stated that an F1-score of 76% was achieved over the features extracted from the posts.

Han et al. [12] developed an approach that predicts the age range from the posts, based on the behavior of young people and adults on social media. The online behavior information of the users is stored in two datasets obtained by user-profiles and the tags of these profiles. It was stated that estimation was made at an accuracy level of 82% by the training models with these datasets.

Companies interested in influencer marketing benefit from various applications [1,2]. These applications collect various information about profiles by making analyzes with artificial intelligence-based approaches and offer the most suitable influencer alternatives for the demanding companies.

Farseev et al. [13] developed an artificial intelligence-based application about influencer discovery, which they named SoMin. The study focuses on analyzing the user profile in the relevant company's market and finding the near- or long-term influencer matches suitable for this profile. They produced a solution called "Individual User Profiling" in order to find out what the attributes of the people who will follow the product will be. They observed all the multimedia shares made on social media regarding the market where the product is located when offering their solution.

The quality of matching potential customers of a product and the influencer must be evaluated for having an intuition about the future income of the product's corporation. Hence, there should be some metrics to handle the quality of the matching process. Engagement rate [14], [15], [16], is a metric that is calculated as the number of users that have interacted with a post (whether they liked, shared, commented, or clicked on the photo or link) divided by the number of the influencer's followers. Sponsored pictures have an impact on the purchase of the customers. Naumanen and Pelkonen [15] propose that the purchase behavior of the followers is affected positively by the sponsored product pictures. They call the influencers micro-influencers having a small number of followers but these influencers have more impact on specific product categories. The engagement rate must be considered to find out the micro-influencers on Instagram.

3. Methodology

Our methodology includes two phases. In the first phase, we detect the fake accounts of the influencer's followers as some of the influencers pay for fake accounts and increase the follower count to be more popular. We must initially ensure that the influencer is followed by real accounts that may be analyzed. After eliminating the fake accounts, we have two prediction operations for classifying the portfolio related to the influencer: gender and age range. These are the two basic properties of the followers to predict their daily expectations about various products. We measure the quality of these classifications via some metrics explained below.

3.1. Fake account prediction

In the detection of fake accounts, a dataset containing the information of real and fake users is created and model training is carried out with this data set. Our dataset contains the following information:

- Follower: The number of followers of the Instagram user.
- Following: The number of accounts followed by the Instagram user.
- Biography Length: The number of characters of the information written in the biography section of the Instagram user's profile.

[1] https://web.archive.org/web/20211223132633/https://hypeauditor.com/
[2] https://web.archive.org/web/20211223133347/https://analisa.io/

- Total Post: The total number of posts made by the Instagram user.
- Has Profile Picture: The information of whether the Instagram user has a profile photo or not. If there is a profile photo, it says "1", if not, it says "0".
- Has External Url: The information about whether the Instagram user has added url information to his profile. If there is url information in their profile, it says "1", if not, "0" is written.
- Full Name Length: The character length of the Instagram user's name and surname.
- Digit Count: The information of how many numbers the Instagram user has in his username.
- Username Length: The information of how many characters the username of the Instagram user consists of.
- FFR: The most important attribute we use for fake account prediction is the ratio of the number of people following to the number of people following (FFR). If FFR is so high, it is an indicator of a potential fake account. It means that a few users or nearly nobody follow this user whereas that account is following lots of accounts.

FFR = Number of people followed / Number of people following

The model is then trained using a data set that includes both real and false account information. The list of Instagram influencers' followers is then gathered, and a data set comprising the information about their followers is constructed. The accuracy of the obtained users is then assessed using the trained model's prediction. As a result, the accounts that we've identified as fake are found, and the phenomenon's profile performance is assessed. As a result of binary classification with Random Forest Classifier during model training, fake and real accounts are separated. An example of a dataset containing real and fake user information that we have prepared is shown below.

Follower	Following	Biography Length	Total Post	Has Profile Pic	Has Ext Url	Full Name Length	Digit Count	User Name Length	FFR	IsFake
145	142	17	0	1	0	17	0	10	0.97	0
288	279	16	0	1	0	16	0	10	0.96	0
420	435	17	0	1	0	17	0	10	1.03	0
183	201	0	0	1	0	0	0	10	1.09	0
208	252	13	0	1	0	13	1	14	1.21	0
0	115	0	6	1	0	0	4	16	115.0	1
0	120	0	6	1	0	0	4	23	120.0	1

Figure 1. *A fragment of the dataset we construct containing real and fake account information*.

3.2. Gender detection and age-range prediction

In our study, two different methods are used for gender prediction to validate the predicted gender information of each account more accurately. Firstly, we predict the gender of the users with image processing. DeepFace[3] library is used as an image processing algorithm in this paper. It is a hybrid face recognition framework wrapping state-of-the-art models: VGG-Face, Google FaceNet, OpenFace, Facebook DeepFace, DeepID, ArcFace, and Dlib. Google FaceNet, VGG-Face, ArcFace, and Dlib are overperforming ones based on experiments. Scores of these models are available in both the Labeled Faces in the Wild and YouTube Faces in the Wild datasets as announced by their creators. The default configuration of DeepFace uses the VGG-Face model [17] and we used DeepFace with that configuration. It is stated that this hybrid model overcomes the accuracy of the human recognition of a face with 97.53% accuracy [18]. The gender and age of the account owners are determined through the Instagram profile picture with DeepFace library. While this library works slower and some of the names give nearly exact information about the gender of a user, a name-based predictor is also used for prediction

[3] https://web.archive.org/web/20211223131917/https://github.com/serengil/deepface

to fasten and verify the overall process. Python's gender_guesser[4] module is our auxiliary solution to predict a gender by name. Gender analysis is performed on the name given by the user in the Instagram profile with the gender_guesser module.

After these analyzes are done, the results of two different methods are compared. When the same outputs are obtained, these results are recorded in the database about an account. In this way, our algorithms work faster by not making repeated analyses for the same user. As image processing via DeepFace takes a huge amount of time for analysis of a set of users, we prevent applying the procedure for the same users repeatedly by storing their information. Therefore, each user may be fetched for different matching operations by skipping these phases and accessing their information from the database.

4. Proposed Metrics

The engagement rate [6] is the most commonly used metric to evaluate the matching quality of the influencers with the potential customers. The metrics shown below are novel that we propose to check the success of matching operations in social media. These are;

Alternative engagement rate: It calculates the influencer's interaction rate on Instagram.

*Alternative engagement rate = (total likes count + total comments count) / (total followers count * total post count)* (1)

Average like rate: It calculates the average number of likes given to the posts of the influencer.

Average like rate = total likes count / total post count (2)

Average comment rate: It calculates the average number of comments written under the posts of the influencer.

Average comment rate = total comments count / total post count (3)

Like rate: It calculates the percentage of the average like rate with respect to the total number of followers.

Like rate = average like rate / total followers' count (4)

Comment rate: It calculates the percentage of the average comment rate with respect to the total number of followers.

Comment rate = average comment rate / total followers' count (5)

Ghost follower rate: There is another concern that needs to be dealt with besides detecting fake accounts. Some of the accounts might be stolen and managed like it is the real owner of the account. In another case, some commercial accounts are compulsorily followed to have a chance of winning an award from a website or that account. In these cases, the follower is not actually interested in the account it follows currently. We propose a metric called ghost follower rate to understand the number of ghost followers of an account. Initially, we gather the list of accounts that like the posts of the influencer and the list of the followers. The accounts which are in the follower list but not in the liked post list are called ghost followers.

*Ghost follower rate = (total users who like posts - total followers count) / (total followers count * total post count)* (6)

4. Results

For the fake account analysis method used in this study, it was first thought to determine the number of likes, comments and story views in Instagram photos. However, the likes, comments, and stories may be generated automatically and be purchased. Hence, observing these attributes affects the analysis negatively. Therefore, the information of the influencer's followers is analyzed and the number of fake followers is determined with our proposed method.

The configuration of the computer where tests are performed in Section 4.1 is:

- Operating System: macOS BigSur
- CPU: Apple M1 chip 8-core CPU with 4 perform-ance cores and 4 efficiency cores
- GPU: Apple M1 chip 7-core GPU, 16-core Neural Engine

[4] https://web.archive.org/web/20211223130832/https://test.pypi.org/project/gender-guesser

- RAM: 8GB

The configuration of the computer where tests are performed in Section 4.2 and Section 4.3 is:

- Operating System: Windows 10 Pro 64 bit
- CPU: Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz (8 CPUs)
- GPU: NVIDIA GeForce GTX 1050 Ti
- RAM: 16GB

4.1. Tests about fake account prediction

An account with 298 followers was split into 234 real followers and 64 fake accounts. The training time of the model is 1.5 sec while the testing time is 0.51 for the prediction phase. After performing the test process on this account with our bot analysis algorithms, 230 accounts are detected (98.3% accuracy) as real and 61 accounts (89.1% accuracy) are detected as fake.

4.2. Tests about gender detection

The number of predicted genders in the dataset involving the followers of an account is not so high compared to the promise of DeepFace as seen in Table 1. DeepFace actually uses the most prominent algorithms and works hybrid for detection. The problem is not related to the algorithms indeed. The deep learning algorithms trained with the given datasets are mostly constituted with preprocessed images whereas the profile pictures of the accounts on Instagram are more complex for prediction. The profile pictures might not involve the face of the account owner or the face of the account owner might be captured from an angle to be predicted accurately. Hence, it is not simply enough to capture the gender profile of the followers of an influencer. Average processing time involves the time of fetching the personal information of the follower and the prediction time of the follower's gender together. The cause of the increase for operation time is related to accessing and fetching the data from Instagram. When it is a time-consuming scheme, we collect each user's information and store it not to process the same user repeatedly.

Table 1. *The results of DeepFace in our dataset.*

# of followers	Avg. processing time including Instagram access (sec)	Avg. detection time
100	141	67
250	316	169
500	736	378

Table 2 indicates the number of predicted genders of the followers with the help of their given names on their accounts. The composing solution approach from the outputs in Table 1 and Table 2 presents us better results to have an intuition about the portfolio of an influencer. The problem in this processing phase is the same in Table 1. It takes a long time to access and gather data from Instagram compared to processing of the names.

Table 2. *The results of the gender_guesser module in our dataset.*

# of followers	# of detected genders of followers	Avg. process time (sec)
100	67	0.4
250	184	0.43
500	357	0.47

The gender detection results via DeepFace and gender_guesser are seen in Figure 2 and Figure 3 in detail.

4.3. Tests about age range detection

Age detection is another important issue for matching the influencers and the potential customers. Because each person within an age range has its own hobbies or interesting products. The advertising companies have a list of profiles to match the products and the potential customers. Age range is an important attribute for these profiles as it is clearly seen. Figure 4 presents the detected age ranges of the users on our dataset collected from Instagram.

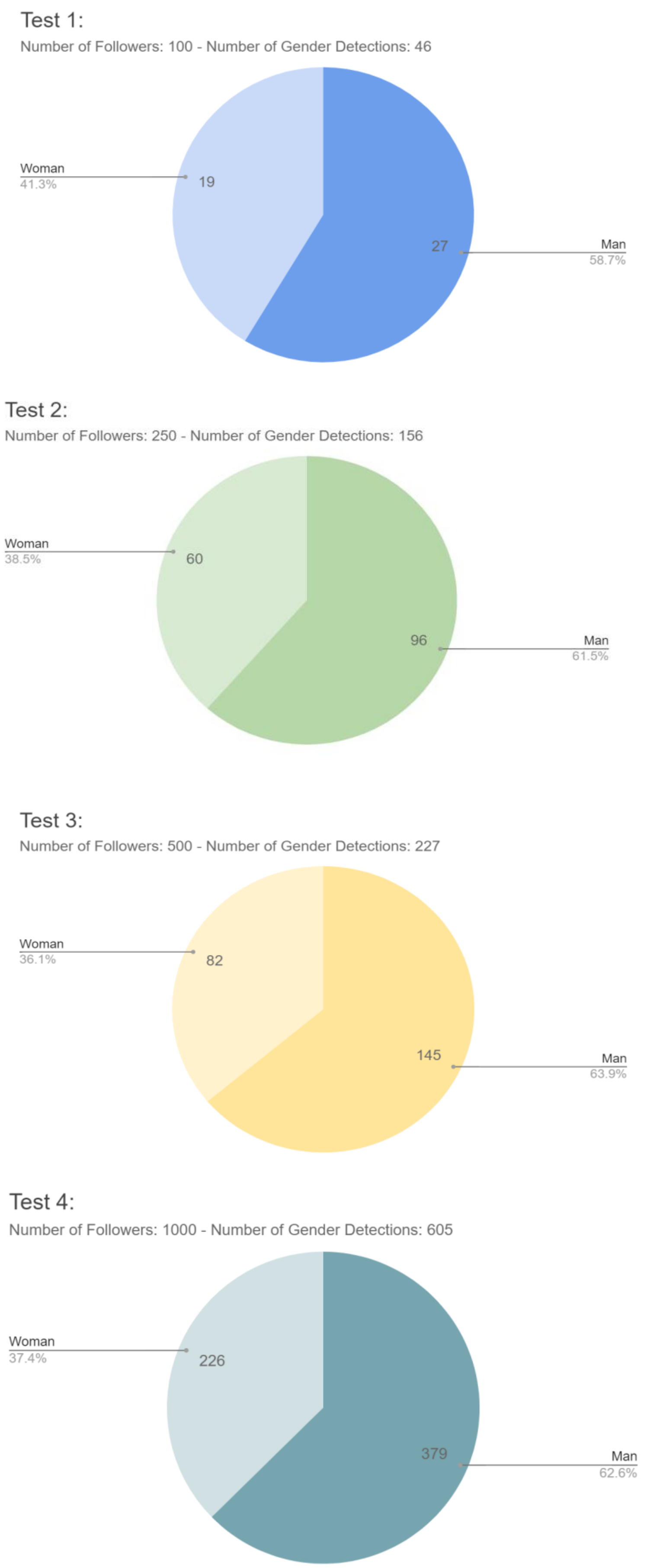

Figure 2. *Tests about gender detection via DeepFace*.

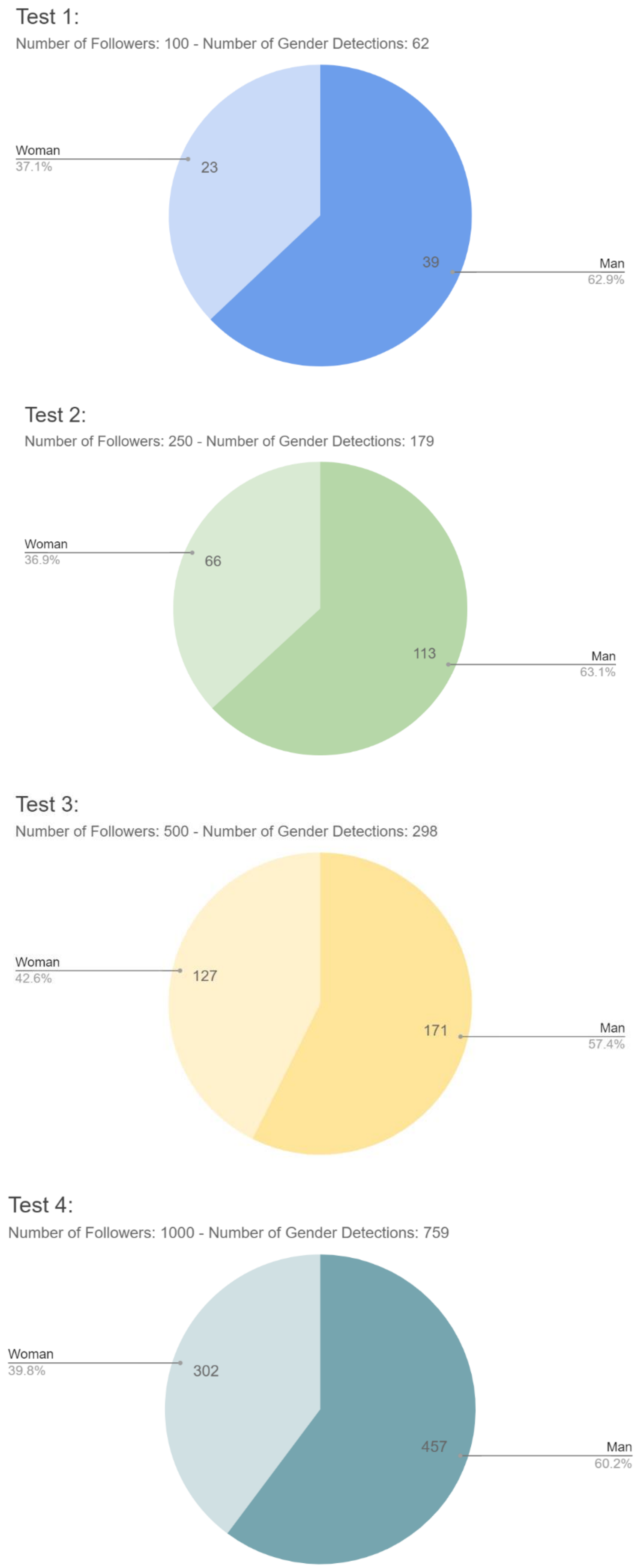

Figure 3. *Tests about gender detection via gender_guesser*.

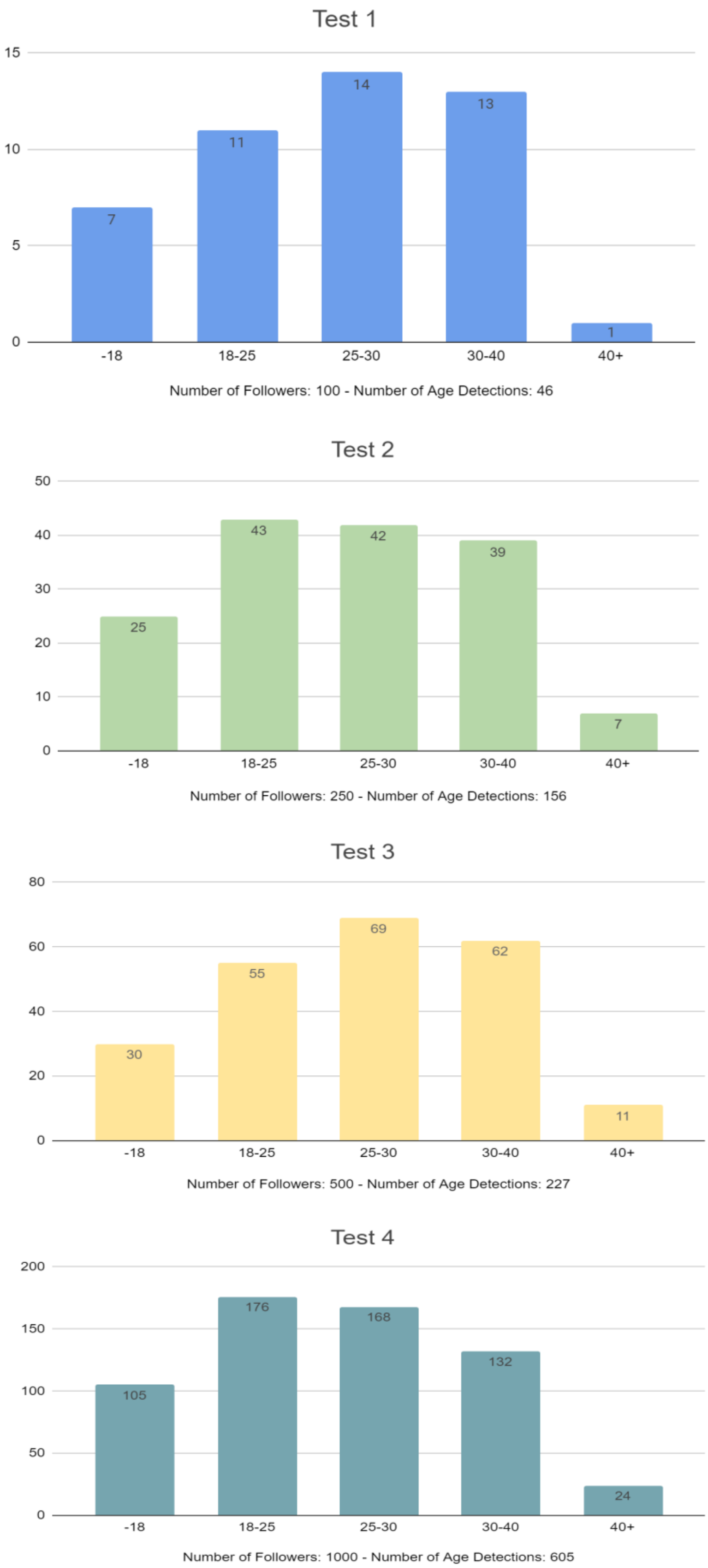

Figure 4. *Tests about age detection* .

5. Conclusion and Future Work

It is known that many influencers have fake followers on their profiles to provide an attraction for potential followers. At the same time, most companies advertise on social media through influencers. Our paper focuses on helping the companies to cooperate with the right influencer and receive the return of their advertisements. Hence, our studies have been carried out to determine whether the followers of an influencer on Instagram are fake. We also find the gender and age range of the influencers to obtain an intuition about the potential market portfolio of that influencer.

Existing solutions in the literature have not paid much attention to fake account detection. Our difference from others is to provide fake account detection and to recommend influencers with real followers to companies that will advertise. In addition to giving the gender and age determination values as percentages, our study also aims to give these values numerically that means we may supply more specific information if needed.

The main problem of our work is to run our application at a less acceptable speed for real-time usage. Our future goal is to run the deep learning solutions on multi-core architectures (GPUs, TPUs) to gain acceptable resulting times. We will also make evaluations with the proposed metrics in Section 4 with the extended dataset to gain more information about the follower of the influencer to offer a better portfolio to the company that requires advertisement for their products.

Declaration of Interest

As authors, we declare that we have no conflict of interest with anyone related to our work.

Acknowledgements

An earlier version of this paper was presented at the ICADA 2021 Conference and was published in its Proceedings (Title of the conference paper: “Sosyal Medyada Fenomenler ve Müşterileri Eşleyen Bir Çözüm Önerisi”).

References

[1] M. Pütter, “The impact of social media on consumer buying intention”, Journal of International Business Research and Marketing, 2017, 3(1), pp. 7-13.

[2] C. Schwemmer and S. Ziewiecki, “Social media sellout: The increasing role of product promotion on YouTube”, Social Media + Society, 2018, 4(3), pp. 1-20.

[3] M. Delbaere, B. Michael and B. J. Phillips, “Social media influencers: A route to brand engagement for their followers”, Psychology and Marketing, 2021, 38(3), pp. 101-112.

[4] M. Bruhn, V. Schoenmueller and D. B. Schäfer, “Are social media replacing traditional media in terms of brand equity creation?”, 2012, Management Research Review, 35(9), pp. 770-790.

[5] E. Constantinides, “Foundations of social media marketing”, Procedia-Social and Behavioral Sciences, 2014, 148, pp. 40-57.

[6] X. J. Lim, A. M. Radzol, J. Cheah and M. W. Wong, “The impact of social media influencers on purchase intention and the mediation effect of customer attitude”, Asian Journal of Business Research, 2017, 7(2), pp. 19-36.

[7] S. Woods, “#Sponsored: The emergence of influencer marketing”, 2016, https://trace.tennessee.edu/utk_chanhonoproj/1976.

[8] T. Gan, S. Wang, M. Liu, X. Song, Y. Yao, and Liqiang Nie, “Seeking Micro-influencers for Brand Promotion”, 2019, In Proceedings of the 27th ACM International Conference on Multimedia (MM '19). Association for Computing Machinery, New York, NY, USA, 1933–1941. DOI:https://doi.org/10.1145/3343031.3351080

[9] S. Wang, T. Gan, Y. Liu, L. Zhang, J. Wu and L. Nie, "Discover Micro-influencers for Brands via Better Understanding," in IEEE Transactions on Multimedia, 2021, doi: 10.1109/TMM.2021.3087038.

[10] F. C. Akyon and E. Kalfaoglu, “Instagram fake and automated account detection” In Proc. IEEE Innovations in Intelligent Systems and Applications Conference, 2019, pp. 1-7.

[11] Y. Jeon, S.G. Jean and K. Han, “Better targeting of consumers: Modeling multifactorial gender and biological sex from Instagram posts”, Journal of User Modeling and User-Adapted Interaction, 2020, vol. 30, pp. 833-866.

[12] K. Han, S. Lee, J. Y. Jang, Y. Jung, and D. Lee, “Teens are from mars, adults are from venus: analyzing and predicting age groups with behavioral characteristics in instagram.”, In Proceedings of the 8th ACM Conference on Web Science (WebSci '16), Association for Computing Machinery, 2016, pp. 35-44.

[13] A. Farseev, K. Lepikhin, H. Schwartz and E. K. Ang., “SoMin.ai: social multimedia influencer discovery marketplace” In Proc. of the 26th ACM International Conference on Multimedia, pp. 1234-1236, 2018.

[14] T. Niciporuc, "Comparative analysis of the engagement rate on Facebook and Google Plus social networks," Proceedings of International Academic Conferences 0902287, International Institute of Social and Economic Sciences, 2014.

[15] E. Naumanen and M. Pelkonen, “Celebrities of Instagram - What Type of Content Influences Followers’ Purchase Intentions and Engagement Rate?”, Master’s Thesis, Aalto University. School of Business, 2017.

[16] R. L. H. Yew, S. B. Suhaidi, P. Seewoochurn and V. K. Sevamalai, "Social Network Influencers' Engagement Rate Algorithm Using Instagram Data," 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), pp. 1-8, doi: 10.1109/ICACCAF.2018.8776755, 2018.

[17] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition", BMVC 2015, 2015, pp. 41.1–41.12, 2015.

[18] S. I. Serengil and A. Ozpinar, "LightFace: A Hybrid Deep Face Recognition Framework, " 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1-5, doi: 10.1109/ASYU50717.2020.9259802, 2020.