Intelligent Identification of Fake Accounts on Social Media

Anastasia Stolbova, Rustam Ganeev
Samara National Research University
Samara, Russia
anastasiya.stolbova@bk.ru

Anton Ivaschenko
Samara State Technical University
Samara, Russia
anton.ivashenko@gmail.com

***Abstract*—The paper presents an original research of fake accounts on social media using an artificial neural network for their identification. Specifically designed and implemented application was used to identify specific features of fake accounts and study the principles and reasons of their generation. Considering the possible use cases and scenarios of its practical use including everyday social media surfing there was examined a possibility to implement a mobile application based on Java in the Android Studio programming environment. Based on the study of 500 real and 500 fake VKontakte accounts there was determined a number of conclusions on the original features of fake accounts. The provided research allowed extending the list of criteria for identifying fake accounts by an original set of patterns. The developed conclusions allow formulating the statements on what criteria can be used for further identification of fake accounts in practical applications.**

I. Introduction

Nowadays social media are recognized as a challenging source of knowledge about social and economic trends in different areas of human activity. In particular, social media is used for decision-making support in marketing, targeted advertising, sociology and politics to determine the feelings and preferences of people and communities. Large volumes of available data allow providing comparatively high accuracy and reliability of established patterns, which makes it reasonable to use this information resource in practice.

Nevertheless, it is pretty understood in professional communities that social media are highly influenced by the human factor. Humans are represented by avatars or digital twins that trend to capture the idealized properties of human nature. This factor can be determined and reduced by deep analysis of social media profile. Frequent and constant online activity leaves no time for hiding and manipulating. As the result, frank and open behavior reflects in authentic profiles.

Fake accounts are a different story. They are initially generated according to manipulative strategy and do not contain any data that can be interesting for fact-based analysis. Fake profiles can look quite realistic and simulate real behavior on both semantic and statistic levels of interception. Therefore, it is highly essentially important to identify them and consider as a separate object for study. In this paper there is presented a new approach of fake accounts identification using an artificial neural network.

II. State of the Art

By a fake account, we mean an imitation of someone's genuine account or an account of a non-existent person or organization for the purpose of deceiving.

In general, fake accounts identification is related to the classification task. In the field of social media analysis it can be solved using several possible approaches. E.g. to classify fake reviews, news, profiles, machine learning methods are successfully used including decision trees [1 – 3], logistic regression [4], [5], neural networks [6], [7], nearest neighbor method [8], support vector machine [9], [10]. The paper [11] explores the methods for detecting anomalous actions and fake Twitter accounts based on a combination of support vector machines and neural networks.

The considered research works propose to identify already existing fake accounts. Another approach is to proactively identify such accounts at the stage of user registration, e.g. based on the analysis of the questionnaire data [12]. The paper [13] is devoted to the early detection of fake accounts based on the analysis of friend requests and responses.

Classification is traditionally performed using the initial data that describes behavioral aspects (frequency of page visits, rate of commenting or updating of profile photos, etc.) and static parameters (number of friends, completeness of filling out a profile, etc.). However, the authors of [14] suggest analyzing user posts and applying preprocessing of the obtained data to highlight keywords, spam words and thematic modeling for subsequent classification. In [15], an original approach is proposed based on calculating the measure of similarity between users of social networks.

Intelligent technologies are frequently used to analyze social media nowadays. In [16], the authors introduce an artificial neural network to train the fake detection classifier on Twitter, while investigating the optimal combinations of activation functions on different layers of the network, which led to high accuracy rates of the classifier.

Within the framework of this research, it is proposed to consider an approach to identify fake accounts using an artificial neural network with a multilayer perceptron topology, as the most commonly used for classification trained on a dataset of social media. This study extends the concept of a combination of statistical and semantic analysis of social media [17 – 19].

The main goal of this research is to determine the main features of fake accounts that can be later used to search and identify them in social media analysis.

III. Methods

Identification of fake accounts is a challenging problem. By definition they are created to simulate the behavior of real accounts in the most realistic way. Therefore, they reproduce not only the content but also the dynamical characteristics.

Due to this feature the problem is being hardly solved not only by ordinary users but also by experts that analyze the social communities. In addition to this, the number of fake accounts can be big and they can even form an unknown proportion of active users in a community.

Consequently, fake accounts identification turns out to be a perspective sphere of application of intelligent technologies. Intelligent identification can be provided on the basis of formal rules or neural networks. The first option allows introducing the knowledge base or fuzzy logic, but requires high expertize to design and develop. Neural networks are better to operate with high uncertainty, but cannot explain the features and reasons of fake accounts appearance, which is critical for decision-making support.

The solution is a combined approach. Neural network is used for an advanced analysis of the raw data imported from social media, which allows formulating the rules for further identification of fake accounts in practical applications. In this paper we propose and demonstrate the possibility of this approach. Expected results include not only fake accounts identification, but also formalization of their features that helps understanding the causes and influencing factors.

For example, in studying the subjectivity of fake accounts it is important to consider their origin. The most interesting are the following situations:

- a human-created digital twin hides the tendency to deviant behavior of the individual;
- a human-created digital twin indicates the presence of personality deviance, but this is an intentional representation;
- a human-created digital twin is normal and does not indicate any abnormalities, however it is a fake.

Understanding the origin allows to make a decision on how to consider the fake accounts in the analyzed dataset. The data model is shown in Fig. 1.

So in this study there was created an initial dataset and implemented an artificial neural network. As the initial data we manually created a dataset consisting of 1000 real accounts of users of the social network "VKontakte", which includes 500 fake accounts and 500 real ones. No specific demographic parameters were considered and no special requirements were used to pick out the accounts.

Initial dataset was marked up manually, using the preliminary defined markers. The following account data were selected as the criteria used to determine whether an account is fake:

- page status (deleted, blocked);
- availability of verification;
- registration date;
- the date of the last visit to the page;
- date of the last page change;
- number of outgoing subscriptions;
- number of subscribers;
- number of friends;
- the number of deleted, blocked friends;
- number of groups;
- number of posts on the wall;
- download dates of the first five records;
- download dates of the last five records;
- number of photos;
- upload dates for the first five photos;
- upload dates of the last five photos;
- the number of duplicates of the main photo.

The training and validation datasets were created by dividing the generated dataset in the ratio of 80% by 20%.

The first stage of fake account identification is collecting data about accounts on social media, which is carried out according to the following algorithm:

Step 1. Entering a link to the considered account;

Step 2. Generation of requests for social media API to obtain basic information about the account;

Step 3. Generation of requests for foaf.php to obtain additional information about the account, for example, registration date, last login and last page change;

Step 4. Generation of queries to determine the originality of the main account photo using the search API;

Step 5. Calling generated requests;

Step 6. Parsing the received responses;

Step 7. Filling in the account information.

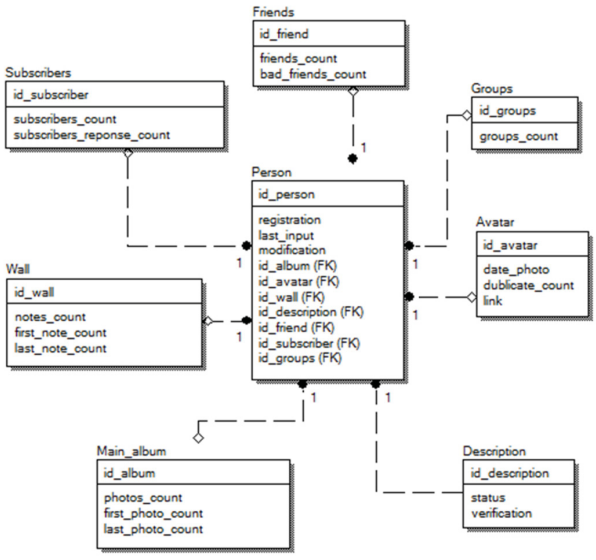

Fig. 1. Data model

To classify the accounts there was used an artificial neural network with a multilayer perceptron topology. The classifier serves as a tool to determine if an account is fake, based on the analysis of the values of the selected criteria. The optimal architecture of a neural network and a combination of activation functions that have been found empirically:

- input layer, consisting of 13 elements;
- 4 hidden layers with dimensions of 64, 32, 16 and 8 neurons;
- output layer with dimension 1;
- hyperbolic tangent to activate hidden layers:

$$\tanh(x) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}, \quad (1)$$

where $z = \sum_i w_i x_i + b$, w_i - weight of the i-th input;

x_i - the input value of the i-th input; b - bias.

- logistic sigmoid to activate the output layer:

$$\tanh(z) = \frac{1}{1 + e^{-z}}. \quad (2)$$

The graph of the error depending on the number of epochs is shown in Fig. 2, and the quality of training is proposed to be measured by the AUC indicator shown in Fig. 3. The error value decreases, and the AUC value increases. The AUC for the validation sample averaged 0.95, and the mean error was 0.22.

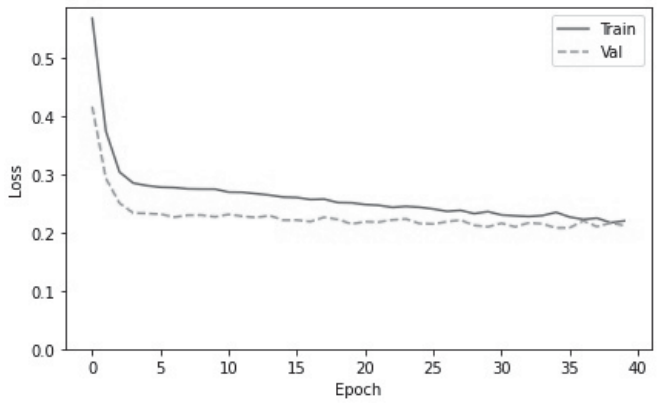

Fig. 2. Error graph

IV. Results

To provide analysis and identification of fake accounts there was developed a software application. Considering the possible use cases and scenarios of its practical use including everyday social media surfing there was examined a possibility to implement a mobile application based on Java in the Android Studio programming environment.

The application uses machine learning platform TensorFlow and APIs such as Keras, “VKontakte API”, “API Search by Yandex.Images”, operating under Android 5.0.

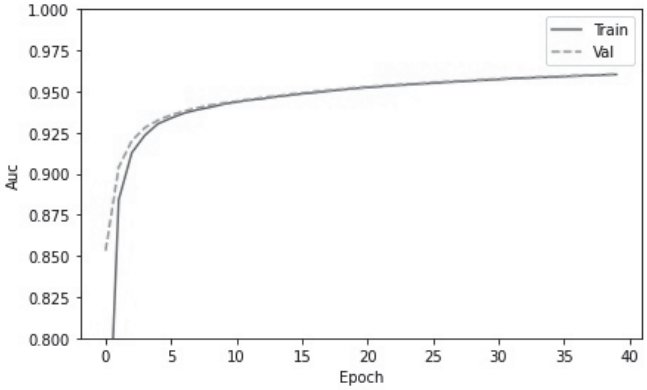

Fig. 3. AUC graph

The test sample includes 100 users. The neural network correctly recognized 94 % of users (2 % of fakes were recognized as real users, and 4 % of real users were falsely recognized as fakes) on the test sample. Table 1 shows the error matrix. The AUC indicator, estimated on the data of the test sample, is 0.96, which indicates the high quality of the account classification.

Identification results are characterized by the Recall of 0.96, Precision of 0.92 and Accuracy of 0.94.

TABLE I. Error matrix

		Actual	
		1	0
Predicted	1	True Positive (TP) 48 %	False Positive (FP) 4 %
	0	False Negative (FN) 2 %	True Negative (TN) 46 %

An example of fake account on the VKontakte social network is presented in Fig. 4. For this account, it was noted that there were a large number of friends and subscribers, there were no photos, advertising posts were posted on the wall and a large number of posts were posted in one day. Therefore, at first glance it may seem to be a real account.

Checking the account in the developed automated system highlighted a list of indicators with values that exceed those for an average account:

- number of friends – 9305;
- number of subscribers – 603;
- number of subscriptions in response – 695;
- the number of “bad” friends –289;
- number of groups – 339;
- number of photos – 1;
- the number of records on the wall is 7402.

The results of the statistical analysis are shown in Fig. 5.

As the result, the classification showed that the considered account is 94% fake.

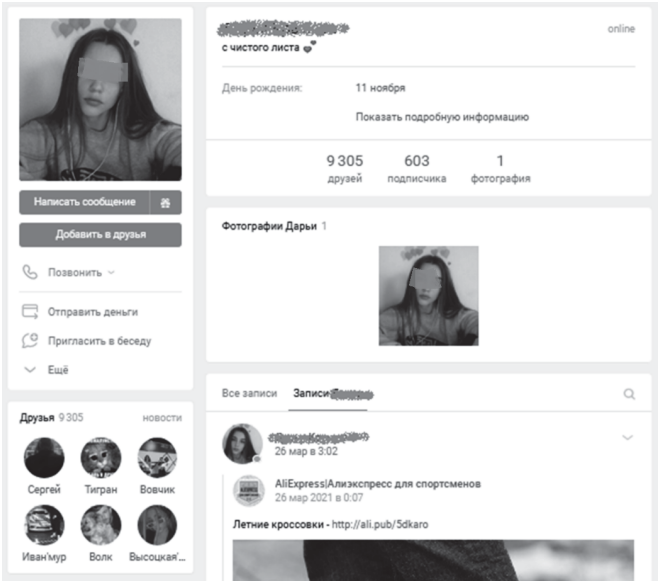

Fig. 4. Fake account example

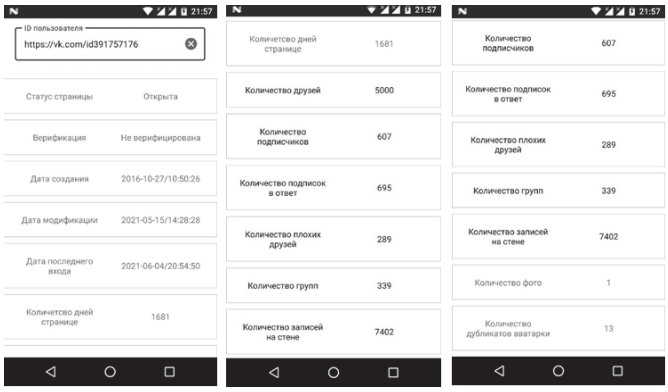

Fig. 5. Results of statistical analysis of the account

The proposed approach was implemented in intelligent software for identification of deviant behavior. This system provides decision-making support on the basis of a knowledge base implemented as an automated Ontology containing the corpus of rules that characterize the degrees of deviance on the basis of finding certain verbal expressions and symbols in posts and comments. These are, for example, thematic and associative words, musical directions, emoticons, etc.

The initial set of rules was created and compiled by expert psychologists. The logic of fake accounts identification allows extending this corpus of rules and thus improving the key algorithms.

Based on the results of the analysis of the activity of a user of social networks, a model was built describing a virtual personality that imitates the social, behavioral, emotional and cognitive qualities of a person. Evaluation of the emotional state of the person is carried out based on the intelligent analysis of the sentiment of posts, and then the emotional state of the person is determined, taking into account the frequency of publications.

One of the system visual components that illustrate the proposed approach with depersonificated data is presented in Fig. 6. The main feature of the proposed approach is an ability to move forward from the semantic analysis of messages by adding the analysis of emotional context and then consider the dynamics of its changes in time. In these conditions identification and reducing the fake accounts is a necessary step to provide the required adequacy and reliability of decisions.

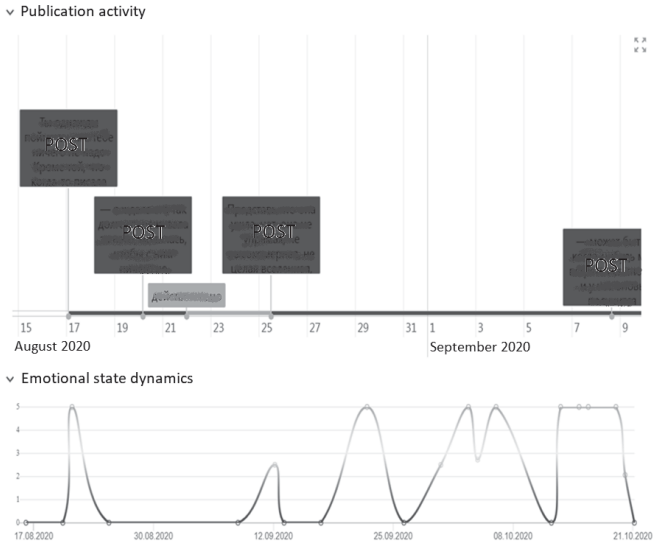

Fig. 6. Social media activity analysis

The proposed approach was probated for social media analysis in social science, marketing, advertising and public opinion research [19]. Positive results were achieved in banking acquiring, where the proposed approach was used to analyze the perspectives of new product placement. Analysis of social media gave new opportunities being practically used in the regional Ministry of social, demographic and family policy.

V. Discussion

Based on the study of 500 real and 500 fake VKontakte accounts there was determined a number of conclusions on the original features of fake accounts. These conclusions allow formulating the statements on what criteria can be used for further identification of fake accounts in practical applications. Details are given in this section below.

Originality of the main profile photo is determined by the presence of duplicates on the Internet. It was revealed that 136 (27.20 %) fake accounts use non-original photos as the main photo. The average number of duplicates for each non-original photo was 62.26 (5.20 %) users with non-original photos were found on real accounts, and the number of duplicates of these photos averaged 73.

This suggests that the owners of real accounts are less likely to put an unoriginal image as the main photo, but if they put it, then this image is popular (a frame from a film, a singer, etc.). In turn, the owners of fake accounts use non-original photos much more often, and the lower number of duplicates indicates

the low popularity of these images, which suggests that such a photo is most likely stolen from another account.

Analysis of the number of friends in accounts showed that fake accounts have more friends than real ones (see Fig. 7).

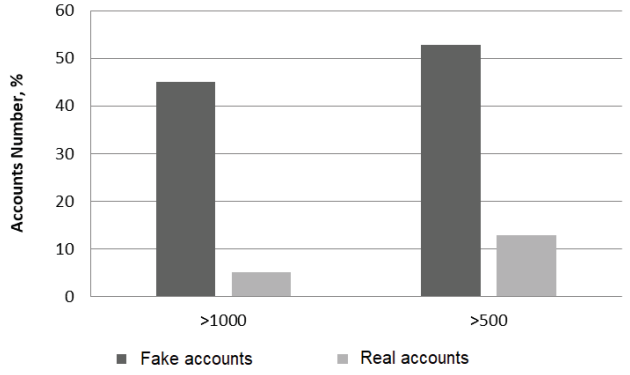

Fig. 7. Number of friends

Analysis of the number of subscribers showed that real accounts have fewer subscribers, and, therefore, they follow the list of subscribers (see Fig. 8).

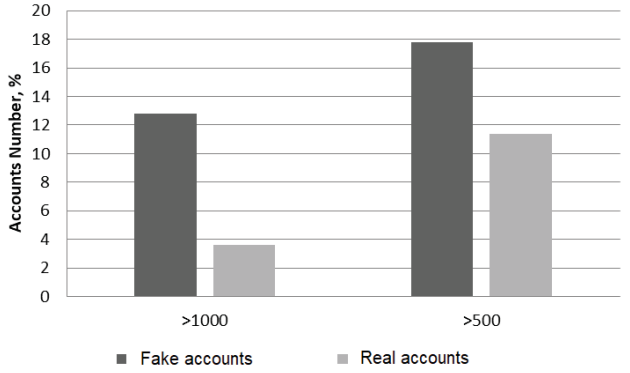

Fig. 8. Number of subscribers

The number of photos posted in the main album at the same time exceeds the mark of 2 photos in 70 (14 %) fake accounts (this means that the user changed the main photo more than twice in one day) and in 40 (8 %) real accounts;

27 (6.14 %) fake accounts out of 440 (other users have restricted access to records on their page) and 22 (7.12 %) real accounts out of 309 lack their own content. The data obtained indicate the fact that real users more often resort to privacy settings.

The presence of advertising records on a page is a separate, large-scale task requiring the implementation of a system for identifying advertising records, and was not considered in this work.

Date of creation of an account on the social network: 79 (15.80 %) fake accounts and 14 (2.80 %) real accounts created their pages within the last month, which indicates the short lifetime of the fake account and the constant need to create new accounts.

The mentioned above criteria were previously used to identify fake accounts. In addition to them there was formulated an original list of additional criteria for identifying fake accounts, introduced below.

The number of “bad” friends described the list of deleted or blocked related accounts. 199 (39.80 %) fake accounts and 30 (6.00 %) real accounts have “bad” friends above 100 users. This is due to the fact that real users follow their friends list and remove inactive users.

The number of outgoing subscriptions shows that real users try not to be in subscriptions with other people, while fake accounts do not monitor this indicator (see Fig. 9).

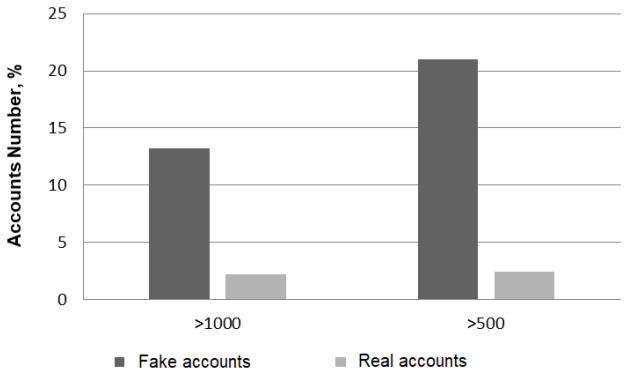

Fig. 9. Number of outgoing subscriptions

Analysis of the number of groups showed that fake accounts are subscribed to more groups than real ones, which indicates the use of fake accounts to fill groups (Fig. 10).

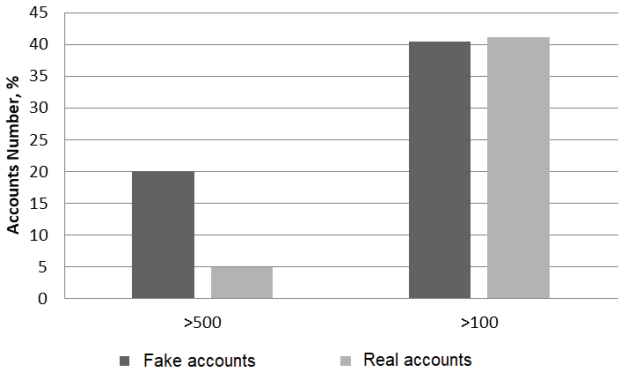

Fig. 10. Number of groups

Analysis of the number of posts on the wall shows that the activity of fake accounts is higher than that of real ones, which is associated with imitation of a real user or the frequent publication of advertising posts (see Fig. 11).

The upload dates of the first five and last five entries are required to get the number of posted posts in one day, since they try to fill the fake account with various contents immediately after creation. The last five records are a trigger for the real account, because situations are not excluded when a newly created, real account is filled with content, and the last five records show whether multiple uploads of records continue in

one day. This verification method improves the system performance, since not all photos are being checked, but only the first five and the last five.

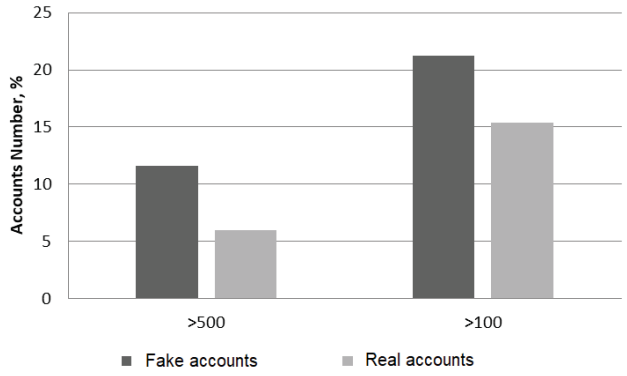

Fig. 11. Number of posts on the wall

The number of photos for 114 (22.80 %) fake accounts and 222 (44.40 %) of real accounts exceeded 20. This result indicates that real accounts are the creators of original content, photos, and fake accounts are often limited to a few photos that could be found on the Internet.

The upload dates of the first five and last five photos are used for the same purpose as the criterion described above for the first five and last entries.

The identified implications can be used as additional classification criteria that correlate with the ones that were originally used to identify fake accounts. Therefore, as being compared to the decision tree method, which provides more precise boundaries for separating real and fake accounts, intelligent classification technology based on neural networks allows generating unexpected criteria.

VI. Conclusion

Thus, within the framework of this work, an approach to identifying fake accounts is proposed. To implement this approach, a set of data was created, consisting of 1000 accounts of the social network "VKontakte", the analysis of which made it possible to identify a list of static and behavioral criteria for an account, which are classification features. The developed classifier based on the multilayer perceptron carries out classification with an accuracy of 0.94.

The provided research allowed extending the list of criteria for identifying fake accounts by an original set of patterns. The developed conclusions allow formulating the statements on what criteria can be used for further identification of fake accounts in practical applications.

Combination of neural networks and rules in the analysis of social media allows identification of fake accounts with a possibility to understand the origin of their appearance, causes and influencing factors. Next research steps are planned to implement the proposed approach in the analysis of social media deviant behavior in various problem domains.

References

[1] K.S. Sanjay and A. Danti, “Detection of fake opinions on online products using Decision Tree and Information Gain”, *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2019, pp. 372-375.

[2] Y. Elyusufi et al, “Social networks fake profiles detection using machine learning algorithms”, *Proceedings of the Third International Conference on Smart City Applications*, Springer, Cham, 2019, pp. 30-40.

[3] S. Hakak et al, “An ensemble machine learning approach through effective feature extraction to classify fake news”, *Future Generation Computer Systems*, Vol. 117, 2021, pp. 47-58.

[4] K.K. Bharti and S. Pandey, “Fake account detection in Twitter using logistic regression with particle swarm optimization”, *Soft Computing*, 2021, pp. 1-13.

[5] M. Granik, V. Mesyura and A. Yarovyi, “Determining fake statements made by public figures by means of artificial intelligence”, *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, IEEE, Vol. 1, 2018, pp. 424-427.

[6] P. Wanda and H.J. Jie, “DeepProfile: finding fake profile in online social network using dynamic CNN”, *Journal of Information Security and Applications*, Vol. 52, 2020, p. 102465.

[7] F.C. Akyon and M.E. Kalfaoglu, “Instagram fake and automated account detection”, *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, IEEE, 2019, pp. 1-7.

[8] P. Kondeti et al, “Fake account detection using machine learning” //Evolutionary computing and mobile sustainable networks. – Springer, Singapore, 2021. – C. 791-802.

[9] P. Karthikeyan et al. “Prevention of shilling attack in recommender systems using discrete wavelet transform and support vector machine” //2016 eighth international conference on Advanced Computing (ICoAC). – IEEE, 2017. – C. 99-104.

[10] K. M. Yazdi et al. “Improving fake news detection using k-means and support vector machine approaches”, *International Journal of Electronics and Communication Engineering*, Vol. 14, No 2, 2020, pp. 38-42.

[11] S. Khaled, N. El-Tazi and H.M.O. Mokhtar, “Detecting fake accounts on social media”, *2018 IEEE international conference on Big Data,* IEEE, 2018, pp. 3672-3681.

[12] D. Yuan, Y. Miao, N.Z. Gong, Z. Yang, Q. Li, D. Song and X. Liang, “Detecting fake accounts in online social networks at the time of registrations”, *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 1423-1438.

[13] A. Breuer, R. Eilat and U. Weinsberg, “Friend or faux: graph-based early detection of fake accounts on social networks”, *Proceedings of The Web Conference* 2020, pp. 1287-1297.

[14] M.M. Swe and N.N. Myo. “Fake accounts detection on twitter using blacklist”, *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, IEEE, 2018, pp. 562-566.

[15] M. Mohammadrezaei, M.E. Shiri and A.M. Rahmani, “Identifying fake accounts on social networks based on graph analysis and classification algorithms”, *Security and Communication Networks*, 2018

[16] M. Simsek, O. Yilmaz, A.H. Kahriman and L. Sabah, “Detecting fake Twitter accounts with using artificial neural networks”, *Artifical intelligence studies*, Vol. 1, No 1, 2018, pp. 26-29.

[17] A. Ivaschenko, A. Stolbova, O. Golovnin, “Data market implementation to match retail customer buying versus social media activity”, *Advances in Intelligent Systems and Computing*, vol 1228, 2020, pp. 363-372

[18] A. Ivaschenko, A. Krivosheev, A. Stolbova, O. Golovnin, “Hybridization of intelligent solutions architecture for text understanding and text generation”, *Applied Sciences*, 11(11):5179, 2021

[19] A. Ivaschenko, A. Krivosheev, A. Stolbova and P. Sitnikov, “Approximate analysis of deviant behavior on social media”, *Lecture Notes in Networks and Systems* 283, 2021, pp. 539-547.