Social Network Users Profiling Using Machine Learning for Information Security Tasks

Elizaveta Dubasova
elsa.dubasova@gmail.com
Artem Berdashkevich
artemberdashkevich@gmail.com
Georgy Kopanitsa
georgy.kopanitsa@gmail.com
Pavel Kashlikov
pavekash@gmail.com
Oleg Metsker
olegmetsker@gmail.com

***Abstract*—The need for bot detection is growing in proportion to the increase in the number of social network users. The robotization of processes has not escaped social networks, with the result that bots, designed to mimic human behavior, create a burden and, in some cases, threats to users, including manipulation and misinformation. Classical information security threats related to bot activity are DDoS, collection and distribution of user data, manipulation of billing systems, and misuse of services. Often bot technology is used for scoring bonus points or using other customer loyalty mechanisms to gain their own benefit, in violation of the service policy. The problem is that it is often hard to confirm the correspondence between a real person and a profile due to the large amount of disparate information about users' activity, as well as the use of modern technologies, including machine learning, to develop bots. This paper focuses on the problem of detecting bots in social networks using machine learning. We propose an automatic, retrainable method for detecting fake accounts on a social network. The study describes the result of developing user classification models based on the activity logs of social network users in the problem of automated user profiling, that is, determining whether a user account is genuine or a bot is hiding behind it. The aim of the work is to develop methods for detecting bots using machine learning and intelligent analysis. In our work to solve the problem we use gradient boosting with an accuracy of AUC = 0.9999.**

I. Introduction

Everyone is deeply involved in enormous networks of social relationships. In order to be able to interact in them regardless of circumstances and time, social networks are actively used. Every year the activity of using social networks increases. According to social media usage statistics, there were 106 million social media users in the Russian Federation in January 2022, equivalent to 72.7% of the total population (importantly, these may be non-unique users), an increase of 7.1% over last year [1].

Any information posted on social media can be found and used by anyone, and not always with good intentions. Therefore, most of the problems in the use of social networks are related to the data that is posted on it. These problems can pose a threat to information security. Every year their number increases, and has increased so much that it has actually created a problem that needs to be solved. For example, Twitter blocked more than 70 million fake accounts in May and June 2018 (The Washington Post). According to Barracuda technology for the first six months of 2021, automated sessions account for nearly two-thirds of Internet traffic.

Often, to neutralize the negative impact of automated Internet traffic generated by bots, methods such as rate limiter, IP blacklisting or blocking by specific identifiers are used. As identifiers here can be http-headers explicitly indicating belonging to a compromised network or use of technologies not typical of the "typical client", such as the browser used by the client. During the rate limiter operation, restrictions are used on the possible number of requests during a certain time, this mechanism as well as the use of identifiers has a major negative effect in the form of possible blocking of real users in case of non-standard situations, such as a regular break of Internet connection, causing multiple reconnections falling under the rate limiter or use of non-specific header by an updated user's browser.

Bad bots are created to carry out malicious actions. On social networks, they can be programmed to scan a platform for keywords and act according to a specific purpose. Bots can attack user accounts, collect personal data, they can mimic people and influence social media discussions or spread political propaganda, fake news, distort content and automatically share content from other profiles, reply to messages, etc.

The article consists of six sections. Section 1 is an introduction. Section 2 analyzes the related work and presents the results of the research conducted. Section 3 is devoted to the description of the methods used in the work. Section 4 gives the results obtained, 5 - interpretation, and finally in section 6 - conclusion, summarizing the results.

II. Related works

An extensive overview of social bots and their role in social media is provided in the following articles [2] [3] [4]. The task of bot detection is a classification task, it is similar to the task of filtering spam among email messages (spam - not spam). In the study [5] a Bayesian classifier was proposed to filter out spam emails and it was quite successful in removing 80% of the incoming unsolicited emails from the user's mail stream. In his article. [6] Wang A. H. argues that for the task of detecting spam bots on the social network Twitter, the best performance is in the Bayesian classifier, a classification

algorithm based on Bayes' theorem. It treats each account as a vector X with feature values that reflect the relationship between followers and users, and features based on the content that the user has posted over the past period of activity. Vectors are classified into two classes Y: spam and non-spam, for each class the posterior probability is calculated. The classification evaluation showed that in terms of precision - 0.917, recall - 0.917 and F-measure - 0.917, the Bayesian classifier has the best overall performance compared to other algorithms: the decision tree (precision - 0.667, recall - 0.333, F-measure - 0.444), neural networks (precision - 1, recall - 0.417, F-measure - 0.588) and support vector machine (precision - 1, recall - 0.25, F-measure - 0.4). Article [7] reviews spam filtering methods using machine learning, classifiers such as Decision Tree, Multilayer Perceptron, Naïve Bayes Classification. The performance of the three models was evaluated based on three criteria: Prediction Accuracy (Prediction Accuracy), Training Time (Training Time), and False Positive Rate (False Positive). In terms of Prediction Accuracy (Prediction Accuracy - 99.3) and False Positive (False Positive - 1), the perceptron classifier outperforms other classifiers: Decision Tree (Prediction Accuracy - 96.6, False Positive - 4) and Naïve Bayes Classification (Prediction Accuracy - 98.6, False Positive - 5). But the multilayer perceptron is inferior in terms of training time (Training time - 138.05), while Naïve Bayes Classification (Training time - 0.15) and Decision Tree (Training time - 0.20).

A similar study was conducted in the work [8] where a methodology for detecting and comparing fake Twitter profiles that are used for defamatory activities with the real profile by analyzing the content of comments was presented. Various machine learning techniques were applied in this work. The experimental results showed that SMO (Sequential Minimal Optimization) and Decision Trees are the most suitable algorithms for this task. The best results are obtained using PolyKernel with 68.47% accuracy and AUC of 0.96. In second and third position, very close, are J48 with 65.81% accuracy and AUC of 0.94 and NormalizedPolyKernel with 65.29% accuracy and AUC of 0.94. The random forest in the fourth position has an accuracy of 66.48% and an AUC of 0.93. Finally, the KNN and naive Bayes algorithms with values ranging from 59.39% to 61.06% accuracy for KNN and 33.91 for naive Bayes in terms of accuracy and 0.89 to 0.92 and 0.90 respectively in terms of AUC.

The paper [9] on detecting bots and assess their influence in social networks presents tools to achieve both goals. To identify bots, an algorithm based on the Ising model was developed to identify coordinated groups of bots. It uses minimal data and is capable of jointly identifying multiple bots with higher accuracy than current algorithms. The observation was that the Ising model algorithm achieves a true positive rate above 60% with a low false positive rate of about 5%. With a similar false positive rate, BotOrNot cannot achieve a true positive rate above 20%. Thus, the Ising model algorithm can achieve higher operating points than BotOrNot. The Ising model algorithm achieves an AUC (0.67 - 0.91) higher than BotOrNot on all but one event. However, the AUC for this event is lower than for the other events for both algorithms, suggesting that bot detection was generally difficult for this event.

In "BotOrNot" [10] to determine whether an account is a bot or not authors use Random Forest, one of the most powerful machine learning methods, which uses a set of decision trees for classification tasks. To classify an account as either a social bot or a human, the model is trained with instances of both classes. A large group of uncorrelated decision trees can produce more accurate and stable results than any of the individual decision trees. This paper states that a ten-fold cross validation yields an AUC of 0.95 (area under the ROC curve). Our work uses the Gradient Boosted Trees method.

In the articles [11] and [12] a comparative study of the most well-known methods of machine learning and intelligent analysis used for classification: decision tree, artificial neural network, k-nearest neighbor method and reference vector method, Bayesian network. The study showed that each method has its advantages and disadvantages, and it is very difficult to find one classifier capable of classifying all data sets with the same accuracy. For each method, there is a data set in which it is very accurate, and another data set in which it is not accurate. Each algorithm has its own implementation domain. None of the algorithms can satisfy all the criteria.

One of the recent interesting studies on the topic of social network bot detection was the "The DAPRA Twitter bot challenge." [13]. Based on the results of the contest participants Kantepe, M. and Ganiz, M. C. in their paper [14] took advantage of the different approaches used. They used several machine learning algorithms to build classification models that best discriminate between bot accounts: Logistic Regression (LR), Multinominal Naïve Bayes (MNB), Support Vector Machines (SVM) and ensemble learning method: Gradient Boosted Trees (GBT). There were 600 suspended accounts and 1200 non-suspended accounts in the study. The data were divided into 70% for training and 30% for testing. The Logistic Regression Algorithm gave 75% accuracy and 72% F1 score, Multinomial Naïve Bayes Algorithm gave 78% accuracy and 77% F1 score, SVM gave 82% accuracy and 75% F1 score, and Gradient Boosted Trees showed the best result with 86% accuracy and 83% F1 score.

The paper [15] developed an automated classification system consisting of four main parts: entropy, spam detection, account properties, and decision maker. The entropy component uses corrected conditional entropy to determine periodic or regular times, a variant of Bayesian classification is used to detect spam, and account properties are used to detect bot deviation from people. The decision maker, based on the Random Forest algorithm, analyzes the features identified by the other three components and decides: human, cyborg, or bot. This classification system can accurately distinguish human from bot. However, it is much harder to distinguish a cyborg from a human or a bot. After averaging the true positives for the three classes with an equal sample size, the overall accuracy of the system can be considered to be 96.0%.

In the paper [16] authors used a gradient-enhanced decision tree classifier (GBDT), an efficient classification of SA and sentiment (SC) Twitter data is proposed. The proposed performance of GBDT is analyzed and compared with Deep CNN (CNN), Artificial Neural Network (ANN), Deep Learning Neural Network (DLNN) and Deep Learning Modified NN (DLMNN) techniques with respect to metrics, namely: a) precision, b) recall, c) F-score, d) execution time, e) accuracy and f) average sentiment score. The analysis of experimental results has shown the highest performance of the proposed method: precision - 90.4534, recall - 93.112, F-score -92.0455.

In 2022, a study [17] was published on the development of a new system for identifying bots on Twitter. It uses a supervised machine learning (ML) framework using the extreme gradient enhancement (XGBoost) algorithm, where hyperparameters are tuned by cross-validation. Shepley's additive explanations (SHAP) are also used to explain the ML model predictions by calculating the importance of the features using Shepley values based on game theory. Experimental evaluation on different datasets demonstrate the superiority of this approach in terms of bot detection accuracy compared to the latest state-of-the-art Twitter bot detection method. As a basic step to create a robust and accurate bot identification system, a model selection procedure was performed by examining the classification accuracy of bots compared to regular users using several state-of-the-art ML algorithms. Random Forest (F1 - 0.908, PR-AUC - 0.955, ROC-AUC - 0.973), Support Vector Machine (SVM) (F1 - 0.889, PR-AUC - 0.941, ROC-AUC - 0.964) and Extreme Gradient Boost Algorithm (XGBoost) (F1 - 0.919, PR-AUC - 0.967, ROC-AUC - 0.979). The XGBoost model gives slightly better results according to the test data than SVM and Random Forest. The results show that the XGBoost model, when trained on a wide range of combined features, spanning from profile and context features to time-based and interaction features, provides the highest bot detection accuracy.

In [18], for the fake news detection problem, gradient-based boosting also showed the best results. Other classification models were analyzed in the paper and the results were simple Logistic Regression Classifiers, Passive Aggressive Classifiers and Random Forest Classifiers. While the models were good at classifying false and partially false and other classes, all models performed very poorly at predicting true classes.

However, the overall results were not unsatisfactory, as the model did classify false news and partially false news very well. The paper concluded that the models could perform better if optimal hyperparameters could be found. Random Forest classifier (Accuracy - 0.534, F1 - 0.502) performed well, but the XGBoost model (Accuracy - 0.571, F1 - 0.543) performed better than Random Forest overall, by a small margin. Passive Aggressive Classifier (Accuracy - 0.542, F1 - 0.489) is also very good in text classification tasks.

Concluding the review, based on the analysis of the related works listed above, we can conclude that to solve the problems of detecting bots in social networks, the gradient boosting method has the greatest efficiency and accuracy – the model used in our work. The advantage of our model is that it shows the best result and gives higher metrics than in the above works. Moreover, this work contributes to the scientific community by a more detailed examination of the model, for example, an analysis of the sensitivity of the model to the data frame size.

II. Methods

A. Data

In this work, historical data of a real social network were used. Activity data was taken from the event log of accessing backend resources. Data was also taken from the relational database tables about users, not including personal data. There were no gaps in the data. Manual markup was performed based on graph analysis and human activity analysis and viewing logs for marking up the training sample. The pandas data package python programming language was used for preprocessing.

B. Machine learning methods

Gradient boosting classifier from the scikit-learn library package was used as a model.

C. Model tuning

A greed search was used to select the best model parameters by the fields scale_pos_weight, colsample_by true, max_depth, learning_rate=0.5, n_estimators.

D. Model evaluation

To evaluate the model, a KFold cross validation method from the sklearn.model_selection library was used. 10 percent of the sample did not participate in the training and was used for the test. The following metrics were used to evaluate the model: accuracy, roc_auc, accuracy, recall, f1.

III. Results

A. Data sampling

The work analyzed 2.5 billion records of activity over 2 years in the social network YARUS. The initial size of datasets – 43011 records of user activity features and user properties features in the time period from 01.06.2021 to 01.06.2022, of which 1100 bots. When preparing the dataset, more than 100 user characteristics were analyzed, including: gender, age, city, user activity data, number of subscribers, number of subscriptions, number of comments, transaction data, device data, etc. A manual partitioning of bots and non-bots was used. Bots were marked accounts with the help of experts from the security department were checked according to closed criteria that do not relate to disclosure data. After that, sampling was done with library scikit-learn with parameter random_state = 5, which resulted in a balanced dataset of 43011 rows and 595 columns. The behavior of bots is described by 384 features with a median pass-fill strategy.

B. Data analysis

The data analysis was carried out in two stages. In the first stage, log data describing bot activity was analyzed. Methods of aggregation were used. At the second stage, a model was developed using machine learning methods.

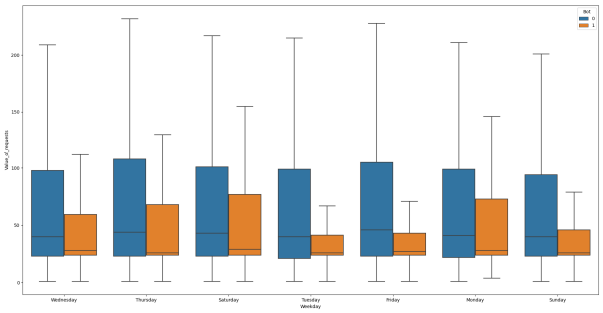

Fig. 1. Boxplot of user and bot activity by day of the week

Fig. 1 shows a boxplot of user and bot activity in the social network YARUS by days of the week in the time period from 01.06.2021 to 01.06.2022. Analyzing this data, we can see that Saturday is the day of the week with the highest activity among bots and the lowest activity among users, with the number of active bots approaching the number of active users on Saturday.

TABLE I. TABLE OF USER AND BOT ACTIVITY BY DAY OF THE WEEK

The difference between the number of active users and the number of active bots (from smaller to larger)	Bot activity (from more to less)	People's activity (from more to less)
Saturday (the number of active bots is roughly equal to the number of active people)	**Saturday**	Sunday
Monday	Monday	Monday
Thursday	Thursday	Thursday
Wednesday	Wednesday	Friday
Tuesday	Sunday	Wednesday
Friday	Friday	Tuesday
Sunday	Tuesday	**Saturday**

Further, Monday and Thursday are the days with the highest activity after Saturday for bots and Sunday for users among both regular users and bots. Also on Monday and Thursday the difference between the number of active users and the number of active bots is the lowest after Saturday compared to other days of the week.

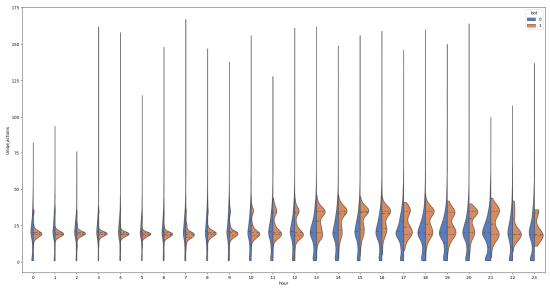

Fig. 2. Violin plot of the number of requests from bots and users as a function of time in a day.

Fig. 2 shows the activity (number of requests) of bots and ordinary users, depending on the time of day. This figure clearly shows that during the night and early morning hours, from 00:00 to 9:00, the bot shows stable activity, making approximately the same number of requests throughout this time. Further, the number of requests gradually increases, but there is a stable repetition of actions, which cannot be said about the activity of ordinary people, who are not inclined to prolonged maintenance of the same manipulations. The highest number of user requests is observed at 7:00, 8:00 in the morning, 8:00 in the evening and 4:00 in the middle of the day and are of a short duration.

C. Gradient boosting model

XGBClassifier gradient boosting was used as a training method for the classification model with parameters scale_pos_weight=100, colsample_bytree=1, max_depth=2, learning_rate=0.5, n_estimators=100, subsample=0.75, strategy for filling gaps by median, Training 5-fold Cross Validation Stratified KFold (n_splits=5, shuffle=True, random_state=42, TreeExplainer interpretation model_output='probability'.

The model was trained on a binary bot class. The ROC curve is shown in Fig. 3.

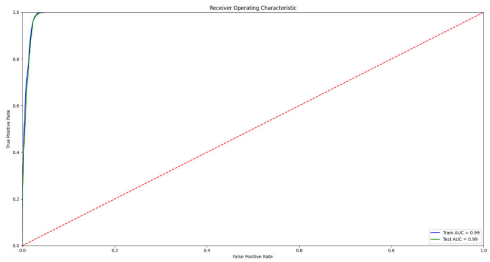

Fig. 3. ROC curve on the "bot"/"non-bot" class

From the graph above you can see that the resulting ROC curve is very close to the best (AUC=0.9999) algorithm.

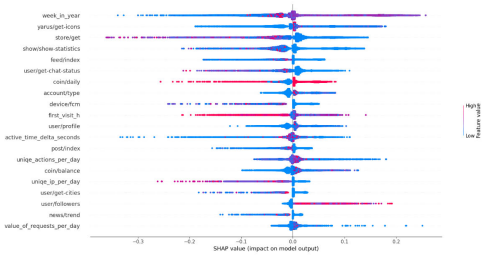

Fig. 4. Features importance (Shepley index) in the gradient boosting model

Among the significant predictors: the number of IP addresses, the time of the first access to the application, the time between requests, access to content.

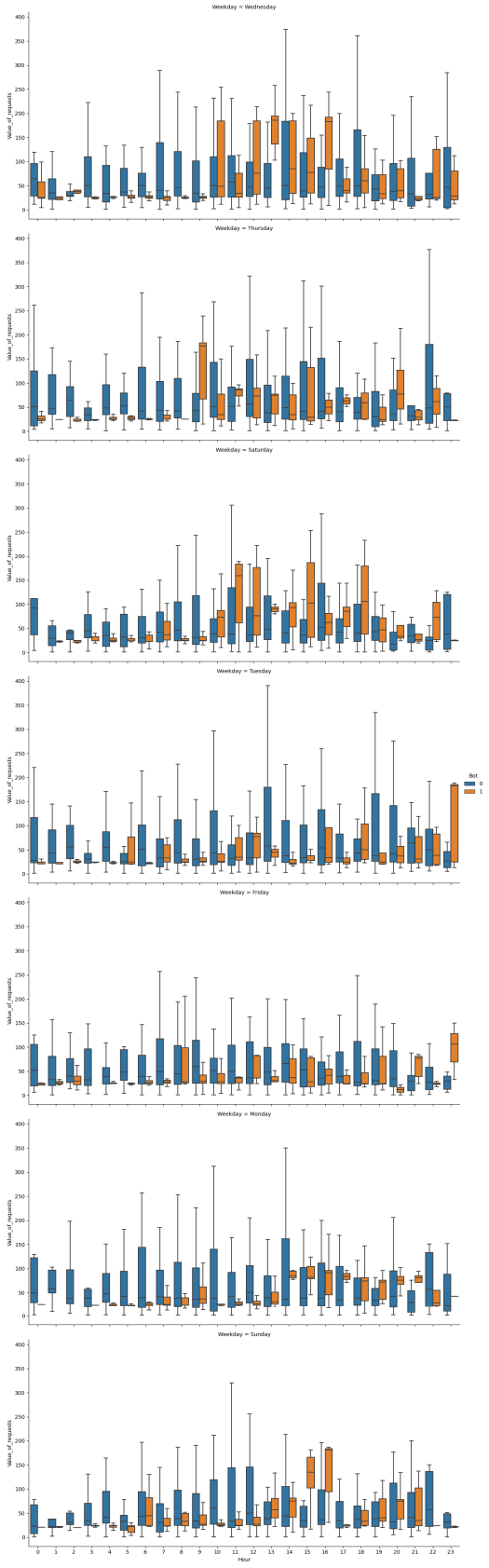

Fig. 5. Plot of the number of requests from bots on the time of day on different days of the week Bot/non-bot activity by hours per week

Fig. 5 shows bot and user activity by hours per week. We can see that the greatest number of bot requests occurs on Wednesday, Thursday and Saturday from 9:00 to 0:00. You can also see that at night people are more active than bots.

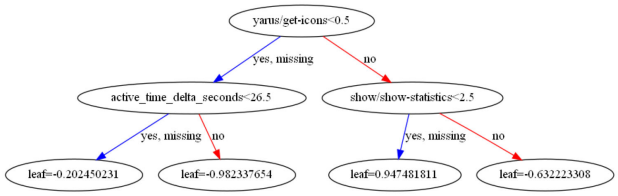

Fig. 6. The decision tree of the gradient boosting on the "bot" class

The sensitivity analysis used Stratified KFold at 5 fouls for cross validation using random datasets from the total set. The dynamics of metrics such as Accuracy, AUC, Precision, Recall and F1 Score were measured.

Fig. 7. Sensitivity analysis of train samples number

From the graph in Fig. 7 we can see that AUC metric stops growing after 2000, while Precision and Recall metrics stop growing only at 10000. From this we can conclude that the optimal number of training instances should be evaluated by Precision and Recall rather than by AUC. According to the dynamics of metrics change we got the optimal number of training data equal to 10000.

V. Discussion

This paper contributes to the field of Intelligence, Social Mining and Web Cloud Computing Systems, and Networks and Applications Algorithms and Modeling as it describes the results of developing models to identify bots from social network data.

This paper focuses on the problem of the presence of malicious bots in social networks. In our work to detect malicious bots in social networks, we apply gradient binning, a machine learning method for regression and classification problems, which generates a prediction model in the form of a

set of decision trees. The main difference from classical trees is in their learning process. The goal is to train multiple trees in stages. In each, we build one tree to correct errors previously established.

Through the application of machine learning and Process mining methods, we have developed a bot classification model that can certainly be used for information security processes in social networks. Since it is increasingly difficult to distinguish between bot and human actions in large volumes manually, the task of their automated detection, solved in this paper with AUC 98% accuracy, is relevant. Manual detection of such accounts is becoming a more time-consuming and costly process. Due to the fact that at the moment there are not enough studies on this topic and their results are somewhat contradictory, our paper is of even greater interest and relevance. In our work, we have developed and applied a new high-precision method effective for the tasks of fake profiles detection.

Having analyzed the related works, we can conclude that for the solution of our task the most effective method is still ensemble learning: Gradient Boosted Trees (GBT). The table below shows the results of two studies using this method and the results of our work. Since the metrics in the related studies were chosen slightly different, it is impossible to make a correct comparison. However, for the metrics that coincide, the results of our algorithm are more successful (Table II).

TABLE II. COMPARISON OF THE RESULTS OF USING THE GRADIENT BOOSTING METHOD IN DIFFERENT WORKS

	Related works			This work
Subject	Detecting bots in the social network	Effective classification of SA and sentiment (SC) Twitter data	Identification of Twitter Bots Based on an Explainable Machine Learning Framework	Detecting bots
Article	[14]	[16]	[17]	This work
The method	ensemble learning method: Gradient Boosted Trees (GBT)	gradient-enhanced decision trees (GBDT)	extreme gradient boosting (XGBoost)	gradient boosting
Accuracy	0.86	-	-	**0.9984**
Recall	0.85	93.1	-	**0.9995**
F1	0.83	92.0	0.919	**0.9926**
Precision	-	90.45	0.967	**0.9859**
AUC	-	-	0.979	**0.9999**

VI. CONCLUSION

Malicious bots in social networks are one of the most dangerous and widespread types of cybercrime, and they are becoming increasingly difficult to detect. Classical methods for detecting their activity have many disadvantages in the form of false positives and constant revision of rules due to changes in bot activity (frequency, types of requests). The problem of regularly adjusting bot systems' activities to the applied information security policies also remains relevant. Therefore, to solve this problem, it is advisable to use machine learning and intelligent analysis techniques. For successful bot detection, we applied Gradient Boosting (GB), an automatic retrainable high-precision method. Our experimental evaluation shows the effectiveness of the proposed classification system. We obtained models with the following characteristics: AUC - 0.9999, Accuracy - 0.9984, Precision - 0.9859, Recall - 0.9995, F1 - 0.9926.

REFERENCES

[1] We Are Social and Hootsuite. Digital 2022: Global overview report // www.datareportal.com. 2022.

[2] Ferrara E. et al. The rise of social bots // Commun ACM. 2016. Vol. 59, № 7.

[3] Wagner C. et al. When social bots attack: Modeling susceptibility of users in online social networks // CEUR Workshop Proceedings. 2012. Vol. 838.

[4] Orabi M. et al. Detection of Bots in Social Media: A Systematic Review // Inf Process Manag. 2020. Vol. 57, № 4.

[5] Sahami M. et al. A Bayesian approach to filtering junk e-mail // Learning for Text Categorization: Papers from the AAAI Workshop. 1998. Vol. WS-98-05, № Cohen.

[6] Wang A.H. Detecting spam bots in online social networking sites: A machine learning approach // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2010. Vol. 6166 LNCS.

[7] Deepika Mallampati. An Efficient Spam Filtering using Supervised Machine Learning Techniques // International Journal of Scientific Research in Computer Science and Engineering. 2018. Vol. 6, № 2.

[8] Galán-García P. et al. Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying // Log J IGPL. 2014. Vol. 24, № 1.

[9] des Mesnards N.G. et al. Detecting Bots and Assessing Their Impact in Social Networks // Oper Res. 2022. Vol. 70, № 1.

[10] Ferrara E. et al. BotOrNot: A System to Evaluate Social Bots Clayton // arXiv preprint arXiv:1407.5225. 2014. № grant 220020274.

[11] Mohamed A.E. Comparative Study of Four Supervised Machine Learning Techniques for Classification // Int J Appl Sci Technol. 2017. Vol. 7, № 2.

[12] Bhavsar H., Ganatra A. A Comparative Study of Training Algorithms for Supervised Machine Learning // International Journal of Soft Computing and Engineering. 2012. Vol. 2, № 4.

[13] Subrahmanian V.S. et al. The DARPA Twitter Bot Challenge // Computer (Long Beach Calif). 2016. Vol. 49, № 6.

[14] Kantepe M., Gañiz M.C. Preprocessing framework for Twitter bot detection // 2nd International Conference on Computer Science and Engineering, UBMK 2017. 2017.

[15] Chu Z. et al. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? // IEEE Trans Dependable Secure Comput. 2012. Vol. 9, № 6.

[16] Neelakandan S., Paulraj D. A gradient boosted decision tree-based sentiment classification of twitter data // Int J Wavelets Multiresolut Inf Process. 2020. Vol. 18, № 4.

[17] Shevtsov A. et al. Identification of Twitter Bots Based on an Explainable Machine Learning Framework: The US 2020 Elections Case Study. 2021.

[18] Utsha R.S. et al. Qword at CheckThat! 2021: An extreme gradient boosting approach for multiclass fake news detection // CEUR Workshop Proceedings. 2021. Vol. 2936.