International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(17s) 2026 pp. 157-167
https://doi.org/10.70917/ijcisim-2026-4712
Article

INTERPRETABLE MACHINE LEARNING FOR DETECTION OF SPAMBOTS AND FAKE FOLLOWERS ON SOCIAL NETWORKS USING FEATURE BASED AND TEXT BASED

Kandukuri Sai Prasad[1], S. V. Achutha Rao[2]

[1] Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Sciences, Sheriguda, Hyderabad, Telangana, India.
Email: saiprasadkandukuri@gmail.com
[2]Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Sciences, Sheriguda, Hyderabad, Telangana, India.
Email: dean.academics@sreedattha.ac.in

Abstract: Social networking sites like Twitter facilitate a wide interaction with humanity but are infiltrated more by robotics which imitating the human behavior propagate false information and control masses. To ensure the information integrity, Spambot detection is necessary, whereas much of the existing techniques rely on opaque, black box models, which restricts transparency and interpretability. In this paper, interpretable machine learning on spambot and false follower classification has been explored on the Cresci-15 and Cresci-17 datasets. Normalization, and tokenization, as well as the removal of extraneous material are part of preprocessing of text-based and feature-based data. SMOTE and SMOTEENN methods eliminate imbalance in the classification, and the RFE method reduces the feature dimension. They are the Decision Tree, the Random Forest, the SVM, the XGBoost, the AdaBoost, the Stacking Classifier and the Voting Classifier. These findings demonstrate that Stacking Classifier has the highest accuracy of 99.9 per cent in Cresci-15 and 99.5 per cent in Cresci-17. Explainable AI tools such as LIME and SHAP show the importance of features, which enhances transparency and decision-making with the model. Such a combination of feature selection, advanced resampling, and ensemble learning algorithms and interpretable methods of robust social network automatic recognition of accounts is demonstrated to be working.
Keywords: Interpretable AI, social network, bot detection, fake followers, spambots

1. Introduction

The new digital information distribution is focused on social networks. X (when it was Twitter) is one of the most popular and significant platforms that enable millions of active users to communicate and discuss in real-time [1]. Its enormous social and economic sway has liftered bad players who have employed its transparency to appeal to the masses and spread lies. This is usually done with bots. Bad bots transmit spam, rumors and even harmful information than good bots, which give instructional blogs and news feeds [2].

Online conversation integrity is crucial to detecting those bogus accounts. Bots can be detected using user metadata, features of user behavior, timestamps, and network architecture [3]. It requires time and familiarity with the subject to feature engineer, or design, and extract these characteristics. Bots are developed to mimic humans, thus becoming more difficult to spot. On the one hand, through a special selection of high-profile accounts, malicious bots cover disinformation such as fake news and hate speech [4].

Automated manipulation can be made difficult by botnets and Sybil accounts. A botnet is a group of synchronized bots, which engage in malicious activities, and Sybil identities are modified identities that aim to corrupt communications [5], [6]. These methods are inappropriate to encourage the authentic engagement and the spread of the information with low credibility and raise issues of trust and reliability in the social media ecosystems.

These risks make machine learning (ML) apply to sports analytics [7], sentiment analysis [8], [9], and false news identification [10]. The social bot detectors using the ML model have been used in recent years to surpass heuristic and network based methods. ML-based detection models tend to be black-box models, which are not very interpretable. Transparency lack rendering renders prediction difficult to monitor as researchers and practitioners cannot discern whether predictions are rooted on actual patterns or overfitted noise.

To avoid these limitations, interpretable machine learning (XAI) is being introduced into bot detection systems. XAI enhances responsibility and trust since it displays the influence of individual variables on model predictions. The application of XAI-based frameworks enhances recognition of bots on X and other similar social networks through developing accuracy and interpretability by closing the disconnect between the human judgment and algorithm judgment.

2. Related Work

Social network bot identification is a very critical research area because of the development of fraudulent profiles that dominate online conversation and information system. To enhance detection accuracy and scalability, advanced machine learning, graph-based methods and interpretable models have become the means used by researchers.

Mbona and Eloff [11] reported that news is provided by many bots, and they can be considered both hostile and benign. Learning strategy with minimal labeled and a lot of unlabeled data is a semi-supervised approach that lowers the annotation costs and enhances bot classification. Yang and others [12] proposed data selection method in ensuring that machine learning models that are being trained on representative datasets perform effectively on different environments, which is critical in the quest to create versatile detection systems.

Bots that are generated based on AI suggest new problems. Lopez-Joy et al. [13] analyzed how massive language model bots are designed and discovered that realistic and contextually relevant content of the chosen type makes them more difficult to recognize than spambots. Their priorities were on updating methods of AI-enhanced opponents. Moreover, Pham [14] proposed a combined framework, which captures latent behavioral features and structural social dynamics based on a pre-trained auto-encoder with graph neural networks (GNNs). This composite approach was superior to monitored techniques in network generalizability.

Terumalasetti and Reeja [15] have come up with a structure of multi-dimensional analytics, which combines the features of content, temporal and relational attributes to enhance the user credibility. Their broad approach saved them against adaptive spambots which proves the necessity of multidimensional models and not feature categories. Paudel et al. [16] demonstrated the development of bot methods in order to reflect human interactions in the network. Primarily with the help of the complex network analysis, they were able to find out that such structural statistics as clustering coefficients and other measures of centrality could be used to detect advanced bots.

The paper of Aljabri et al. [17] has critiqued the literature on machine learning-based bot detection to synthesize the advancements. They observed the feature engineering dependency, scalability problems, black box model transparency. Their evaluation emphasized the need to make a tradeoff between detection accuracy, interpretability and flexibility. The Botshape tool which was developed by Wu et al. [18] examines various behavioral patterns such as the frequency of posting and the type of temporal activity. Their method was better than human timing bots which simulation of dynamic behavior.

Graph techniques are also prevalent. Bebensee et al. [19] proposed the application of node neighborhoods and egraph topology to characterize variation in relationships of social graph microstructures that do not possess created characteristics. They also proved that graph-based learning can be used to identify anomalies in contextualized network frameworks. Finally, Dimitriadis et al. [20] suggested Caleb, a conditional adversarial system reinforced against adaptive bot algorithms. They used the strategy to generalize and to be stronger, training on adversarial cases, solving the detection system-adversary arms race.

3. Materials And Methods

By combining both the feature based and textual information of Cresci-15 and Cresci-17 datasets, the suggested approach involves the design of a powerful and explainable machine learning model to be used in the detection of social media spambots and fake followers. Preprocessing of the data includes missing values, encoding of categorical variables, standardization of numbers in the form of a fraction, cleaning of texts (HTML tags or URLs or non-printable characters), normalization, tokenization, and TF-IDF representation. SMOTE and SMOTEENN deal with disparity of classes. The decision tree model, the random forest model, the SVM model, and the XGBoost model, Naive Bayes, AdaBoost, The Stacking / Voting Classifier are all used to build various predictive models. LIME and SHAP are interpretable: meanwhile, Recursive Feature Elimination is not. It is possible to detect the decision of the models and to interactively analyse them in real time with Flask.

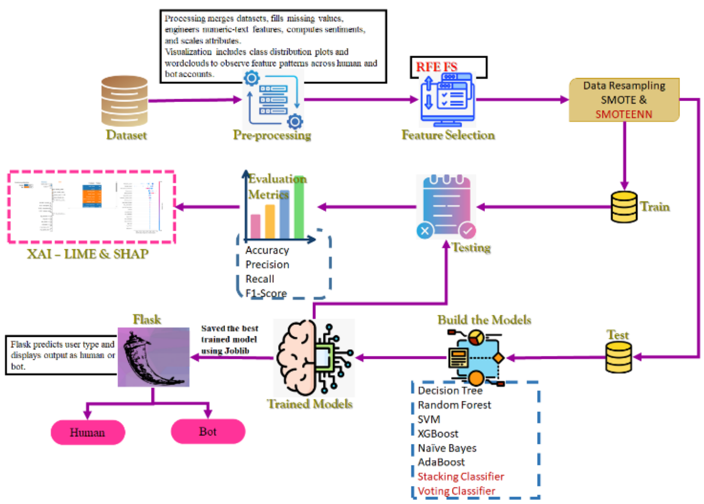


Fig. 1. System Architecture

The Cresci dataset is cleaned and prepared in the proposed system architecture through data preparation. The features engineering and selection are extracted to obtain the representative tweet, user, and sentiment characteristics. Introduction of processing information into a training, validation and testing achievement produces modeling feature vectors. Fine-tuned classifiers are used to differentiate on human and bot accounts. Explainable AI techniques (LIME, SHAP) increase the trust and interpretability of the bot decisions by providing the explanations of the model choices.

a) Dataset Collection:

i) Cresci-17: Cresci-17 dataset was a study of social spambots, conventional spambots, false followers and actual Twitter accounts, which comprised of 10,894 bots and 3,474 people at the user level and 3,798,254 items and 2,839,362 items at the bot level. There were several different types of accounts: social, conventional and false followers. Removing, aggregating, and labeling profile and tweet characteristics of every user formed a balanced dataset to use in a supervised machine learning task and translatable artificial intelligence-based bot detection.

	user_id	followers_count	friends_count	statuses_count	favourites_count	listed_count	verified	default_profile	created_at	account_age_days	description	description_len
0	1502026416	208	332	2177	265	1	0	0	2013-06-11 11:20:35	4474	15years ago X.Lines24	
1	2492782375	330	485	2660	3972	5	0	1	2014-05-13 10:37:57	4138	[illegible]	
2	293212315	166	177	1254	1185	0	0	0	2011-05-04 23:30:37	5243	Let me see what your best move is!	
3	191839658	2248	981	202968	60304	101	0	0	2010-09-17 14:02:10	5472	20. menna: #farida #nyc and the 80s actually y...	
4	3020965143	21	79	82	5	0	0	1	2015-02-06 04:10:49	3870	Cosmetologist	

Fig.2 Cresci-17 dataset

ii) Cresci-15: Cresci-15 dataset is a combination of 5, 301 individuals and 2, 827, 757 tweets of E13, TFP, FSF, INT and TWT. Human accounts and bot accounts are 1, 950 and 3,351 respectively. All users have 35 raw features, which consist of follows, friends, statuses, favorites, list count, verified, profile information, and date of account creation. Tweets are enhanced by word count, hashtags, mentions, URLs, emoticons, punctuation, emotion and engagement analytics. The ability to perform comprehensive bot identification and behavioral analysis can be done using 29 user-level variables.

	user_id	followers_count	friends_count	statuses_count	favourites_count	listed_count	verified	default_profile	created_at	account_age_days	description	description_length
0	3610511	5470	2385	20370	145	52	0	0	2007-04-06 10:58:22	6733	Founder of http://www.screenweek.it & http://w...	151
1	5656162	506	381	3131	9	40	0	0	2007-04-30 15:08:42	6709	BSc degree (cum laude) in Computer Engineering...	104
2	5682702	264	87	4024	323	16	0	0	2007-05-01 11:53:40	6708	Cogito ergo bestemmio.	22
3	6067292	640	622	40586	1118	32	0	0	2007-05-15 16:55:16	6694	Se la vita ti dÃ sarde, scapocciale!	37
4	6015122	62	64	2016	13	0	0	0	2007-05-13 19:52:00	6695	Je me souviens	14

Fig.3 Cresci-15 dataset

b) Pre-Processing:

The Cresci-15 and Cresci-17 data sets had been preprocessed to prepare, convert, and optimize user and tweet information in order to have robust and accurate bot and human accounts.

1. Data Processing: The user and tweet data of Cresci-15 and Cresci-17 were combined with each other in all the sub-datasets. Missing followers, friends and statuses and likes were being filled with zeros. Datetime account creation dates were used to estimate age in days of an account. To ensure the consistency in account identification in both datasets to be used in downstream feature engineering and model training, standardized labels were used (0=human, 1=bot).

2. Feature Engineering: The user and tweet data were mined to provide meaningful characteristics. The metrics within the user were followed to friends ratio, description length, presence of a description, and sentiment ratings. Characteristics at the tweet level include word counts, hashtags, mentions, URLs, emoticons, punctuations, unique word ratio, ratio of capitalization, length of tweets, and emotion. To give a comprehensive set of features to be modeled, average hashtags, mentions, emojis, interaction, sentiment per user were calculated.

3. Visualization: Patterns, ranges and variability in both feature distributions were studied. Histograms, boxplots and scatter graphs were used to plot followers, friends, length of tweet and age of the account. Outliers, skewed distribution, and bot-human difference were detected through a visual analysis, and the results influenced the preprocessing and normalization choices to influence the feature selection and model development.

4. Feature Selection: Numeric and aggregated characteristics of tweets were selected to enhance their predictability. We dropped unnecessary identifiers and degree and columns that were not useful. Facilities that are retained are followers, friends, followers-friends ratio, description sentiment, average hashtags, mentions, emoticons, engagement analytics, and tweet-level data. The models of feature selection involved the most informative features so that efficiency, reduction of noise and accuracy of the classification of the data were increased in both datasets.

5. Data Balancing: In databases, humans were outnumbered by bots. This was solved by over sampling the minority group or fabricating data to even out human and bot numbers. Similar datasets minimized the dominance of one of the classes and improved prediction. User-level aggregated characteristics were consistent and offered equal training data to use in accurate Cresci-15 and Cresci-17 categorization.

6. Scaling: Normalization was done using min-max or z-score to scale the followers, friends, statuses, likes, tweet counts and engagement measures. Scaling transformed range of the features, minimized the extreme values, and made all the characteristics equal to the model training. This amplified the convergence of gradient-based classifier and feature matching, elevating the level of prediction accuracy on Cresci-15 and Cresci-17 datasets.

c) Training and Testing:

In order to maintain the distribution of classes, processed and feature-engineered data were stratified and split into training and testing sets. The classification of supervised classifiers is trained using user-level and aggregated tweets. The accuracy, precision, recall, and F1-score were used as model performance measures. It made it possible to strongly generalize across Cresci-15 and Cresci-17 datasets and make full feature representations that recognize bot and human accounts.

d) Algorithms

Decision Tree: Further, Decision Tree recursively divides data based on the most useful characteristics in order to investigate feature-based and text-based features. The interpretable structure can be classified to detect spambots and false followers easily as it can classify accounts and handle category and numerical data.

$$I\,(i) = 1 - \sum_{i=1}^{k} \quad p_i^2 \quad (1)$$

Random Forest: RF also constructs a number of decision trees and averages the outcomes of these trees to avoid noise and overfitting. It looks at numerous features at the same time to enhance generality and precision and identify automated accounts which have complicated, nonlinear social network activities.

SVM: That is done to obtain the best hyperplane to separate the human and the automated account classes, obtained through SVM. Its high-dimensional analysis of the kernel functions enables to identify exactly the difference of features and yet resist overfitting and a noisy dataset.

XGBoost: The Gradient optimization builds casing improved decision trees in XGBoost which reflects the interaction of features and aids better performance compared to standardized models. It copes with the imbalance, analyses both structured and text-based properties in good time, scales to enormous data, and works well in detecting spambots.

$$\hat{y}_i = \sigma\left(\sum_{k=1}^{K} \quad f_k(x_i)\right), f_k \in F \quad (2)$$

Naïve Bayes: Naive Bayes relies on the conditional independence of its features to compute categorization that is based on probabilities. It quickly and ideas classifies large volumes of reading and orchestral data and information by probability. It is also basic and effective, offering a baseline of lightweight and reliable bot identification.

$$P(B) = \frac{P(A).P(A)}{P(B)} \quad (3)$$

AdaBoost: The sequential weak classifiers in AdaBoost pay more weight to the misclassified cases so that they can be detected. Enhances consistency in prediction and reduces bias, minimizes imbalance whilst preserving interpretability and produces a composite model of spambot detection in corrupt data.

Stacking Classifier: It uses predictions of a base model to combine them with a meta-classifier. It tries to detect false and automated accounts based on evaluating feature-based and text-based characteristics that can capture different levels of strengths in algorithms and minimize weaknesses and enhance detection rates.

$$y\hat{} = g(Y_{base}) = g\big(f_1(x), f_2(x), \ldots, f_m(x)\big) \quad (4)$$

Voting Classifier: It combines majority consensus with the outputs of classifiers to improve the stability of the decision-making and reduces model errors. The balance and reliable results with complementary algorithmic views are detected with easy to understand results that are easy to understand by the user.

$$\hat{y} = argmax_c \left(\sum_{i=1}^{n} \quad II\ (\hat{y}_i = c) \right) \quad (5)$$

e) Integration of XAI and Flask Framework:

Explainable Artificial Intelligence (XAI) is a machine learning framework that uses the flask to make the application transparent and reachable. Such XAI-based algorithms as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) can present feature contributions and decision-making mechanisms in model predictions. Such explanations give a consumer the idea that this is a bot or a human because such a model has selected an account and they create confidence and responsibility.

Such models and illustrations are deployed with the flask which is a light weight python web framework. It provides an easy way to combine both the AI models running in the back and the user-interfaces running in the front. The system is user-friendly, intuitive, and transparent to generate input data, receive forecasts, and have complete XAI explanations in real-time.

4. Experimental Results

Accuracy: The accuracy of a test is its ability to identify the sick and healthy cases. Computation of fraction of true positive and true negative in aggregate number of cases analyzed is calculated to determine accuracy of test. Mathematically, this is:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

Precision: Precision has to do with the percentage of positive samples or cases that are correctly categorized. Precision can be determined as follows:

$$Precision = \frac{True\ Positive}{True\ Positive\ + False\ Positive} \quad (7\ )$$

Recall: Machine learning recall is used to assess how a model can identify all the pertinent examples of a class. It presents the thoroughness of the model in volumizing occurrences of a class involving contrasting accurately predicted positive observations with the sum of positives.

$$Recall = \frac{TP}{TP\ +\ FN} \qquad (8)$$

F1-Score: The accuracy of the machine learning model is given in F1 score. A combination of model recall and preciseness. Accuracy statistic is given as the predicted frequency of the model on the dataset.

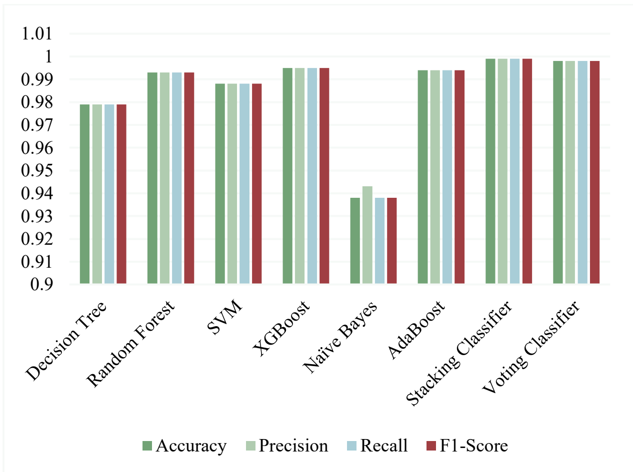
$$F1\ Score = 2 * \frac{Recall\ X\ Precision}{Recall + Precision} * 100 \quad (9)$$

Table.1 Performance Evaluation – Cresci – 15

ML Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.979	0.979	0.979	0.979
RF	0.993	0.993	0.993	0.993
SVM	0.988	0.988	0.988	0.988
XGBoost	0.995	0.995	0.995	0.995
Naïve Bayes	0.938	0.943	0.938	0.938
AdaBoost	0.994	0.994	0.994	0.994
Stacking Classifier	0.999	0.999	0.999	0.999
Voting Classifier	0.998	0.998	0.998	0.998

In Table.1, comparative results highlight that the Stacking Classifier achieved the highest overall performance among all evaluated algorithms.

Fig.4 Comparison Graph – Cresci – 15



In fig.4 highlights the Stacking Classifier as the highest-performing algorithm, where Accuracy is green, Precision is dark green, Recall is light blue, and F1-Score is red across all models.

Table.2 Performance Evaluation – Cresci – 17

ML Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.975	0.975	0.975	0.975
RF	0.990	0.990	0.990	0.990
SVM	0.954	0.955	0.954	0.954
XGBoost	0.991	0.991	0.991	0.991
Naïve Bayes	0.689	0.799	0.689	0.657
AdaBoost	0.987	0.987	0.987	0.987
Stacking Classifier	0.995	0.995	0.995	0.995
Voting Classifier	0.987	0.988	0.987	0.988

In Table.2, results show that the Stacking Classifier achieved the highest overall performance compared to other evaluated machine learning models.

Fig.5 Comparison Graph – Cresci – 17

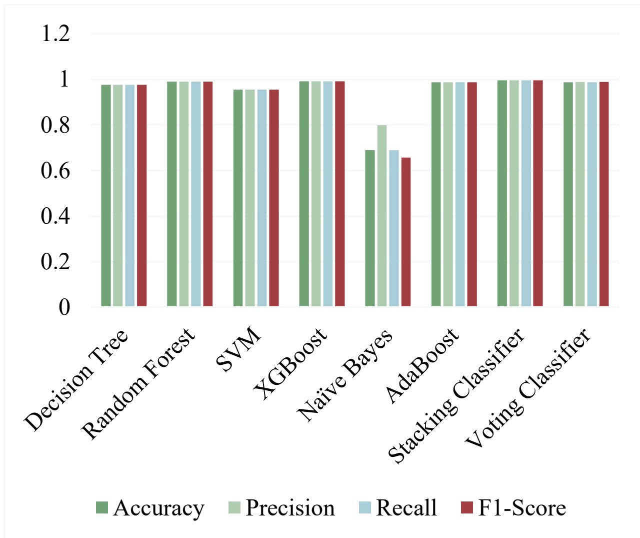


Fig. 5 indicates how a Stacking Classifier is the most successful among eight classifiers on Accuracy (green), Precision (light green), Recall (light blue), and F1-Score (brown).

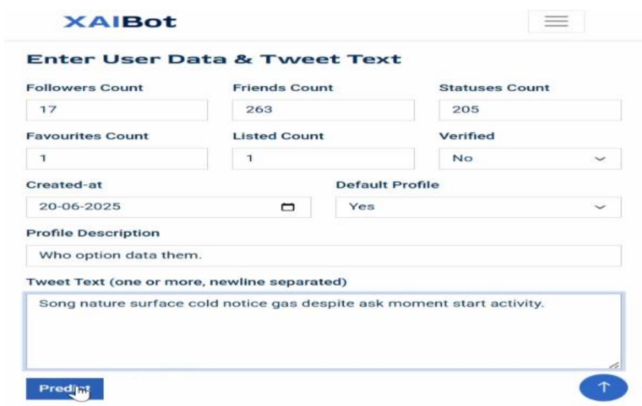


Fig.6 Enter Input Data

Figure 6 features the twitter interface that allows users to input personal data and tweet content, including followers, friends and status numbers.

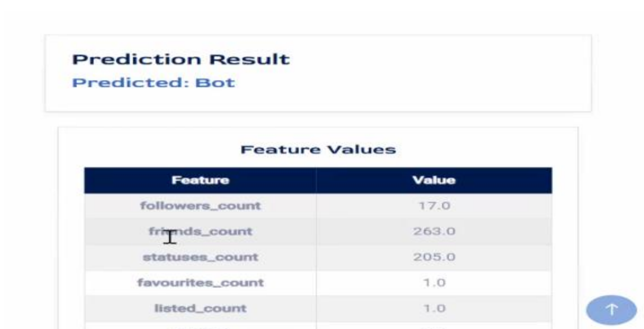


Fig.7 Predicted Results

Figure 7 depicts the output of the prediction of Bot and the values of features that were typed in to analyse and be classified.

Fig.8 Enter Input Data

In Fig. 8, users insert a username in order to analyze whether an account is a human or a bot.

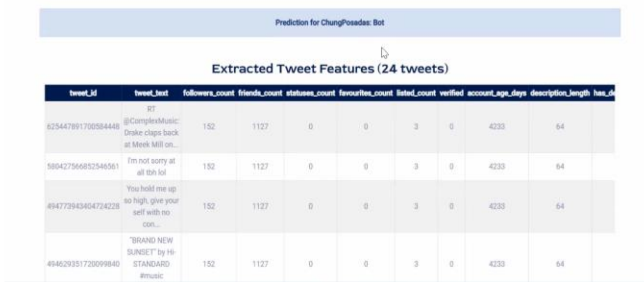


Fig.9 Predicted Results

Fig.9 illustrates as anticipated intimates Bot on both the calendar and account characteristic data.

5. Conclusion

The report indicates that social media spambots and fake following can be identified by focusing on features and utilizing text-based data along with an efficient machine learning. This was done by the models through the Recursive Feature Elimination (RFE) method which helped in the reduction of features and with the SMOTE and SMOTEENN methods which helped the models to resolve the issue of class imbalance. The ensemble learning was the most effective achieving 99.9% accuracy on the Cresci-15 dataset and 99.5% accuracy on the Cresci-17 dataset which proved the consistency of the algorithm combination. Explainable AI methods such as LIME and SHAP give a thorough information of feature importance, which guarantee a certain degree of interpretability and accuracy. The higher ensemble models based on appropriate chosen characteristics and balanced datasets enhance automated account detection, whereas interpretable algorithms enhance the understanding of the decisions. The results demonstrate that feature selection, sampling methods, and explainable machine learning can enhance the accuracy of detection, transparency, and reliability, which is a full solution to monitoring and countering spambots and fake followers within the social network setting.

The detection system can be upgraded to perform more sophisticated automated accounts that adapt their behaviour to escape in the subsequent research. Models can identify spambot behavior by real-time feeds of numerous social networks to enhance relevant scalability and reactivity. Language models that are based on transformers and more sophisticated deep learning techniques may provide an ability to capture more intricate textual patterns to feature-based signatures. There are more multimodal data as photos, videos, and network interactions can enhance detection accuracy. The explainable AI methods will help users and administrators to assess predictions with more complex model by guaranteeing transparency. It will also be relevant to enhance the efficiency of the computing systems in the implementation of social networks on large scales.

REFERENCES

1. Alkathiri, N., & Slhoub, K. (2025). Challenges in machine learning-based social bot detection: a systematic review. *Discover Artificial Intelligence, 5*(1), 214.
2. Akhtar, M. M., Bhuiyan, N. S., Masood, R., Ikram, M., & Kanhere, S. S. (2025). BotSSCL: Social Bot Detection with Self-Supervised Contrastive Learning. *Online Social Networks and Media, 48*, 100318.
3. Lopez-Joya, S., Diaz-Garcia, J. A., Ruiz, M. D., & Martin-Bautista, M. J. (2024). Exploring social bots: a feature-based approach to improve bot detection in social networks. *arXiv preprint arXiv:2411.06626*.
4. Nguyen, H. D., Nguyen, D. Q., Nguyen, C. D., To, P. T., Nguyen, D. H., Nguyen-Gia, H., ... & Quan, T. (2024). Supervised learning models for social bot detection: Literature review and benchmark. *Expert Systems with Applications, 238*, 122217.
5. Sallah, A., Agoujil, S., Wani, M. A., Hammad, M., Maleh, Y., & Abd El-Latif, A. A. (2024). Fine-tuned understanding: Enhancing social bot detection with transformer-based classification. *IEEE Access, 12*, 118250-118269.
6. Mou, G., & Lee, K. (2020, October). Malicious bot detection in online social networks: arming handcrafted features with deep learning. In *International conference on social informatics* (pp. 220-236). Cham: Springer International Publishing.
7. Shevtsov, A., Tzagkarakis, C., Antonakaki, D., & Ioannidis, S. (2022, May). Identification of twitter bots based on an explainable machine learning framework: The US 2020 elections case study. In *Proceedings of the international AAAI conference on web and social media* (Vol. 16, pp. 956-967).
8. Deshmukh, A., Moh, M., & Moh, T. S. (2024, December). Bot Detection in Social Media Using GraphSage and BERT. In *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 804-811). IEEE.
9. Pham, P., Nguyen, L. T., Vo, B., & Yun, U. (2022). Bot2Vec: A general approach of intra-community oriented representation learning for bot detection in different types of social networks. *Information Systems, 103*, 101771.
10. Javed, D., Jhanjhi, N. Z., & Khan, N. A. (2023, July). Explainable Twitter bot detection model for limited features. In *IET Conference Proceedings CP837* (Vol. 2023, No. 11, pp. 476-481). Stevenage, UK: The Institution of Engineering and Technology.
11. Mbona, I., & Eloff, J. H. (2023). Classifying social media bots as malicious or benign using semi-supervised machine learning. *Journal of Cybersecurity, 9*(1), tyac015.
12. Yang, K. C., Varol, O., Hui, P. M., & Menczer, F. (2020, April). Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 01, pp. 1096-1103).
13. Lopez-Joya, S., Diaz-Garcia, J. A., Ruiz, M. D., & Martin-Bautista, M. J. (2025). Dissecting a social bot powered by generative AI: anatomy, new trends and challenges. *Social Network Analysis and Mining, 15*(1), 7.
14. Pham, P. (2025). An Integrated Pre-Trained Auto-Encoder and Graph Neural Network for General Social Bot Detection. *Annals of Data Science*, 1-21.
15. Terumalasetti, S., & Reeja, S. R. (2024). Enhancing social media user’s trust: A comprehensive framework for detecting malicious profiles using multi-dimensional analytics. *IEEE Access*.
16. Paudel, P., Nguyen, T. T., Hatua, A., & Sung, A. H. (2019, August). How the tables have turned: Studying the new wave of social bots on Twitter using complex network analysis techniques. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 501-508).
17. Aljabri, M., Zagrouba, R., Shaahid, A., Alnasser, F., Saleh, A., & Alomari, D. M. (2023). Machine learning-based social media bot detection: a comprehensive literature review. *Social Network Analysis and Mining, 13*(1), 20.
18. Wu, J., Ye, X., & Mou, C. (2023). Botshape: A novel social bots detection approach via behavioral patterns. *arXiv preprint arXiv:2303.10214*.
19. Bebensee, B., Nazarov, N., & Zhang, B. T. (2021). Leveraging node neighborhoods and egograph topology for better bot detection in social graphs. *Social Network Analysis and Mining, 11*(1), 10.
20. Dimitriadis, I., Dialektakis, G., & Vakali, A. (2024). Caleb: a conditional adversarial learning framework to enhance bot detection. *Data & Knowledge Engineering, 149*, 102245.
21. Guo, S., Wang, J., Wang, Z., Yu, G., & Wu, S. (2025). BotICC: enhancing social bot detection through implicit connection computation. *Multimedia Systems, 31*(3), 210.
22. Qiao, B., Li, K., Zhou, W., Li, S., Lu, Q., & Hu, S. (2025, April). Identifying Bots on Social Media through Coordinated Group Perception. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
23. Moghaddam, S. H., & Abbaspour, M. (2022). Friendship preference: Scalable and robust category of features for social bot detection. *IEEE Transactions on Dependable and Secure Computing, 20*(2), 1516-1528.
24. Ellaky, Z., Benabbou, F., Ouahabi, S., & Sael, N. (2021, June). A survey of spam bots detection in online social networks. In *2021 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA)* (pp. 58-65). IEEE.
25. Wu, J., Ye, X., & Man, Y. (2023, May). Bottrinet: A unified and efficient embedding for social bots detection via metric learning. In *2023 11th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.
26. Najari, S., Salehi, M., & Farahbakhsh, R. (2022). GANBOT: a GAN-based framework for social bot detection. *Social Network Analysis and Mining, 12*(1), 4.
27. Sánchez-Corcuera, R., Zubiaga, A., & Almeida, A. (2024). Early Detection and Prevention of Malicious User Behavior on Twitter Using Deep Learning Techniques. *IEEE Transactions on Computational Social Systems*.
28. Kouvela, M., Dimitriadis, I., & Vakali, A. (2020, November). Bot-Detective: An explainable Twitter bot detection service with crowdsourcing functionalities. In *Proceedings of the 12th International Conference on Management of Digital EcoSystems* (pp. 55-63).
29. Li, Y., Li, Z., Gong, D., Hu, Q., & Lu, H. (2024). BotCL: a social bot detection model based on graph contrastive learning. *Knowledge and Information Systems, 66*(9), 5185-5202.
30. Kumar, A. S., Kumar, N. S., Devi, R. K., & Muthukannan, M. (2023). Analysis of deep learning-based approaches for spam bots and cyberbullying detection in online social networks. *AI-Centric Modeling and Analytics*, 324-361.