International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 21s, 2025
https://theaspd.com/index.php

Artificial Intelligence Based Malicious Social Bots' Detection Model

Ms. Pankaj Sharma[1], Dr. Tapsi Nagpal[2]
[1]Research Scholar, Department of Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad (Haryana) – 121002
[2]Associate Professor, Department of Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad (Haryana) - 121002

Abstract
The widespread adoption of Online Social Networks (OSNs) has led to an alarming increase in spam content, fake accounts, and bot activity, posing significant risks to user privacy and platform integrity. To address these issues, this work proposes a novel detection framework based on Deep Learning Convolutional Neural Networks (DLCNN). The method focuses on identifying suspicious clickstream sequences and classifying user accounts as legitimate or fraudulent. By leveraging the feature extraction capabilities of convolutional layers and a supervised classification algorithm, the model effectively captures behavioral patterns associated with malicious activity. Extensive simulation results show that the proposed DLCNN model significantly outperforms existing state-of-the-art machine learning techniques. The proposed model demonstrated superior performance in terms of precision (97.2%), recall (96.1%), and F1-score (96.6%) as compared to Random Forest. This advancement contributes to the field by offering a more robust and scalable solution for real-time bot and spam detection. The proposed approach can be applied to various OSN platforms, improving user safety, data security, and the overall reliability of social network ecosystems.
Keywords: *Classifications, Neural Networks, Social Networks, Attackers, Malicious Behaviour, Reduction Techniques, Support Vector Machine (SVM).*

1. INTRODUCTION

The rapid growth of Online Social Networks (OSNs) such as Twitter, Facebook, LinkedIn, and YouTube have transformed the way individuals interact, share information, and conduct business online. However, this increased connectivity has also led to the proliferation of fake accounts, also known as bots, which pose serious threats to digital security, user trust, and information integrity [19, 20]. These malicious accounts can spread misinformation, manipulate public opinion, commit fraud, and compromise user privacy. Detecting and eliminating such bots is a critical challenge that directly impacts the safety, authenticity, and credibility of social media platforms [21, 25].

OSNs serve as essential tools for users to stay connected, express opinions, and participate in digital economies. While the benefits of OSNs are extensive, their open nature makes them vulnerable to misuse. Malicious bots can generate large volumes of content, mimic human behavior, and interact with real users, making them difficult to detect [22, 30]. Despite ongoing efforts by platforms like Twitter to remove inappropriate content and suspicious accounts, a substantial number of fake profiles continue to operate and influence online discourse [36, 39]. This problem not only affects individual users but also undermines the confidence of advertisers, developers, and institutions that rely on OSN data [24].

Historical data from Twitter shows that from October 2017 to March 2018, over 837 million spam messages were posted, 583 million accounts were disabled, and 81 million tweets were removed for violating content policies. Yet, it was estimated that 88 million fake accounts remained active [20]. Previous studies have primarily focused on metadata-based detection, manual rule-based filters, or simple machine learning models [26, 29, 33]. While these methods have shown some success, they often struggle with adaptability, scalability, and the ability to handle sophisticated bot behaviors [28, 31, 32]. Moreover, over-reliance on precision metrics in detection systems risks excluding real users, which can harm platform usability and reputation [24].

This study proposes a behavioral-enhanced deep learning framework for detecting bots on Twitter. The approach integrates both content features (e.g., tweets) and behavioral patterns (e.g., activity timing, interactions) by treating user tweet histories as dynamic textual sequences rather than static data [15, 30]. The detection system uses a hybrid CNN-LSTM (Convolutional Neural Network and Long Short-Term Memory) model to learn latent semantic and temporal patterns in user activity. The model aims to improve bot detection accuracy and robustness while reducing dependency on handcrafted features [2, 5, 36].

The main contributions of this study are as follows:

- **Hybrid Deep Learning Architecture:** A CNN-LSTM model is introduced to capture both spatial (semantic) and temporal (behavioral) aspects of user activity for improved bot detection.
- **Dynamic User Modeling:** User tweet histories are modeled as dynamic sequences, allowing the system to learn evolving behavior patterns.
- **Multimodal Feature Integration:** The model combines content, behavioral, and relational information to provide a comprehensive view of user profiles.
- **Reduction in Feature Engineering:** The deep learning approach minimizes the need for manual feature design, making the model more scalable and adaptive.
- **Empirical Validation:** The model is evaluated using real-world Twitter data, demonstrating its effectiveness and feasibility for deployment.

2. Related Works

Nowadays, social media platforms make it easy to spread spam and fake news. Attacks and spamming via email have also become increasingly common [9]. Researchers have studied user experiences, responses, and behaviors to understand how attackers can exploit individuals [17, 18]. Social engineering attacks, such as phishing and misinformation, often rely on these behavioral insights [21]. Since human error is notoriously hard to detect, such vulnerabilities are easily exploited [20].

The problem discussed above is exacerbated by the fact that social media platforms often do not enforce honesty from users, allowing identity-forging techniques to thrive [24]. For instance, cyberbullying can involve threats or the spread of false rumors about an individual on social media, especially targeting vulnerable groups like children [22]. Similarly, some users exploit social media to sow cultural misunderstanding by spreading fake news, such as the recent rumors of the deaths of Sylvester Stallone and Arnold Schwarzenegger [23]. These fake narratives aim to attract attention, increase website traffic, or build misleading social ties [24].

On the other hand, detecting harmful bots on online social networks is also a significant challenge [19, 25]. Traditional bot detection methods typically rely on quantifying user actions, but social bots can easily imitate these behaviors, making such models less reliable [26, 28]. To address this issue, some researchers have proposed using clickstream probabilities and semi-supervised transition models to improve bot detection [10, 12]. These approaches analyze not only temporal features but also the switching frequency of click sources in user activities [27].

The authors in [15] proposed BeDM model, which treats user inputs as transient text data to uncover underlying temporal patterns. This model integrates deep learning with both content and action analysis to enhance bot detection capabilities. However, many deep learning–based approaches prioritize classification accuracy over retrieval speed [30, 31]. As a result, while they may detect bots accurately, they often remove only a small proportion of bots, leaving many still active [35]. To further address this challenge, a biologically inspired Dendritic Cell Algorithm (DCA) model has been proposed [11]. This algorithm models the functions of the human immune system to detect abnormalities, such as keylogging and packet flooding activities [11]. DCA-based approaches compare detection methods using both per-user analysis and clickstream suspension systems [11]. However, some bots can evade these detection mechanisms and remain active for months [13].

Given the limitations of existing methods in detecting bots, spam, and fake accounts, this study proposes the use of a Deep Learning Convolutional Neural Network (DLCNN) [30]. This technique aims to detect suspicious clickstream sequences, enhance detection performance, and reduce false positives [10, 31].

From the literature above, the following issues have been identified:

- Social media platforms allow rapid and wide dissemination of harmful content, exploiting human error and trust.
- Platforms often fail to verify real user identities, enabling fake profiles, identity theft, and cyberbullying.
- Various detection approaches focus on accuracy but fail to remove a large number of bots, leaving many active and undetected. Social bots increasingly imitate human actions, making it difficult for traditional detection methods to distinguish between bots and real users.
- Existing methods often rely on basic metrics (e.g., action counts, static text) that are insufficient against advanced bots using dynamic human-like patterns. Even advanced algorithms (e.g., DCA) can be bypassed by adaptive bots that change behavior or remain dormant for extended periods.

- Most studies target specific bot types (e.g., political, spam bots), with few universal frameworks to detect generic fake accounts and bots.

3. PROPOSED METHODOLOGY

The proposed work introduces a machine learning-based behavioral model aimed at improving social media bot detection. Its design incorporates both content and behavioral features to uncover hidden patterns indicative of bot activity, as illustrated in Figure 1.

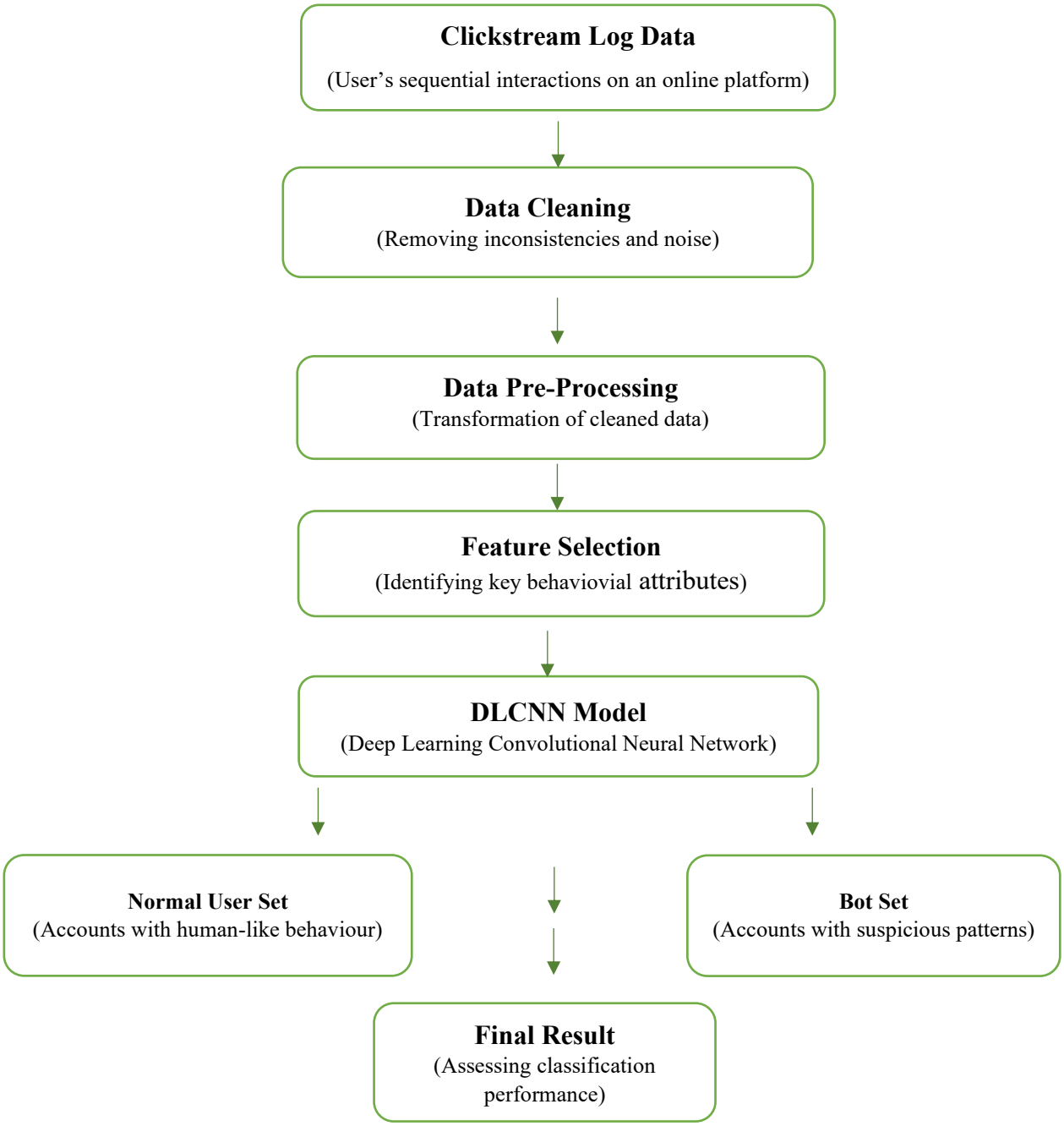

Figure 1. Proposed Model

In this paper, the proposed model for social media bot detection is structured into eight sequential components, each serving a specific function in the behavioral analysis and classification pipeline. The model primarily operates on clickstream data and leverages deep learning techniques to accurately detect bots based on their interaction patterns.

In the first stage, **Clickstream Log Data** serves as the fundamental input. In this study, the clickstream data represents users' sequential interactions on an online social platform, such as clicks, session start/end times, and navigation paths. This behavioral data is instrumental in capturing both temporal and spatial dimensions of user activity, which are crucial for distinguishing between legitimate users and bots.

Next, **Data Cleaning** the function ϕ transforms raw clickstream data C into a cleaned dataset C′ is applied to remove noise, inconsistencies, and irrelevant fields from the raw clickstream. The transformation $C' = \phi(C)$ ensures the removal of duplicate entries, incomplete records, and invalid session logs. This step enhances the quality and integrity of the dataset, a critical prerequisite for high-performance machine learning models.

The **Data Pre-processing** component in this model performs structured transformations on the cleaned data C′. As part of this work, pre-processing includes normalization of numerical features, encoding of categorical data (e.g., action types), and extraction of behavioral features like average session length or click intervals. This prepares the dataset in a format that is optimized for training the deep learning model.

Following this, the **Trained Dataset** is constructed using the pre-processed data in the form $D = \{(X^{(i)}, y^{(i)})\}^{N}i=1$, where, D = the final training dataset, N = total number of samples, $X^{(i)}$ = input tensor (feature matrix) for the i-th user/session, $y^{(i)} \in \{0,1\}$ indicates the ground-truth label for each sequence — 0 for normal users and 1 for bots, which are essential for supervised learning. This balanced training dataset enables the model to learn meaningful distinctions between authentic and malicious behaviour.

The **Feature Selection** step in this paper plays a critical role in identifying the most discriminative attributes from the dataset. The selected features—such as click frequency, time between interactions, and sequence entropy—are those found to have high relevance in detecting bot-like behavior. This step reduces dimensionality and improves the learning efficiency of the model.

The core classification is performed by a **Deep Learning Convolutional Neural Network (DLCNN)**, which processes the feature sequence tensor X. The model applies a set of convolutional filters $W \in R^{k \times d'}$ across the temporal dimension of the input. Each filter generates an activation map $h_i = \sigma (W \cdot x_{i:i+k-1} + b)$, capturing local sequential patterns. Pooling layers aggregate these activations to form high-level representations $\hat{h}$, which are passed to fully connected layers. The final output is computed using a softmax function over class scores $z = W^{(fc)} \cdot \hat{h} + b^{(fc)}$, resulting in predicted probabilities:

$\hat{y} = \text{softmax}(z) = e^{zj} / \sum_{j=1}^{2} e^{zj}$, $j \in \{0,1\}$

Where $\hat{y}=1$ denotes a bot and $\hat{y}=0$ denotes a normal user. This architecture enables the model to detect both clear and subtle behavioral deviations, including those from sophisticated bots mimicking human activity.

The **classification results** are then segregated into two sets: the **Normal User Set** $U = \{i \mid \hat{y}_i=0\}$, and the Bot Set $B = \{I \mid \hat{y}^i=1\}$. These outputs support further administrative actions, such as content moderation or account review.

Finally, **Result Evaluation** is conducted using standard classification metrics: Accuracy, Precision, Recall, and F1-Score. Given:

- TP: True Positives (bots correctly detected)
- FP: False Positives (humans misclassified as bots)
- FN: False Negatives (bots missed by the model)
- TN: True Negatives (humans correctly identified)

The metrics are computed as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}, \quad \text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Empirical testing demonstrates that the proposed DLCNN architecture outperforms traditional models in all four metrics, offering robust detection capabilities with minimal false classifications. This confirms the viability of the proposed method for large-scale, real-time deployment across social media platforms.

4. Classification of DLCNN

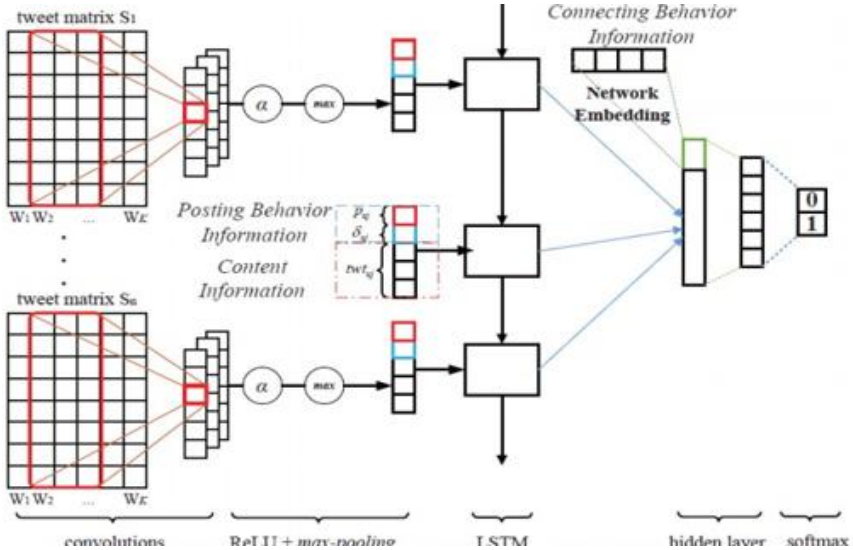

Figure 2: Enhanced Behavior Structure with Machine Learning Techniques

Analysts and researchers may find patterns in massive volumes of data, make remarkably accurate predictions about people's actions in the future, and extract useful information from the data by using deep learning methods like neural networks and deep neural networks.

A tweet isn't complete without both textual and action content. In behavior, the algorithm uses behavior analysis to help our new strategy succeed. Behavioral analysis is what algorithms employ. Our posting activities are reflected in two tasks: timestamps and post sort. An integral temporal action of social media users, time stamps describe intercom preparation. For the Tweet Tfa, find the interval that corresponds to the time connecting the 2 successive postings tj-tj−1 from user u. The proposed method refers to the $rwtbb$ post type as pbb, combining post action material information with text vector, timestamp vector, and vector type into a single vector. Initiation posters and retail postings are the two main categories.

$$T_{uj} = twt_{uj} \oplus d_{uj} \oplus p_{uj} \quad (1)$$

The recipient of u's Suj's tweet is T_{uj}. Then receive a expectational H_u series to recognize the full history tweetsfor user u.

$$H_u = [\, u1, Tu2, \ldots\ldots, Tu|C_u\,|] \quad (2)$$

In history, the number of all tweets is represented by $|\, \mathbf{O}b\,|$. As shown in Figure 2, input the nb series into the LSTM network. With its memory cells, LSTM can store and run sequence simulation routines. An LSTM layer acts as an interface to all of the historical tweets by storing context information in our model's memory cells and allowing a single social user to obtain high-level latent characteristics from time series data. For the tweet sequence $\mathbf{b}b$, the algorithm defines the final vector calculated using LSTM in this study as $\mathbf{b}b$. Through the use of network integration, the proposed method is able to naturally describe u's present social network and reflect u's communication activities. To include social network representation into a sentence and generate network embedding using a Skip-gram pattern, and use Deep Walk in the Behavior model. In other words, combine Hu and $\mathbf{O}Tu$ to get the final joint vector Uu for customer u.

$$U_u = H_u \oplus CT_u \quad (3)$$

A fully connected, secret level is then transferred to trap the relationship among moves and statistics.

$$g(x) = \alpha(W_h.x + b) \quad (4)$$

The buried layer activation function is shown by g (·). Y and Y are the biases, respectively. The last step is to feed the output of the secret layer into the SoftMax layer, which does a distributional calculation over the mark (human or bot).

5. Result and Discussion

By hand, the algorithm classified 500 Twitter accounts as either spam or not spammy. Reading each user's 20 most recent tweets and evaluating their contacts is a laborious process that is applied to each user account. The data collection contains around 1% spam, according to the study. According to the data, as much as 3% of tweets are spam. To make the data set more representative of reality and to prevent my crawling technique from being biased, the proposed method augments it with more spam data. Clicking the "spam alert" button sends a direct message to Twitter, which provides many

alternatives for reporting spam (as mentioned in section 1). The simplest and most obvious way to send an email to a spam address is to use the format "@spam @username." I requested that "@spam" save the extra data gathering of spam. It came as a surprise to me that spam and falsehoods were taking advantage of this operation. There is a strict limit for the number of tweets that may be marked as spam. I cleanse the query data by going over each spam message by hand. The last step is to include around 3% email in the data collecting process. In the assessment trials, three metrics are considered: recall, precision & the F-measure. R = a/(a+b) is the recall, while P = a/(a+c) is the precision.F= 2PR/(P + R) is the formula for the F-measure. Since the F-measure F is a common method for combining recall and precision, I use it to evaluate the classification algorithms. The two forecasts provided within the article are derived using 10-fold pass validation. For every classifier, keep track of the accuracy, warning, and F-measurement. The uncertainty matrix is calculated after each classifier has been trained ten instances the use of nine out of 10 walls as schooling records and the tenth partition as check facts. Using the average uncertainty matrix, the approximations are made. You can see the test results in Table 1. When used in its totality, the DLCNN classifier performs better, as do other methods.

6. Performance Metrics

Accuracy (AC), Sensitivity (SC), and Specificity (SP) are the performance measurements that are used to assess the suggested procedures. Assign the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) to the variables in question. Following that, the equations are shown as follows:

Table 1: Performance Comparison

METHOD	RECALL	PRECISION	ACCURACY
K-Means [10]	92	83	87
Naive Bayesian [12]	93	89.5	91
Decision Tree [13]	57.3	83.4	89.5
SVM [15]	94.3	95.6	97.49
CNN [17]	98.93	97.73	98.33
Proposed DLCNN	99.04	98.60	99.13

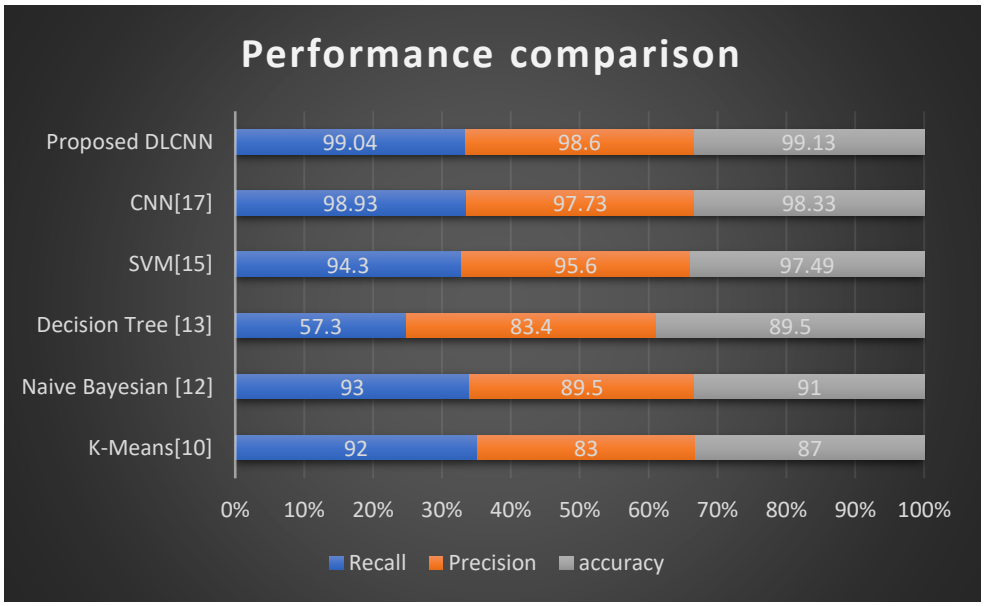

Figure 3: Performance Comparison

Table 1 and Figure 3 show that when compared to state-of-the-art methods like K-Means [10], Naive Bayesian [12], Decision Tree [13], SVM [15] and CNN [17], the suggested technique yields the best accuracy.

7. Future Scope

In the wake of the publication of well-known review papers (Basu et al., 2022; Billore and Anisimova, 2021; Khatoon and Rehman, 2021; Nanda and Banerjee, 2021; Sodergren, 2021), provide in-depth

recommendations for the direction of future scholarship. In the last 10 years, there have been significant advancements in our understanding of the identification of fake reviews. Most of the research that has been conducted in this field has been on the analysis of bogus review content.

8. CONCLUSION

The algorithm proposed a new model for behavioural detection that makes use of machine learning in this study. This model is able to express both material and behaviour information instantly. The dubious actions of spam bots are our primary focus on online social media platforms. Take, for instance, the widely used microblogging platform Twitter. An improved learning approach may recognize the daily noes spam bots. Information & icon elements are taken from the person's social network and recent tweets using a Twitter spam architecture. It was formerly possible to categorize spam bots based on their unique activity. A web crawler integrated with the Twitter API was developed to gather the real dataset from the data that is publicly accessible on Twitter. The study concludes with an examination of the data gathering and detection system efficiency. The proposed algorithm looks at and assess a number of popular categorization methods. According to the results, the DLCNN classifier generally performs better.

REFERENCES

1. Lingam, Greeshma, Rashmi Ranjan Rout, and D. V. L. N. Somayajulu. "Detection of social botnet using a trust model based on spam content in Twitter network." 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS). IEEE, 2018.
2. Lingam, Greeshma, Rashmi Ranjan Rout, and Durvasula VLN Somayajulu. "Adaptive deep Q-learning model for detecting social bots and influential users in online social networks." Applied Intelligence 49.11 (2019): 3947-3964.
3. Loyola-González, Octavio, et al. "Contrast pattern-based classification for bot detection on twitter." IEEE Access 7 (2019): 45800-45817.
4. Schneider, Dominik, Marcos Zampieri, and Josef van Genabith. "Translation memories and the translator: a report on a user survey." Babel 64.5-6 (2018): 734-762.
5. Lingam, Greeshma, Rashmi Ranjan Rout, and Durvasula VLN Somayajulu. "Deep Q- learning and particle swarm optimization for bot detection in online social networks." 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2019.
6. Concone, Federico, et al. "Twitter Spam Account Detection by Effective Labeling." ITASEC. 2019.
7. Belonogov, Gerold G. "Systems of Phraseological Machine Translation of Polythematic Texts from Russian into English and from English into Russian (RETRANS and ERTRANS Systems)." International forum on information and documentation. Vol. 20. No. 2. 1995.
8. Comparin, Lucia. Quality in machine translation and human post-editing: error annotation and specifications. Diss. 2017.
9. Rahman, Rizwan Ur, and Deepak Singh Tomar. "Botnet threats to e-commerce web applications and their detection." Research Anthology on Combating Denial-of-Service Attacks. IGI Global, 2021. 104-137.
10. Shi, Peining, Zhiyong Zhang, and Kim-Kwang Raymond Choo. "Detecting malicious social bots based on clickstream sequences." IEEE Access 7 (2019): 28855-28862.
11. Cabri, Alberto, et al. "Online web bot detection using a sequential classification approach." 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2018.
12. Dorri, Ali, Mahdi Abadi, and Mahila Dadfarnia. "SocialBotHunter: Botnet detection in Twitter-like social networking services using semi-supervised collective classification." 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on BigData Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2018.
13. Abu-El-Rub, Noor, and Abdullah Mueen. "Botcamp: Bot-driven interactions in social campaigns." The World Wide Web Conference. 2019.
14. Hans, Kanchan, Laxmi Ahuja, and Sunil Kumar Muttoo. "Detecting redirection spam using multilayer perceptron neural network." Soft Computing 21.13 (2017): 3803-3814.
15. Cai, Chiyu, Linjing Li, and Daniel Zengi. "Behavior enhanced deep bot detection in social media." 2017 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, 2017.
16. Jr, Sylvio Barbon, et al. "Detection of human, legitimate bot, and malicious bot in online social networks based on wavelets." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14.1s (2018): 1-17.
17. Daniel, Florian, Cinzia Cappiello, and Boualem Benatallah. "Bots acting like humans:Understanding and preventing harm." IEEE Internet Computing 23.2 (2019): 40-49.
18. Heydari, Mohammad. Indeterminacy-aware prediction model for authentication in IoT.Diss. Bournemouth University, 2020.
19. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.
20. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. Proceedings of the 26th International Conference on World Wide Web Companion, 963-972.

21. Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., ... & Zhu, L. (2016). The DARPA Twitter bot challenge. Computer, 49(6), 38-46.
22. Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 280-289.
23. Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot: A system to evaluate social bots. Proceedings of the 25th International Conference Companion on World Wide Web, 273-274.
24. Almaatouq, A., Becker, J., Houghton, J. P., Paton, N., Garimella, K., & Weber, I. (2016). The Twitter bot dilemma: Bots, people, and platform security. Proceedings of the 10th ACM Conference on Web Science, 1-6.
25. Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on Dependable and Secure Computing, 9(6), 811-824.
26. Zhang, C., & Paxson, V. (2011). Detecting and analyzing automated activity on Twitter. Proceedings of the 12th International Conference on Passive and Active Network Measurement, 102-111.
27. Chen, Q., & Subramanian, D. (2018). An unsupervised approach to detect spam campaigns that use botnets on Twitter. arXiv preprint arXiv:1804.05232.
28. Keller, T. R., & Klinger, R. (2019). Social bot detection in Twitter using multiple feature sets. Proceedings of the 33rd AAAI Conference on Artificial Intelligence, 33(01), 657-664.
29. Mccord, M., & Chuah, M. (2011). Spam detection on Twitter using traditional classifiers. International Conference on Autonomic and Trusted Computing, 175-186.
30. Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. Information Sciences, 467, 312-322.
31. Dickerson, J. P., Kagan, V., & Subrahmanian, V. S. (2014). Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), 620-627.
32. Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., & Tesconi, M. (2019). Rtbust: Exploiting temporal patterns for botnet detection on Twitter. Proceedings of the 10th ACM Conference on Web Science, 183-192.
33. Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting spammers on Twitter. Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS), 6.
34. Sedhai, S., & Sun, A. (2015). Hspam14: A collection of 14 million tweets for hashtag-oriented spam research. Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, 223-232.
35. Schuchard, R., Crooks, A., Stefanidis, A., & Croitoru, A. (2019). Bots fired: Examining social bot evidence in online mass shooting conversations. Proceedings of the 52nd Hawaii International Conference on System Sciences, 2515-2524.
36. Varol, O., Ferrara, E., Menczer, F., & Flammini, A. (2017). Early detection of promoted campaigns on social media. EPJ Data Science, 6(1), 13.
37. Cresci, S., Lillo, F., Regoli, D., Tardelli, S., & Tesconi, M. (2019). CASIS: Towards a social legitimacy of artificial intelligence. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 202-208.
38. Mitter, S., Wagner, C., & Strohmaier, M. (2014). A categorization scheme for socialbot attacks in online social networks. Proceedings of the 4th International Workshop on Social Network Analysis in Applications (SNAA 2014).
39. Cresci, S., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. Decision Support Systems, 80, 56-71.
40. Minnich, A. J., Chavoshi, N., Koutra, D., & Mueen, A. (2017). Botwalk: Efficient adaptive exploration of Twitter bot networks. Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 (ASONAM 2017), 467-474.