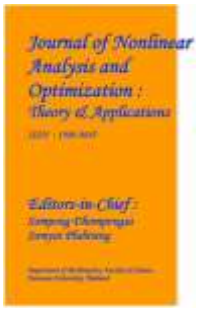
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**



MACHINE LEARNING BASED SOCIAL BOT DETECTION WITH TRANSFORMER - BASED CLASSIFICATION

K. JAYA KRISHNA[1], M.VENKATA SURESH BABU[2]

[1]Associate Professor, Dept. of MCA, QIS College of Engineering and Technology, Ongole, Andhra Pradesh.

[2]PG Scholar, Dept. of MCA, QIS College of Engineering and Technology, Ongole, Andhra Pradesh.

ABSTRACT— In recent years, the proliferation of online communication platforms and social media has given rise to a new wave of challenges, including the rapid spread of malicious bots. These bots, often programmed to impersonate human users, can infiltrate online communities, disseminate misinformation, and engage in various activities detrimental to the integrity of digital discourse. It is becoming more and more difficult to discern a text produced by deep neural networks from that created by humans. Transformer based Pre-trained Language Models (PLMs) have recently shown excellent results in challenges involving natural language understanding (NLU). The suggested method is to employ an approach to detect bots at the tweet level by utilizing content and fine-tuning PLMs, to reduce the current threat. Building on the recent developments of the BERT (Bidirectional Encoder Representations from Transformers) and GPT-3, the suggested model employs a text embedding approach. This method offers a high-quality representation that can enhance the efficacy of detection. In addition, a Feedforward Neural Network (FNN) was used on top of the PLMs for final classification. The model was experimentally evaluated using the Twitter bot dataset.

Index Terms— Online social networks, NLP, transfer learning, Bot classification, transformers.

I. INTRODUCTION

Since bots pose a greater threat to free will and opinion, the problem of bot identification on Twitter has grown in interest as the social network becomes more widely used. Due to the rising complexity of bots, which goes beyond the traditional study of individual features, researchers have concluded that it is necessary to classify them at a behavioral level on the platform. Numerous individuals worldwide now rely heavily on social media platforms as their primary source of information. Twitter is widely recognized as the most renowned microblogging platform. The usage of bots (@HundredZeros or @TayTweets) is one example of social media manipulation. User accounts that are managed by software algorithms instead of human users are commonly referred to as bot accounts. Bots are programmed to perform specific tasks or actions on digital platforms, often to automate processes or provide certain functionalities. These accounts can interact with other users, generate content, or perform actions based on predefined instructions. The sophistication of current social media bots ranges widely; some are relatively basic, primarily engaging in retweeting content they find interesting, whilst others are more complex and may communicate with actual users. Twitter is one of these platforms that facilitates the rapid dissemination of information throughout its user community. In the field of text generation, the latest neural language models have reached a state where their output is remarkably grammatically accurate, fluent, and coherent. As a result, it has become challenging to differentiate between text generated by these models and text written by humans. Consequently, there is a growing need to explore the effectiveness of existing detection methods proposed in the research literature to distinguish human-generated text from text generated by neural language models. This is especially crucial due to emerging evidence suggesting that humans find this task exceedingly difficult. Twitter postings are character-restricted writings that are limited to 280 characters. Because short messages authored by bots are harder to discern from human-generated texts than lengthier texts, this format is perfect for text-generating algorithms. There has been a significant increase in academic curiosity and study given to identifying and detecting social media bots in recent years.

The growing engagement and resultant effect of these automated accounts on numerous social media platforms are driving this increased emphasis. According to statistics

published in March 2023 the most recent statistics from an internal study of Twitter bot percentages revealed that fewer than 5% of its users are fraudulent or spam bots. The objective is to create a robust model capable of producing cutting-edge bot detection findings. We investigated various standard word embedding approaches in this context, including Word2Vec and Global Vectors for Word Representation (Glove). In this study, we explore and empirically assess the performance of pre-trained word embeddings and language models tailored for text analysis from social media.

II. LITERATURE SURVEY

A. Contextual string embeddings for sequence labelling

Recent advances in language modeling using recurrent neural networks have made it viable to model language as distributions over characters. By learning to predict the next character on the basis of previous characters, such models have been shown to automatically internalize linguistic concepts such as words, sentences, subclauses and even sentiment. In this paper, we propose to leverage the internal states of a trained character language model to produce a novel type of word embedding which we refer to as contextual string embeddings. Our proposed embeddings have the distinct properties that they (a) are trained without any explicit notion of words and thus fundamentally model words as sequences of characters, and (b) are contextualized by their surrounding text, meaning that the same word will have different embeddings depending on its contextual use. We conduct a comparative evaluation against previous embeddings and find that our embeddings are highly useful for downstream tasks: across four classic sequence labeling tasks we consistently outperform the previous state-of-the-art. In particular, we significantly outperform previous work on English and German named entity recognition (NER), allowing us to report new state-of-the-art F1-scores on the CoNLL03 shared task.

B. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information

Sentiment analysis on social media such as Twitter has become a very important and challenging task. Due to the characteristics of such data tweet length, spelling errors, abbreviations, and special characters the sentiment analysis task in such an environment requires a non-traditional approach. Moreover, social media sentiment

analysis is a fundamental problem with many interesting applications.

Most current social media sentiment classification methods judge the sentiment polarity primarily according to textual content and neglect other information on these platforms. In this paper, we propose a neural network model that also incorporates user behavioral information within a given document (tweet).

C. Bot prediction on social networks of Twitter in altmetrics using deep graph convolutional networks

In the context of smart cities, it is crucial to filter out falsified information spread on social media channels through paid campaigns or bot-user accounts that significantly influence communication networks across the social communities and may affect smart decision-making by the citizens. In this paper, we focus on two major aspects of the Twitter social network associated with altmetrics: (a) to analyze the properties of bots on Twitter networks and (b) to distinguish between bots and human accounts. Firstly, we employed state-of-the-art social network analysis techniques that exploit Twitter's social network properties in novel altmetrics data.

III. PROPOSED SYSTEM

The overview of our proposed system is shown in the below figure.

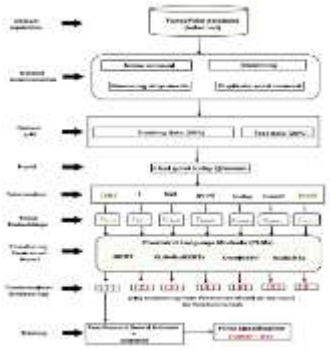


Fig. 1: System Overview

Implementation Modules

Service Provider:

- In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train and Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy, View All Remote Users.

Remote User:

- In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name

and password. Once Login is successful user will do some operations like, Predict Tweet account Type, View Your Profile.

Train and Test Model

- In this module, the service provider split the Used dataset into train and test data of ratio 70 % and 30 % respectively. The 70% of the data is consider as train data which is used to train the model and 30% of the data is consider as test which is used to test the model.

IV. RESULTS

Fig.4: Models Accuracy

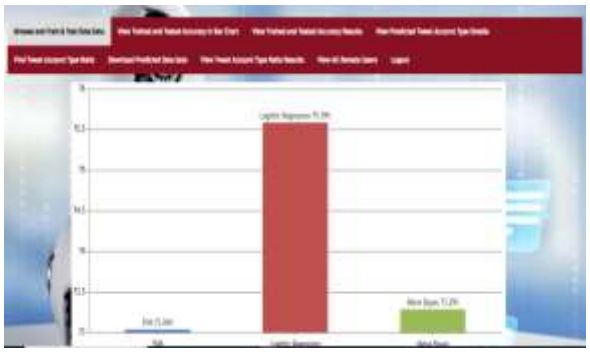


Fig.5: Models Accuracy Result

Fig.6: Models accuracy results in Line Chart.

V. CONCLUSION

In this project, a robust solution for detecting Bots in Twittera ccounts has been described. In particular, this study has taken advantage of Transfer learning techniques via powerful state-of-the-art NLP models such as Transformers to extract compact multilingual representations of the text-based features associated with user accounts. By doing so, several constraints presented in previous studies related to process text-based features to improve the input feature vector from multiple languages were mitigated. Furthermore, by employing the text encodings along with additional metadata on top of a dense-based neural network, a final classifier named as Bot-DenseNet has been trained and validated using a large set of samples collected via the Twitter API. More

specifically, several experiments were conducted using different combinations of Word Embeddings, document embeddings (Pooling and LSTMs) and Transformers to obtain a single vector regarding the text-based features of the user account. Subsequently, a detailed comparison of the performance of the proposed classifier when using these approaches of Language Models as part of the input has been presented to investigate which input vector provides the best result in terms of performance simplicity in the generation of decision boundaries and feasibility.

REFERENCE

[1] Ž. Agić and I. Vulić, ‘‘JW300: A wide-coverage parallel corpus for lowresource languages,’’ in Proc. 57th Annu. Meeting Assoc. Comput.Linguistics. Florence, Italy: Association for Computational Linguistics, Jul. 2019,pp. 3204–3210.

[2] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and A. Vollgraf,‘‘FLAIR: An easy-to-use framework for state-of-the-art NLP,’’ in Proc.Conf. North Amer. Chapter Assoc. Comput. Linguistics, Demonstrations,2019, pp. 54–59.

[3] M. A. Shaik, A. Kethireddy, S. Nerella, S. Pinninti, V. Kathare and P. Pitta, “Sound Wave Scribe: Bridging Spoken Language and Written Text”, 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT), Delhi, India, 2024, pp. 413-417, doi: 10.1109/IC2SDT62152.2024.10696694.

[4] A. S. M. Alharbi and E. de Doncker, ‘‘Twitter sentiment analysis witha deep neural network: An enhanced approach using user behavioural information,’’ Cogn. Syst. Res., vol. 54, pp. 50–61, May 2019.

[5] M. A. Shaik and N. L. Sri, “A Comparison of Stock Price Prediction Using Machine Learning Techniques”, 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2024, pp. 1-5, doi: 10.1109/ICESC60852.2024.10689767.

[6] M. Arora and V. Kansal, ‘‘Character level embedding with deep convolutional neural network for text normalization of unstructured data forTwitter sentiment analysis,’’ Social Netw.Anal. Mining, vol. 9, no. 1, p. 12,Dec. 2019.

[7] A. Balestrucci, R. De Nicola, O. Inverso, and C. Trubiani, ‘‘Identificationof credulous users on Twitter,’’ in Proc.

34th ACM/SIGAPP Symp. Appl.Comput., Apr. 2019, pp. 2096–2103.

[8] M. A. Shaik, M. Parveen and I. Qureshi, "Leveraging Machine Learning and Drone Technology for Effective Insect Pest Management in Agriculture", 2024 Asia Pacific Conference on Innovation in Technology (APCIT), MYSORE, India, 2024, pp. 1-8, doi: 10.1109/APCIT62007.2024.10673597.

[9] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, ''Detecting automation ofTwitter accounts: Are you a human, bot, or cyborg?'' IEEE Trans. Depend.Sec. Comput., vol. 9, no. 6, pp. 811–824, Nov./Dec. 2012.

[10] M. A. Shaik, Y. Sahithi, M. Nishitha, R. Reethika, K. S. Teja and C. P. Reddy, "Realtime Emotion Recognition from Images to Understand Facial Expressions", 2024 Asia Pacific Conference on Innovation in Technology (APCIT), MYSORE, India, 2024, pp. 1-5, doi: 10.1109/APCIT62007.2024.10673486.

AUTHORS Profile

Mr. K. Jaya Krishna is an Associate Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai, and his M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). With a strong research background, he has authored and co-authored over 90 research papers published in reputed peer-reviewed Scopus-indexed journals. He has also actively presented his work at various national and international conferences, with several of his publications appearing in IEEE-indexed proceedings. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

M. Venkata Suresh Babu has received degree in B.sc computers Acharya Nagarjuna University 2023. Pursuing MCA at QIS College of Engineering and Technology affiliated to JNTUK in 2023-2025.