Journal of Next Generation Technology (ISSN: 2583-021X)
Vol. 4, Issue 2, May 2024

Recognition of Counterfeit Profiles on Communal Media using Machine Learning Artificial Neural Networks & Support Vector Machine Algorithms

Dr. Ranga Swamy Sirisati[1], A. Kalyani[2], V. Rupa[3],
Dr.Pradeep Venuthurumilli[4], Md Ameer Raza[5]
[1]Associate Professor, Department of CSE (AI&ML), [2]Assistant Professor, Department of CSE (DS), [3]Assistant Professor, Department of Information Technology,
[1,2,3]Vignan's Institute of Management and Technology for Women, Kondapur, Telangana
[4]Associate Professor, Department of CSE (DATA SCIENCE), Mallareddy Engineering College for Women, Maisammaguda, Secunderabad
[5]Assistant Professor, Department of CSE, Sri Vasavi Institute of Engineering and Technology, Nandamuru, Pedana, Andhra Pradesh

Email id : sirisatiranga@gmail.com[1], anipidikalyani2957@gmail.com[2] , rupa@vmtw.in[3] , pradeepvenuthuru@gmail.com[4], raza681@gmail.com[5]

Article Received: 08 Mar 2024 *Article Accepted: 1 May 2024* *Article Published: 30 May 2024*

Citation
Dr. Ranga Swamy Sirisati, A. Kalyani, V. Rupa, Dr.Pradeep Venuthurumilli, Md Ameer Raza (2024). "Recognition of Counterfeit Profiles on Communal Media using Machine Learning Artificial Neural Networks & Support Vector Machine Algorithms", Journal of Next Generation Technology (ISSN: 2583-021X), 4(2), pp. 19-27. May 2024.

Abstract

Internet users rely on social networks to help them with daily tasks including exchanging material, reading news, sending messages, reviewing products, and talking about events. Social media platforms also attract people who send different types of spam at the same time. These internet criminals include trolls, online fraudsters, sexual predators, and advocates for advertising. These people are fabricating profiles in order to disseminate their stuff and conduct con games. The consumers and the service providers are both at great risk from all of these fraudulent identities. Determine if accounts are real or fraudulent by identifying them from the social media service providers. We introduced several categorization algorithms in this paper, including neural networks and support vector machines. These formulas assist in to detect fake profiles.

Keywords: Social Media, Artificial Intelligence, Machine Learning, Artificial Neural Networks and Support Vector Machine.

I. INTRODUCTION

Every member of society in the current generation is now connected to social media. The way we pursue our social lives has drastically changed as a result of social media. In this paper, we will utilize Artificial Neural Networks to determine if the account data provided are from real or fraudulent individuals. An artificial neural network (ANN) using the SVM

technique will be trained on all of the real and false account data from prior users. If we are given fresh test data, the ANN train model will be applied to the new test data. It will be deciding whether the newly provided data for accountings are real or fraudulent [1][2].

• Social media these days has a hazardous effect on people's mental health.

• A lot of people are becoming vulnerable to false profiling.

• We consider fraudulent accounts to be our community of concern, and we believe that we have a categorization or clustering issue.

• Fake profiling detection can offer a better way to prevent such issues[3].[4].

In a social network, every profile (or account) has a lot of information, such as gender, the number of friends and comments, education, employment, and so on. others of this data is accessible to the public, while others is private[5].

Given that confidential data isn't available Therefore, in order to identify the phony profiles on the social network, we have only utilized publicly available information. But if the social networking businesses themselves utilize our suggested plan, they won't breach any privacy laws because they would use the private information from the profiles for detection.[6]. We have taken into account these details as profile attributes in order to distinguish between authentic and fraudulent profiles. The procedures we've used to identify fraudulent profiles [7][8]. The purpose of the paper is to determine the correctness of the profile and if it is real or fraudulent[9].

II. RELATED WORKS

Numerous pieces of input data, such as name, sexual orientation, friends, followers, interests, and area codes, are added to social media accounts online. Both public and private data make up half of this input[10]. Since private data is inaccessible, we must utilize public input to identify fake profiles for interpersonal organization[11]. A description of the predicted results for a classification task is called a confusion matrix [15–20]. Depend values are used to summarize the number of right and wrong forecasts, which are then broken down by each elegance. The confusion matrix's key is that[12]. The techniques by which your classification model becomes confused when generating predictions are shown by the confusion matrix[13][14].

III. METHODOLOGY

The Application Domain of the subsequent Paper was Community detection. Detection in the community is is vital to understand the structure of networking complexities with an ultimately extract information from them. During this Paper, a framework is used through which a fake profile is detected using a machine learning algorithm so that the social lifetime

of people become secured [21-27]. Fig 1 depicts the system to detect an the abstract of an overall process of the software system and the relationships, the constraints, and overall the boundaries between components. It is a crucial tool since it offers a comprehensive overview of the Fig. 2 software system's physical deployment as well as its future development plan. A user makes a friend request to an unknown individual in Figure 1, and that individual's complete data is saved in a database. An ANN classifier preprocesses and verifies the user's status from the database, and an admin uploads the data, after which the ANN classifier generates the result.

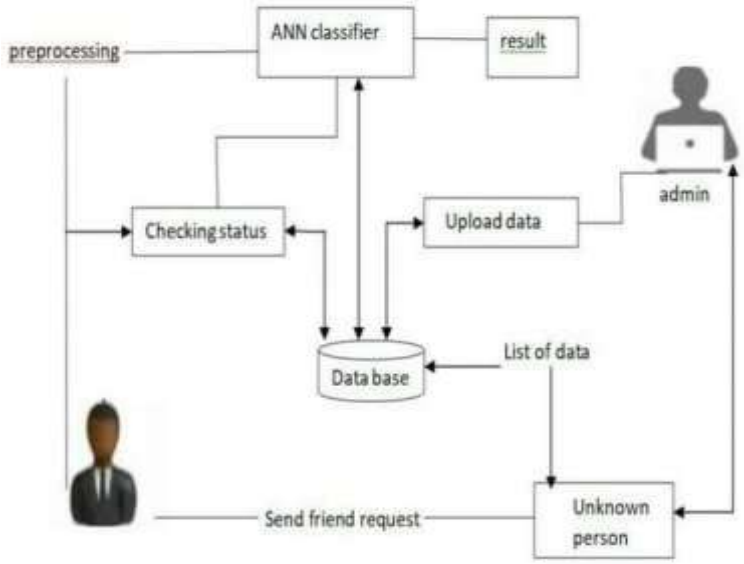

Figure 1: Architectural Diagram

Selecting the profile to be classed is the first step in the classification process. Upon selecting the profile, the pertinent elements are taken out with the intention of classifying. The trained classifier receives the extracted characteristics after that. After the classification algorithm's results are confirmed, the classifier receives feedback.

IV. PROCEDURE TO DETECT FAKE PROFILE USING NEURAL NETWORKS.

The following procedures have been taken into consideration for the use of neural network technology in the identification of false profiles on online social networking sites:

Step 1: First, the sys, csv, os, date/time, math, NumPy, pandas, and matlab libraries were imported.
Step 2: After this, gender detection libraries are being loaded to compute the information about the gender. Plotting the matrix has been done by integrating sklearn packages for preprocessing and data validation. The evaluation metric offers details on the various confusion matrix variables. Area under cover and accuracy have been used to assess the classifier.

Step 3: Next, import the Pybrain library so that the datasets may be trained. It is an open-sourced, publicly accessible library for machine learning algorithms. This library is implemented in conjunction with many utility tools.
Step 4: The next step is to define a method called read_datasets() in order to read the dataset. For this, comma separated value files, or CSV files, are utilized. The default for datasets to read must be set. It is necessary to determine the duration of users after merging the phony and real users.
Step 5: After that, a different function is built to determine a person's gender based on their given name. To compute the model, the person's first name is declassified into components. In addition, other related elements will be merged with status and follower counts, among other things.
Step 6: Plotting of the confusion matrix then starts, integrating the plot according to the phony and authentic profile accounts.
Step 7: The ROC curve definition function has been implemented for additional computing.
Step 8: A function is declared to train the dataset using the neural network. Read_datasets() has been used for this.
Step 9: The output will like this once the data has been read.
Step 10: The training datasets' characteristics are extracted.
Step 11: The confusion matrix graph is displayed without normalization.
Step 12: To define the fake and authentic profile precision index, recall, f1-score, and support vector, the reports have been classified.
Step 13: The experiment's final results, which define the real positive and false positive values, are as follows.

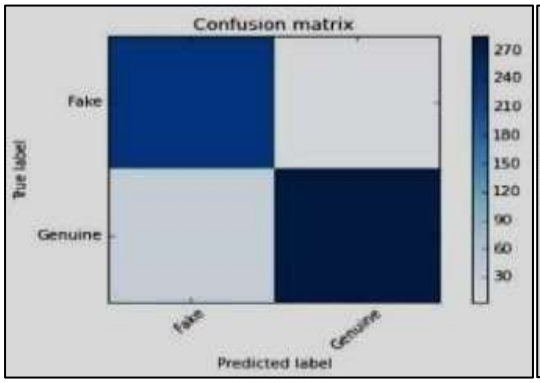

Figure 3: After doing the normalization of the confusion matrix.

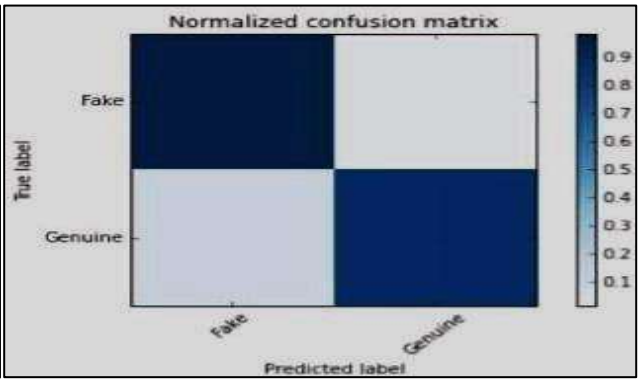

Figure 3: Normalized Confusion Matrix

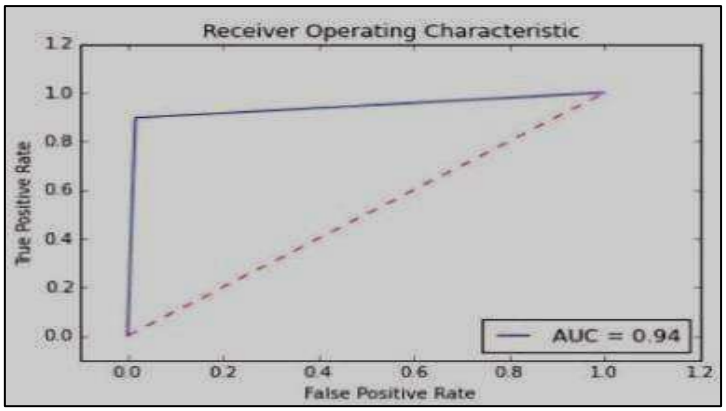

Figure 4: Receiver Operating Characteristic Curve

V. EXPERIMENTS AND RESULTS

Procedure to detect Fake Profile using Support Vector Machine

Step 1: A number of libraries, including MATLAB, sys, csv, datetime, and others, must be integrated in order to use SVM for the detection of phony profiles on social media sites. To read the CSV datasets and plot the matrix, these libraries are necessary.
stage 2: Reading datasets is part of the second stage. The files fusers.csv and users.csv are used in this instance to train and test the model, respectively. A Real users are been kept in the user.csv file, but fictitious or fraudulent users are kept in fusers.
Step 3: Create a function that uses the name provided in the dataset to retrieve gender information.
Step 4: We have been declared the function extract features for feature extraction.
Step 5: After that, we'll sketch the learning curve graphic, which will reveal details about certain characteristics that are employed to analyze the vectors.
Step 6: The confusion plot matrix linked to the profiles of Real and Fake users comes next. We'll also set the color value for the same plot.
Step 7: A function has been defined in order to plot the receiver operating characteristic, or ROC.
Step 8: The function with the name oftrain and the SVM classifier have been declared in order to train the dataset using a support vector machine.
Step 9: Reading and extracting the features from the dataset is the next step.
Step 10: After the characteristics are extracted, various values are displayed, such as the number of status updates, followers, listed, language code, and so forth for more
Step 11: The Splitting of the data-sets to train till Test.
Step 12: The training data of learning curve will be displayed in red color and the cross-validation in green color. The score data has been mentioned along with the training.
Step 13: Predictive labeling of confusion matrix is being performed before normalization.
Step 14: The ROC curve with True Positive and False Positive features is displayed last.

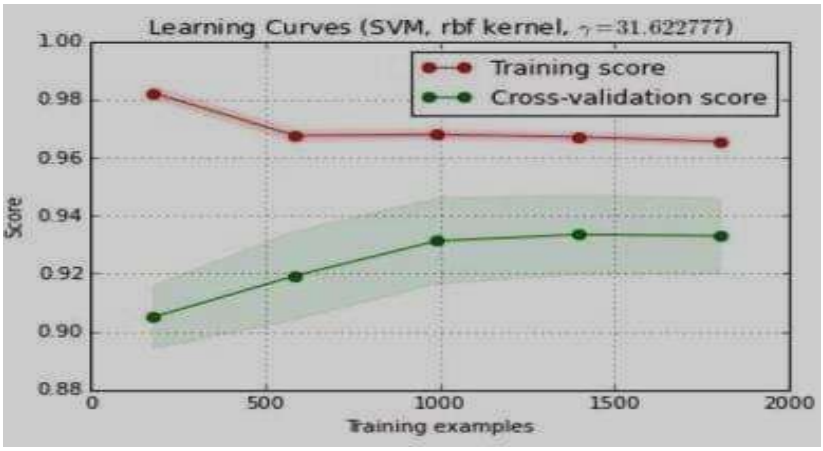

Figure 5: Training Examples

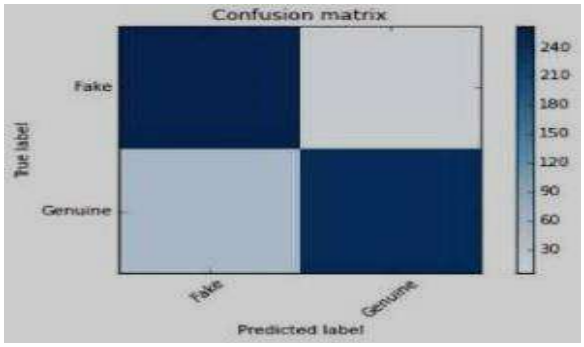

Figure 6: Confusion Matrix

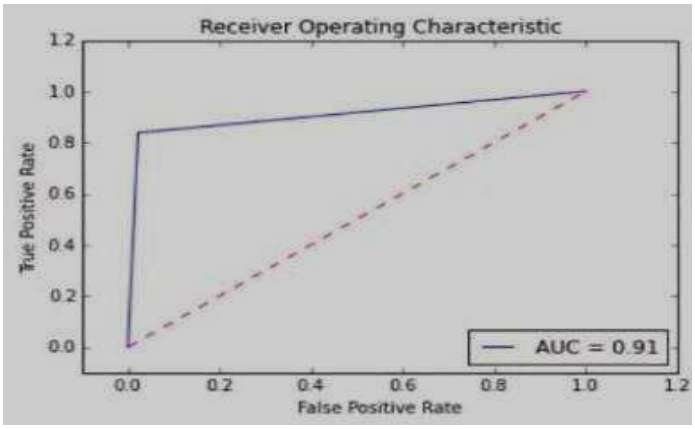

Figure 7: Receiver Operating Characteristic Curve

VI. CONCLUSIONS AND FUTURE SCOPE OF WORK

Social media is taking up in practically every industry, and businesses are using it as their primary platform to present their goods and services to potential clients. Public relations firms profit millions from social media by promoting information associated with various entities, such as political parties, celebrities, institutions, etc. Fake identities are used on social media to promote fake news, and they are linked to fraudulent reviews, comments, and

other content. fraudulent news is spreading quickly with the aid of fake accounts. Social media behemoths like Facebook and Twitter work tirelessly to identify and eliminate phony accounts, but the issue really only becomes worse as social media usage increases. To identify the fraudulent social media profiles, we have used Python and machine learning techniques. Three distinct various methods that are , namely Support Vector Machines (SVM) and Neural Networks (NN), are employed. The results indicate that NN outperforms SVM in terms of AUC. Researchers are always trying to lessen, if not completely eradicate, this significant social network issue, and as AI capabilities continue to advance, they anticipate that this problem will become less of a concern in the future.

References

[1]. Breuer, A. , Eilat, R. , Weinsberg, U. (2020, April). Friend or Faux: Graph- Based Early Detection of Fake Accounts on Social Networks. In Proceedings of The Web Conference 2020 (pp. 1287- 1297).

[2]. Balaanand, M. Karthikeyan, N. Karthik, S. Varatharajan, R. Manoharan, G. Sivaparthipan, C. B. (2019). An enhanced graph based semi-supervised learning algorithm to detect fake users on Twitter. The Journal of Supercomputing.

[3]. Jiang, X. , Li, Q. , Ma, Z. , Dong, M. , Wu, J. , Guo, D. (2019). Quick Squad: A new single- machine graph computing framework for detecting fake accounts in large-scale social networks. Peer-to Peer Networking and Applications, 12(5), 1385-1402.

[4]. Sahoo, S. R. , Gupta, B. B. (2020). Real-Time Detection of Fake Account in Twitter Using Machine-Learning Approach. in Computational Intelligence and Communication Technology (pp. 149- 159). Springer, Singapore.

[5]. Pakaya, F. N, Ibrohim, M. O. , Budi, I. (2019, October). Malicious Account Detection on Twitter Based on Tweet Account Features using Machine Learning.

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

[7]. El-Mawass, N. Honeine, P. Vercouter, SimilCatch: Enhanced social spammers detection on Twitter using Markov Random Fields. Information Processing Management, 57(6), 102317.

[8]. ESTEE VAN DER WALT and JANELOFF Using Machine Learning to Detect Fake Identities: Bots vs Humans" Received December 5, 2017, acknowledged January 12, 2018, date of production January 23, 2018, date of current rendition Walk 9, 2018.

[9]. Swamy, S. Ranga, et al. "Dimensionality reduction using machine learning and big data technologies." Int. J. Innov. Technol. Explor. Eng.(IJITEE) 9.2 (2019): 1740-1745.

[10]. Swamy, R. S., S. C. Kumar, and G. A. Latha. "An efficient skin cancer prognosis strategy using deep learning techniques." Indian Journal of Computer Science and Engineering (IJCSE) 12.1 (2021).

[11]. Sirisati, Ranga Swamy, et al. "An Enhanced Multi Layer Neural Network to Detect Early Cardiac Arrests." 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2021.
[12]. Sirisati RS, Prasanthi KG, Latha AG. An aviation delay prediction and recommendation system using machine learning techniques. InProceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020 2021 (pp. 239-253). Springer Singapore.
[13]. Sirisati, Ranga Swamy, et al. "Cancer Sight: Illuminating the Hidden-Advancing Breast Cancer Detection with Machine Learning-Based Image Processing Techniques." 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA). IEEE, 2023.
[14]. Sirisati, Ranga Swamy, et al. "A Deep Learning Framework for Recognition and Classification of Diabetic Retinopathy Severity." Telematique 23.01 (2024): 228-238.
[15]. Praveen, S. P., Vellela, S. S., & Balamanigandan, R. (2024). SmartIris ML: Harnessing Machine Learning for Enhanced Multi-Biometric Authentication. Journal of Next Generation Technology (ISSN: 2583-021X), 4(1).
[16]. Madhuri, A., Jyothi, V. E., Praveen, S. P., Altaee, M., & Abdullah, I. N. (2023). Granulation-Based Data Fusion Approach for a Critical Thinking Worldview Information Processing. Journal of Intelligent Systems and Internet of Things, 9(1), 49-68.
[17]. Dedeepya, P., Sowmya, P., Saketh, T. D., Sruthi, P., Abhijit, P., & Praveen, S. P. (2023, February). Detecting cyber bullying on twitter using support vector machine. In 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 817-822). IEEE.
[18]. Praveen, S. P., Suntharam, V. S., Ravi, S., Harita, U., Thatha, V. N., & Swapna, D. (2023). A Novel Dual Confusion and Diffusion Approach for Grey Image Encryption using Multiple Chaotic Maps. International Journal of Advanced Computer Science and Applications, 14(8).
[19]. Jyothi, V. E., Kumar, D. L. S., Thati, B., Tondepu, Y., Pratap, V. K., & Praveen, S. P. (2022, December). Secure data access management for cyber threats using artificial intelligence. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology (pp. 693-697). IEEE.
[20]. Praveen, S., Nguyen, H., Swapna, D., Rao, K., & Kumar, D. (2020). The efficient way to detect and stall fake articles in public media using the blockchain technique: Proof of trustworthiness. International Journal on Emerging Technologies, 11(3), 158-163.
[21]. Swapna, D., & Praveen, S. P. (2020). An exploration of distributed access control mechanism using blockchain. In Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 2 (pp. 13-20). Springer Singapore.
[22]. Vemulakonda, R., Meka, S., Ketha, V., Surapaneni, P. P., & Kondapalli, S. V. (2016). An algorithm for basic IoT Architectures.

[23]. Sindhura, S., Praveen, S. P., Rao, N., & ArunaSafali, M. (2021, September). An energy efficient novel routing protocol in wireless sensor networks (WSN). In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 319-324). IEEE.

[24]. Marrapu, B. V., Raju, K. Y. N., Chowdary, M. J., Vempati, H., & Praveen, S. P. (2022, January). Automating the creation of machine learning algorithms using basic math. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 866-871). IEEE.

[25]. S. Phani Praveen , Thulasi Bikku, P. Muthukumar, K. Sandeep, Jampani Chandra Sekhar, V. Krishna Pratap. (2024). Enhanced Intrusion Detection Using Stacked FT-Transformer Architecture. Journal of , 13 (2), 19-29 (Doi : https://doi.org/10.54216/JCIM.130202)

[26]. Bikku, T., Chandolu, S. B., Praveen, S. P., Tirumalasetti, N. R., Swathi, K., & Sirisha, U. (2024). Enhancing Real-Time Malware Analysis with Quantum Neural Networks. Journal of Intelligent Systems and Internet of Things, 12(1), 57-7.

[27]. Aruna, R., Kushwah, V. S., Praveen, S. P., Pradhan, R., Chinchawade, A. J., Asaad, R. R., & Kumar, R. L. (2024). Coalescing novel QoS routing with fault tolerance for improving QoS parameters in wireless Ad-Hoc network using craft protocol. Wireless Networks, 30(2), 711-735.