Temporal Sentiment Analysis (TSMFPMSM) Model for Multimodal Social Media Fake Profile Detection

Bhrugumalla L. V. S. Aditya[1(✉)], Sachi Nandan Mohanty[1], and Yalamanchili Salini[2]

[1] School of Computer Science and Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India
{aditya.22phd7023,sachinandan.m}@vitap.ac.in

[2] Information Technology, VR Siddhartha Engineering College, Andhra Pradesh, India

Abstract. Today's social media sites should be able to spot fake profiles. Most social media accounts (around 25%) are fake or managed by automated software. Therefore, advanced models are required to detect and remove these fake profiles. Anomalies in login, usage, and non-functional elements have inspired researchers to construct pattern analysis models. This book expands on existing models using temporal sentiment analysis to spot fake profiles in social networks with multiple interface types. Massive datasets are gathered from social media platforms like Twitter, Facebook, and others and used in the model. TSPs are derived from these data sets using a unique ensemble sentiment analysis engine. Using Afinn, GloVe, and Word2Vec, the sentiment analysis engine labels statements as "positive," "negative," or "neutral." These TSPs teach a 1D CNN to identify fake profiles with high accuracy. The true-human behavioral theory influenced the model, which predicts that false users' periodic data will always converge. User perspectives are accounted for in the model. Therefore, the model achieves a 5.9%, 4.5%, and 3.2% improvement over state-of-the-art techniques for identifying bogus profiles, respectively. Multimodal social media interfaces benefit from this method.

Keywords: social media · Temporal · Afinn · 1D CNN · GloVE · Word2Vec · Interfaces · Sentiments

1 Introduction

Almost everyone today has a social media account. It's a great way to brag about expensive camera equipment, stalk your favorite stars, and stay in touch with your pals. We can meet new people and expand our knowledge thanks to these channels. There are negative aspects. Conflicts arise through social media. Twitter's 145 million DAUs and 330 million MAUs pale compared to Facebook's 500,000 DAUs and 6 SAUs added every minute. Twitter's constant updates include fresh news, itineraries, trending topics, and hashtags. A user's ability to like, remark, share, and respond is limited to 280 characters. Despite the serious topics that tend to dominate social media conversations, rumors can

O. Castillo et al. (Eds.): PerSOM 2023, LNICST 517, pp. 329–341, 2024.
https://doi.org/10.1007/978-3-031-66044-3_25

potentially spread and amplify societal tensions. Significant issues include privacy invasion, exploitation, cyberbullying, and fake news [1–3]. False information is an important issue today. Artificial neural networks (ANNs) can distinguish between human, bot, and cybernetic identities [4–6]. Computers currently handle “cyborg” accounts despite their human origin. These sham profiles spread hate speech and graphic images of abuse. They could potentially affect public opinion by applying anti-vaccination theories. Fake accounts plague Internet message boards. Fake accounts are used in spamming [7–9]. Cybercriminals are the ones who made these accounts. Concerns over identity theft and data breaches have led researchers to develop methods for detecting fake accounts, some of which are Collaborative Filtering (CF) [10] and C4.5 Decision Tree (CDT) [12]. User’s personal information is sent to malicious third-party services when they visit URLs associated with these fake accounts. Fake accounts often impersonate legitimate businesses or individuals, damaging credibility and influencing how many people follow a profile.

2 Temporal Sentiment Analysis (TSMFPMSM) Model for Multimodal Social Media Fake Profile Detection

Researchers have presented a wide range of pattern analysis algorithms after studying methods for detecting fake profiles. These models estimate login, usage, and non-functional aspects to identify unusual user activity. Based on these models, the temporal sentiment analysis model can identify phony profiles on social media platforms that use several input methods. The model’s operation is depicted in Fig. 1. As a first step, the model gathers extensive statistics from numerous social media platforms. These data sets will be used as the basis for further investigation. Temporal sentiment patterns (TSPs) are produced by running the collected datasets through a novel ensemble sentiment analysis engine. This engine classifies words as “positive,” “negative,” or “neutral” using a mixture of Afinn, GloVe, and Word2Vec. The obtained TSPs efficiently detect false profiles by training a 1D Convolutional Neural Network (1D CNN). Periodic data sent by false users always converge, according to the true-human behavioral theory, which is incorporated into the suggested model. This theory applies to user feelings, for which it is accurate. The first stage of the methodology is gathering ‘Fake’ and ‘Genuine’ social media posts from various sources, including Twitter, Facebook, and others. Then, the Afinn, GloVe, and Word2Vec models categorize the posts as positive, negative, or neutral. The Afinn sentiment analyzer uses a pre-tagged vocabulary to determine how an individual feels about a statement by comparing the words used. Word2Vec performs a semantic analysis, considering contextual meaning and synonyms, before transforming the words into vectors. The GloVe engine uses the training corpus to assign global frequency aggregates to individual words.

The sentiment analysis conducted by Afinn evaluates sentiments by assigning a final sentiment score (*Sout*) ranging from −1 to 1, indicating negative, neutral, and positive sentiments, respectively. The estimation of this score follows Eq. 1, which is presented below.

$$S_{out} = \frac{\sum_{i=1}^{N} \left|W_i == W_p\right| - \sum_{i=1}^{N} \left|W_i == W_n\right|}{N} \tag{1}$$

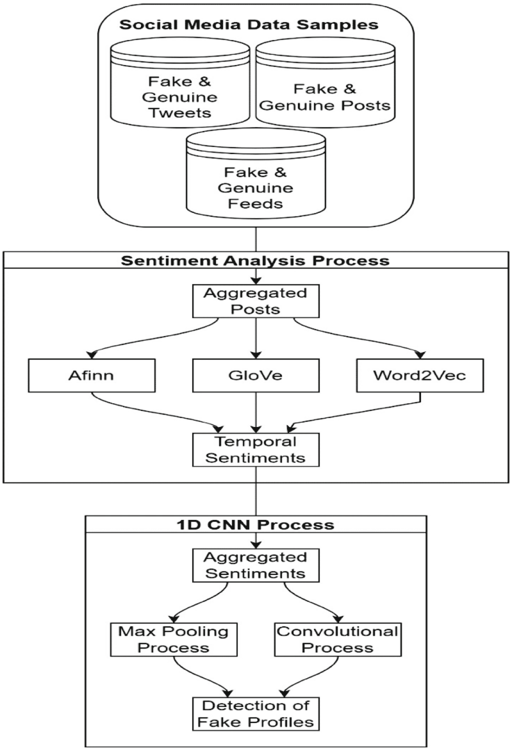

Fig. 1. Proposed process for fake profile detection

Wi is the input word, Wp and Wn are the positive and negative words in the corpus, and N is the number of terms used in the equation. Figure 2 depicts how the Word2Vec model, also used for sentiment analysis, uses many constructors.

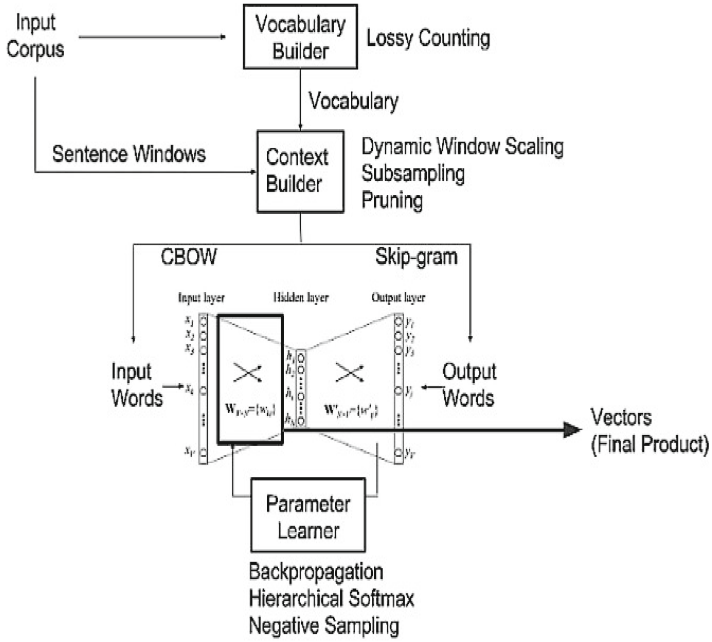

Fig. 2. Word2Vec process of sentiment analysis

Word2Vec's approach to sentiment analysis consists of many tools, such as a thesaurus, a context generator, and a continuous bag of words (CBoW) engine. Key takeaways can be gleaned from closely reading the document's many sections. To get started, the program harvests "action words" from existing sentences to generate a vocabulary. The context builder receives this vocabulary and uses it to create word pairings to find sets of words that go together. These variations are used by the CBoW engine and skip-gram models to improve feature extraction performance. A two-layer neural network processes these features to generate feelings from input word pairs. As a result of the overall sentiment analysis, the model generates sentiment linked with each word pair while considering the context.

The engine's emotion classification performance is impressive, but training can take time. With a few tweaks, GloVe uses a similar design. One-hot encoding is used to represent words in GloVe's model, which requires more time in training and more frequent splits between the training and testing sets. The number and quality of the training corpus affect the GloVe model's prediction accuracy. To further understand how one-hot encoding works, the word embeddings generated for different input texts are graphically represented below (Fig. 3).

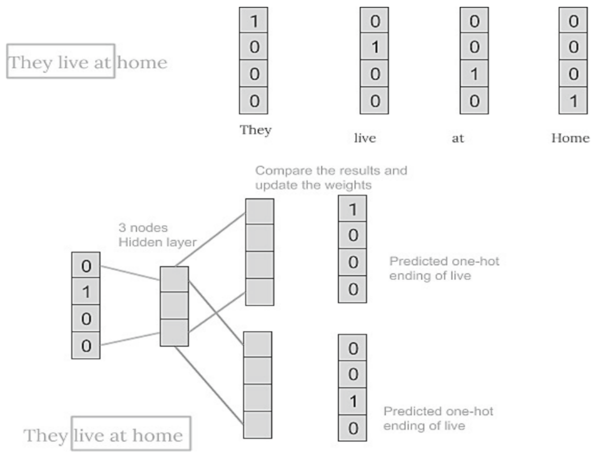

Fig. 3. Sentiment Analysis for GloVe Embedding Model

Following encoding, the values are fed into a neural network that, like Word2Vec, operates on vector representations of words. Finally, the network's sentiment scores are tested across various data sets. Combining the results of these techniques yields sentiment scores that are then used to generate a 1D temporal sentiment feature vector. A 1D Convolutional Neural Network (CNN) model is then used to classify these vectors, which helps detect fraudulent profiles. Figure 4 depicts the CNN model's flow, with the various layers and their sizes laid out for practical applications.

Convolutional layers, Max Pooling layers, and Dropout layers are all a part of the suggested model for efficient feature extraction and the selection of richly diverse feature sets. These layers are linked in a cascading fashion to improve feature enhancement, each with its unique window and stride parameters. A SoftMax-based fully connected layer is then used to categorize the enriched feature sets, allowing for the easier identification of distinct profile kinds.

The model uses Eq. 2 to extract convolutional features, which helps improve and expand sentiment feature sets.

$$C_{feats} = \sum\nolimits_{a=-\frac{m}{2}}^{\frac{m}{2}} SV(i-a) \times ReLU\left(\frac{m+a}{2}\right) \tag{2}$$

In Eq. 2, the variables m and a denote the sizes of the convolutional window and stride, respectively. SV represents the sentiment vectors, while $ReLU$ represents the rectilinear unit used to activate these features. The extracted features are then passed to a variance maximization layer. Equation 3 is utilized to estimate the variance of the convolutional features, enabling the removal of features with lower variance levels.

$$v = \sqrt{\frac{\left(\sum\nolimits_{i=1}^{D_s}\left(x_i - \sum\nolimits_{j=1}^{D_s}\frac{x_j}{D_s}\right)^2\right)}{D_s+1}} \tag{3}$$

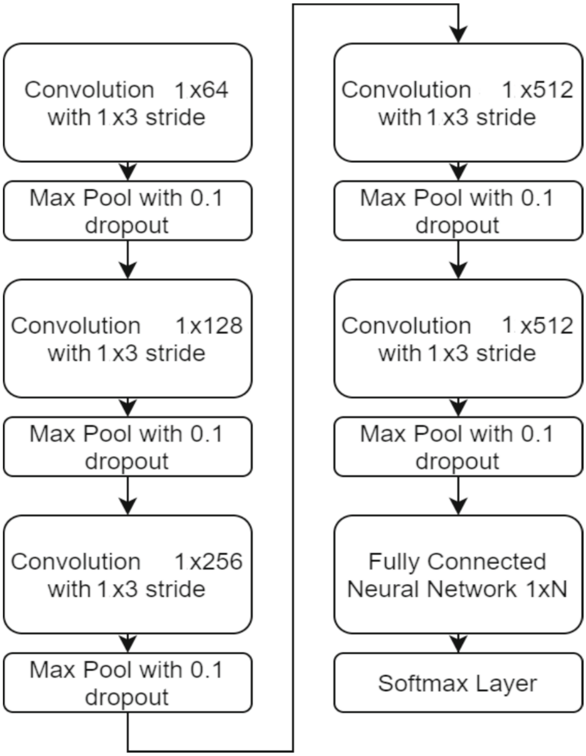

Fig. 4. Identification of fake profiles with 1D CNN Model

In Eq. 4, the variables x and Ds represent the feature value and the total samples extracted by the convolutional layers, respectively. These features are classified using feature-level weights (w) and biases (b), enabling the identification of different profile types.

$$C_{out} = softMax\left(\sum_{i=1}^{N} f_i \times w_i + b_i\right) \tag{4}$$

In Eq. 5, Nf represents the total number of features extracted by the preceding convolutional layers. By employing the customized 1D CNN model, the proposed technique achieves high accuracy in identifying fake profiles. The performance of the proposed model was evaluated on various sample sets and compared with standard models, which will be discussed in the subsequent section of this text.

3 Result Analysis and Comparison

The suggested approach starts with collecting data from several social feeds, which is subsequently processed by Afinn, GloVe, and Word2Vec models to isolate temporal feature sets. Classifying these feature sets with a 1D CNN can quickly determine profile kinds. The following datasets were used to assess the quality of this model:

https://www.kaggle.com/datasets/free4ever1/instagram-fake-spammer-genuine-accounts.

Social Network Fake Account Dataset, which is available at https://www.kaggle.com/datasets/bitandatom/social-network-fake-account-dataset

Our World Dataset, which is available at https://ieeexplore.ieee.org/document/9458194

The combined datasets consisted of a total of 5,000 social media profiles. Among these, 70% of the profiles were used for training, 10% for testing, and 20% for validation. The performance of the proposed model was compared to that of ANN [5], CDT [12], and DFB [16] as benchmarks to assess its efficiency compared to standard implementations. Table 1 provides the accuracy (A) of fake profile detection about the total test entries (TTE), illustrating the results obtained from this approach

Table 1. Fake profile identification for different data samples Estimated accuracy.

TTE	A (%) ANN [5]	A (%) CDT [12]	A (%) DFB [16]	A (%) TSM FPM SM
250	85.52	84.77	89.00	93.25
500	86.15	85.43	89.65	93.89
750	86.61	85.90	90.13	94.35
1000	86.92	86.20	90.45	94.65
1250	87.12	86.42	90.67	94.86
1500	87.29	86.63	90.84	95.05
1875	87.45	86.86	91.00	95.23
2250	87.65	87.14	91.22	95.46
2500	87.88	87.44	91.45	95.71
3000	88.11	87.72	91.69	95.96
3750	88.36	88.00	91.94	96.25
5000	88.68	88.34	92.27	96.60

The evaluation and analysis, as shown in Fig. 5, demonstrate that the proposed model achieved a significant improvement in fake profile detection accuracy compared to ANN [5], with an increase of 8.3%. Furthermore, it outperformed CDT [12] by 8.5% and DFB [16] by 3.9% in accuracy. These results highlight the model's effectiveness across various real-time use cases. The improved accuracy can be attributed to a high-efficiency 1D CNN classifier trained to optimize classification performance for different data types. Furthermore, Table 2 provides the precision values, which can be examined as follows.

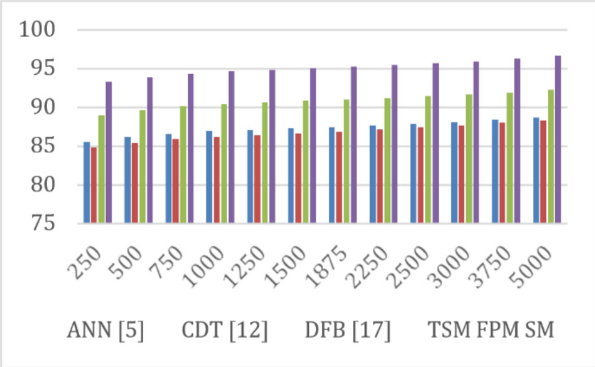

Fig. 5. Fake profile identification for different data samples Estimated accuracy

Table 2. Fake profile identification for different data samples Estimated precision

TTE	P (%) ANN [5]	P (%) CDT [12]	P (%) DFB [16]	P (%) TSM FPM SM
250	81.63	80.91	84.95	89.00
500	82.23	81.55	85.58	89.62
750	82.67	81.99	86.03	90.06
1000	82.97	82.28	86.34	90.35
1250	83.16	82.49	86.54	90.55
1500	83.31	82.69	86.71	90.73
1875	83.48	82.90	86.87	90.91
2250	83.67	83.17	87.07	91.12
2500	83.89	83.46	87.29	91.36
3000	84.10	83.73	87.51	91.60
3750	84.34	84.00	87.76	91.87
5000	84.65	84.33	88.08	92.20

Based on the evaluation depicted in Figure 6, it is evident that the proposed model exhibited a substantial improvement in precision for fake profile detection. Compared to ANN [5], it achieved a 7.5% higher precision. Additionally, it outperformed CDT [12] by 8.3% and DFB [16] by 4.5% in precision. These results emphasize the model's effectiveness across various real-time use cases. The enhanced precision is attributed to the use of temporal sentiments and the high-efficiency 1D CNN classifier, which has been trained to optimize precision performance for different data types. Furthermore, Table 3 provides insights into the recall of the classification, which can be examined as follows.

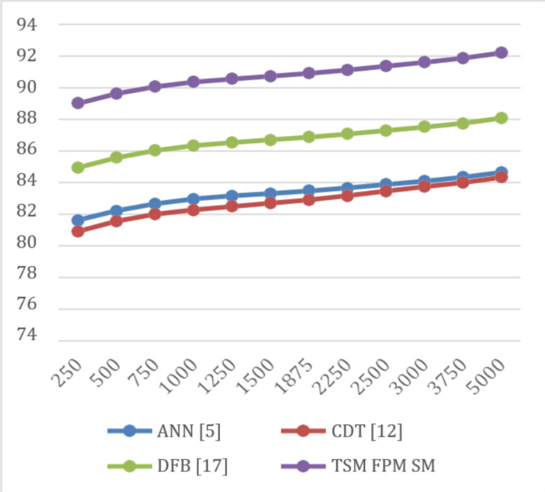

Fig. 6. Fake profile identification for different data samples Estimated precision

Table 3. Fake profile identification for different data samples Estimated recall

NT	R (%) ANN [5]	R (%) CDT [12]	R (%) DFB [16]	R (%) TSM FPM SM
250	83.02	82.29	86.40	90.51
500	83.63	82.93	87.03	91.14
750	84.08	83.38	87.49	91.58
1000	84.37	83.68	87.80	91.88
1250	84.57	83.89	88.01	92.09
1500	84.73	84.09	88.17	92.26
1875	84.89	84.31	88.34	92.45
2250	85.09	84.58	88.55	92.67
2500	85.30	84.88	88.77	92.91
3000	85.53	85.15	89.00	93.15
3750	85.77	85.42	89.25	93.43
5000	86.08	85.75	89.57	93.77

Figure 7 displays the estimated recall for fake profile identification across different data samples. Based on the evaluation and analysis depicted in Figure 8, the proposed model notably improves recall for fake profile detection. Compared to ANN [5], it achieved a 6.5% higher recall. Similarly, it outperformed CDT [12] by 7.4% and DFB [16] by 5.5% in the recall. These results underscore the model's effectiveness in various real-time use cases. The enhanced recall can be attributed to the utilization of temporal sentiments and the high-efficiency 1D CNN classifier, which has been trained to optimize recall performance for different data types. Furthermore, Table 4 provides information on the delay of classification, which can be examined as follows.

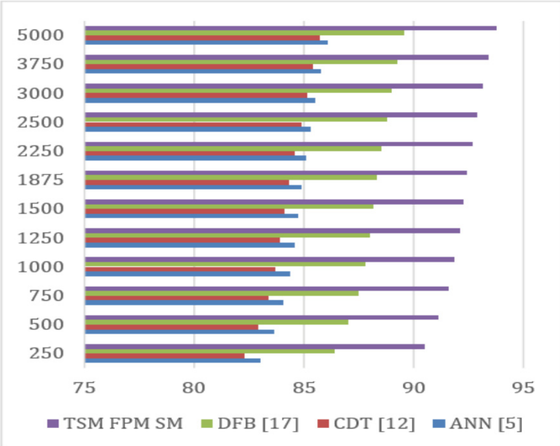

Fig. 7. Fake profile identification of different data samples Estimated recall

Table 4. The estimated delay needed for fake profile identification under different data samples

NT	D (ms) ANN [5]	D (ms) CDT [12]	D (ms) DFB [16]	D (ms) TSM FPM SM
250	168.76	167.27	175.70	149.63
500	170.00	168.58	176.92	150.59
750	170.90	169.49	177.85	151.30
1000	171.51	170.10	178.49	151.77
1250	171.91	170.53	178.90	152.11
1500	172.24	170.94	179.24	152.40
1875	172.56	171.39	179.57	152.71
2250	172.97	171.94	179.99	153.08
2500	173.41	172.54	180.46	153.47
3000	173.86	173.09	180.92	153.88
3750	174.35	173.64	181.42	154.35
5000	174.98	174.32	182.06	154.90

Figure 8 shows the evaluation results, which show that the suggested model significantly reduces the identification delay for detecting bogus profiles. Its identification delay was 10.5% lower than ANN's [5]. In terms of identification latency, it also fared

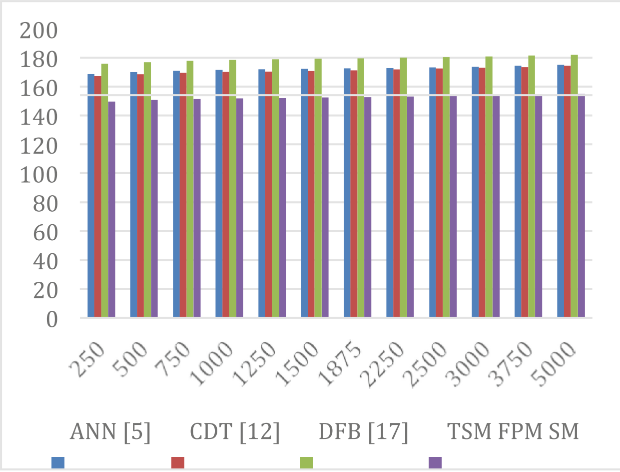

Fig. 8. Time required, on average, to spot a phony profile based on a variety of data sets

better than CDT [12] (9.5%) and DFB [16] (8.3%). These findings demonstrate the model's versatility for various high-velocity applications. Time savings result from using temporal sentiments and a very effective 1D CNN classifier tuned to maximize speed performance across a wide range of data types. This enhanced functionality makes the suggested model suitable for real-time social media settings to identify false profiles.

4 Conclusion and Future Scope

Temporal feature sets from various social feeds are identified using Afinn, GloVe, and Word2Vec models. A 1D CNN then classifies these feature sets to determine profile categories efficientlyQ4. The proposed model's accuracy was 8.3% higher than that of ANN [5], 8.5% higher than that of CDT [12], and 3.9% higher than that of DFB [16], making it suitable for real-time applications. One way to enhance accuracy is to use a multi-purpose 1D convolutional neural network (CNN) classifier.The proposed model had an accuracy of 7.5%, which was higher than ANN [5], CDT [12], and DFB [16]. A high-efficiency, multi-input-type, one-dimensional convolutional neural network classifier is used to increase the accuracy of temporal feelings. The proposed model outperformed ANN [5], CDT [12], and DFB [16] in terms of recall in real-world settings by 6, 5, and 5.5 percentage points, respectively. A high-efficiency 1D convolutional neural network (CNN) classifier tuned for recall across data types improves recall when applied to temporal feelings. The proposed model is suitable for high-speed use cases because its identification latency is 10.5% shorter than ANN [5], 9.5% lower than CDT [12], and 8.3% lower than DFB [16]. Reduce latency using temporal sentiment analysis and a solid 1D CNN classifier tuned for speed performance across many data types.

References

1. Pizarro, J.: Profiling bots and fake news spreaders at PAN'19 and PAN'20: bots and gender profiling 2019, profiling fake news spreaders on Twitter 2020. In: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pp. 626–630 (2020). https://doi.org/10.1109/DSAA49011.2020.00088
2. Krishnan, P., Aravindhar, D.J., Bhanu Prakash Reddy, P.: Finite automata for fake profile identification in online social networks. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1301–1305 (2020). https://doi.org/10.1109/ICICCS48265.2020.9121086
3. Harris, P., Gojal, J., Chitra, R., Anithra, S.: Fake Instagram profile identification and classification using machine learning. In: 2021 2nd Global Conference for Advancement in Technology (GCAT), pp. 1–5 (2021). https://doi.org/10.1109/GCAT52182.2021.9587858
4. Tiwari, V.: Analysis and detection of fake profile over the social network. In: 2017 International Conference on Computing, Communication, and Automation (ICCCA), pp. 175–179 (2017). https://doi.org/10.1109/CCAA.2017.8229795
5. Hajdu, G., Minoso, Y., Lopez, R., Acosta, M., Elleithy, A.: Use of artificial neural networks to identify fake profiles. In: 2019 IEEE Long Island Systems, Applications and Technology Conference (LISAT), pp. 1–4 (2019).https://doi.org/10.1109/LISAT.2019.8817330
6. Rathod, S.: Exploring author profiling for fake news detection. In: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1614–1619(2022). https://doi.org/10.1109/COMPSAC54236.2022.00256
7. Kulkarni, V., Aashritha Reddy, D., Sreevani, P., Teja, R.N.: Fake profile identification using ANN. In: 4th Smart Cities Symposium (SCS 2021), pp. 375–380 (2021). https://doi.org/10.1049/icp.2022.0372
8. Patel, K., Agrahari, S., Srivastava, S.: Survey on fake profile detection on social sites by using machine learning algorithm. In: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1236–1240 (2020).https://doi.org/10.1109/ICRITO48877.2020.9197935
9. Durier da Silva, F.C., Cristina Bicharra Garcia, A., Matsui Siqueira, S.W.: A systematic literature mapping on profile trustworthiness in fake news spread. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 275–279 (2022). https://doi.org/10.1109/CSCWD54268.2022.9776232
10. Parihar Devanand, P., Kumar, N.: Fake profile detection from the social dataset for movie promotion. In: 2021 Sixth International Conference on Image Information Processing (ICIIP), pp. 495–498 (2021). https://doi.org/10.1109/ICIIP53038.2021.9702684
11. Roy, P.K., Chahar, S.: Fake profile detection on social networking websites: a comprehensive review. IEEE Trans. Artif. Intell. **1**(3), 271–285 (2020). https://doi.org/10.1109/TAI.2021.3064901
12. Sowmya, P., Chatterjee, M.: Detection of fake and clone accounts in Twitter using classification and distance measure algorithms. In: 2020 International Conference on Communication and Signal Processing (ICCSP), pp. 0067–0070 (2020). https://doi.org/10.1109/ICCSP48568.2020.9182353
13. "Keynote Speech 2: Detecting fake news and profiling fake news spreaders and conspiracy propagators," 2021 12th International Conference on Information and Communication Systems (ICICS), p. 2 (2021). https://doi.org/10.1109/ICICS52457.2021.9464534
14. Revathi, S., Suriakala, D.M.: Profile similarity communication matching approaches for detection of duplicate profiles in online social network. In: 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 174–182 (2018). https://doi.org/10.1109/CSITSS.2018.8768751

15. Shu, K., Wang, S., Liu, H.: Understanding user profiles on social media for fake news detection. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 430–435 (2018). https://doi.org/10.1109/MIPR.2018.00092
16. Wani, M.A., Agarwal, N., Jabin, S., Hussain, S.Z.: Analyzing real and fake users in Facebook network based on emotions. In: 2019 11th International Conference on Communication Systems & Networks (COMSNETS), pp. 110–117 (2019). https://doi.org/10.1109/COMSNETS.2019.8711124
17. Conti, M., Poovendran, R., Secchiero, M.: FakeBook: detecting fake profiles in online social networks. In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 1071–1078 (2012). https://doi.org/10.1109/ASONAM.2012.185
18. Narayanan, A., Garg, A., Arora, I., Sureka, T., Sridhar, M., Prasad, H.B.: IronSense: towards the identification of fake user-profiles on Twitter using machine learning. In: 2018 Fourteenth International Conference on Information Processing (ICINPRO), pp. 1–7 (2018).https://doi.org/10.1109/ICINPRO43533.2018.9096687
19. Singh, N., Sharma, T., Thakral, A., Choudhury, T.: Detection of fake profile in online social networks using machine learning. In: 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), pp. 231–234 (2018).https://doi.org/10.1109/ICACCE.2018
20. Raja, E.V.S., Aditya, B.L., Mohanty, S.N.: Fake profile detection using logistic regression and gradient descent algorithm on online social networks. EAI Endorsed Trans. Scalable Inf. Syst. **11**(1) (2023). https://doi.org/10.4108/eetsis.4342
21. Aditya, B.L.V.S., Mohanty, S.N.: Heterogenous social media analysis for efficient deep learning fake-profile identification. IEEE Access **11**, 99339–99351 (2023). https://doi.org/10.1109/ACCESS.2023.3313169
22. Aditya, B.L.V.S., Rajaram, G., Hole, S.R., Mohanty, S.N. (2023). F2PMSMD: Design of a Fusion Model to Identify Fake Profiles from Multimodal Social Media Datasets. In: Nandan Mohanty, S., Garcia Diaz, V., Satish Kumar, G.A.E. (eds) Intelligent Systems and Machine Learning. ICISML 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 471. Springer, Cham