Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: https://jaraset.com/
ISSN: 2462-1943

A REVIEW ON MACHINE LEARNING HYBRID MODEL FOR SOCIAL MEDIA THREAT DETECTION AND PREDICTION

Rashmi Tiwari [1*], Dr. Gaurav Aggarwal [2]

[1] Research Scholar, Faculty of Engineering & Technology , Jagannath University, Jhajjar, India; rt45720@gmail.com
[2] Professor, Faculty of Engineering & Technology , Jagannath University, Jhajjar, India; gaurav.aggarwal@jagannathuniversityncr.ac.in

ARTICLE INFO

Article history:
Received: 09-04-2024
Received in revised form: 10-05-2024
Accepted: 10-06-2024
Available online: 23-08-2024

Keywords:

Machine Learning, Deep Learning, social media, Hybrid Model

ABSTRACT

The exponential expansion of social media has resulted in a flood of user-generated information on previously unseen scales, making it difficult to maintain user privacy & security. This research provides a thorough analysis of hybrid machine learning (ML) models developed to identify & threats social media security risks. The review begins by exploring the landscape of social media threats, encompassing diverse categories such as cyberbullying, hate speech, misinformation, and malicious activities. Because of the potential for increased attack variety brought on by the pandemic, cyber-security will continue to be a crucial industry in the years to come. There are numerous methods to offer cyber security, including firewalls, IDS, and authentication and encryption (intrusion detection system). Subsequently, it surveys the technology on standalone ML models employed for threat detection and identifies their limitations in handling the evolving nature of threats on social media.

INTRODUCTION

With so many people utilizing the services of the WWW for their daily activities, the Internet has grown to be a significant part of today's society. Individuals access the internet for both personal & professional purposes, & large percentage of tasks includes sharing data, retrieving data, transferring files, interacting with friends or coworkers on social media, & most relevantly in e-commerce, which necessitates storing passwords & credit/debit payment information. Organizations as well as

[*] *Corresponding author.*
E-mail address: rt45720@gmail.com

individuals rely greatly on the Internet or its services. Network attacks in the form of malicious traffic, according to M. Uma et al. (2013), result in data loss, invasions of people's privacy, negative financial, political, & economic effects on major businesses, & halted functioning for all of their shareholders. Working from home is now considered the new norm as a result of Covid-19 & ongoing pandemic. Due to less complex protection methods on personal devices compared to those of the company, this has made them vulnerable. Because of the potential for increased attack variety brought on by the pandemic, cyber-security will continue to be a crucial industry in the years to come. There are numerous methods to offer cyber security, including firewalls, IDS, and authentication and encryption (intrusion detection system). IDS has established itself as a superior option to previous strategies by offering social media monitoring and behavior analysis, as well as further identifying attacks from network flow [E. Vasilomanolakis 2015]. Cyberattack detection works similarly to a classification technique in that it classifies assaults into benign and malicious categories. By categorizing network traffic, traditional ML techniques, commonly referred to as shallow learning, have been employed for intrusion detection [A. L. Buczak 2015]. Due to ML techniques' excessive reliance on the characteristics chosen by human experts, performance of these techniques degrades when real-world data grows over time, creating a high dimensional space.

MACHINE LEARNING

The process of creating a machine learning algorithm consists of two distinct phases: training & testing. The methods used to train a given model can vary. Frameworks like scikit-learn [F. Pedregosa 2011], Tensorflow, PyTorch, Matlab, & Weka are commonly used to handle this procedure & creation of machine learning models. Optimisation of the algorithm or the amount of available parameters are both profoundly affected by the framework. Classification, regression, & reconstruction are three tasks that machine learning models can accomplish that are of special importance for intrusion detection. Using a variety of criteria, such as "normal" vs "attack" or "attack family," classification places data into distinct groups. The chance that a given input represents an assault can be calculated using regression (also known as "prediction"), which is used to ascertain continuous quantities. Finally, only a certain kind of neural network is suitable for reconstruction. To make the network learn the features (illustration learning), this task compresses & decompresses the input data.

DEEP LEARNING

There are numerous definitions of deep learning that have been proposed by various scholars. However, they all share traits & terminology in common, such as "nonlinear data transformations,"

"unsupervised machine learning," "complex architectural data model," & "learning several layers." All of these key terms have a tight connection to neural networks and pattern recognition [Deng2014]. Deep learning typically doesn't need pre-selected features, which solves the feature selection problem [Mrazova et al. 2012; Tian Tian et al. 2020]. Deep learning automatically derives meaningful features from raw input to solve the given problem. Deep learning models often comprise many processing layers with different levels of abstraction, allowing the system to learn different data features. The network may remember distinctive traits because of the various levels that are present. In many fields, including topic classification, sentiment analysis, signal processing, natural language processing, face recognition, speech recognition, image recognition, and natural language processing, deep learning has been widely acknowledged as a method that produces promising results [Bjornson et al. 2020]. Additionally, a wide range of deep learning architectures, including DBNs, RNNs, & CNNs, have been created.

CONVOLUTIONAL NEURAL NETWORK

According to Yamashita et al. (2018), a CNN is a deep learning model that processes data like images & based on how the animal visual cortex is organized. In low to high level patterns, it is principally made to automatically & adaptively learn the spatial hierarchies of features [Yamashita et al. 2018]. Several tasks, including face identification, object identification, and traffic sign detection, have been successfully completed using it, most notably in robots & self-driving automobiles [Zhang et al. 2020]. The most crucial component of a CNN is reducing the amount of parameters in an ANN, which has encouraged researchers & developers to concentrate on building larger models that could be applied to address challenging problems—something that is not achievable with conventional ANNs [Albawi et al. 2017].

SENTIMENT ANALYSIS (SA)

SA of authoritative content published on SM sites is crucial to critically analyse user opinion/sentiment conveyed on SM sites. SA utilise text analysis & computational linguistics to sift through marketing and support databases in search of and extract evaluative data. Its goal is to examine how people feel about various items. SA aims to mine effective feelings, categorize polarity, and extract usable knowledge, essential for making better productbased judgments. Text SA scales might be document-level or sentence-level. Document-level SA is the classification of a document's sentiment polarity (positive and negative) (W. Zhang 2009). Sentence-level SA extracts sentiment polarity at sentence level (B. Liu 2012). Sentences in a text may have varying degrees of positive or negative sentiment, making it difficult to make accurate estimates. Furthermore, a phrase

is considered objective if descriptive and does not communicate feelings, but subjective sentences can be classed depending on sentiment polarity. Figure 1 depicts SA at various levels.

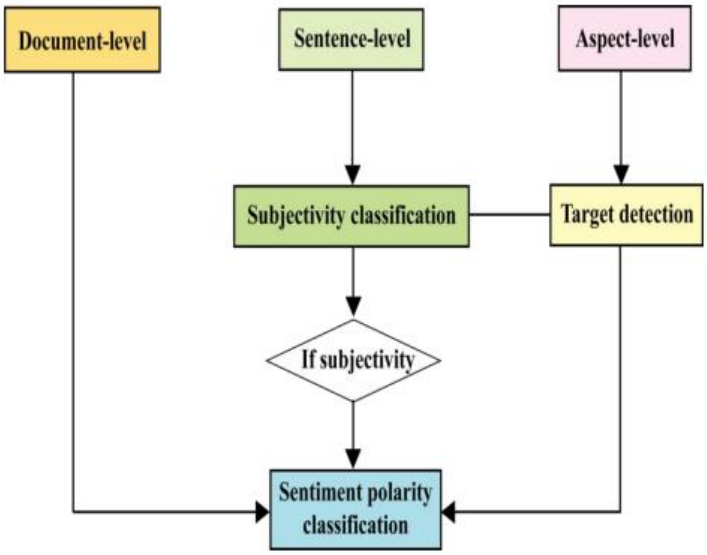

Figure 1 Concept Map of SA

In analysis where SA plays a significant part in comprehending the meaning of text data, helpful knowledge may be derived from many types of text datasets. Y. Yu, (2016) suggested a method for evaluating sentiment polarization in micro-blogs that combined text and visual information. Real-world problems necessitate efficient systems that can handle a variety of data carriers. Stock markets (M. Hagenau 2013), tourism (A.I. Kim Boes 2012), Twitter and Microblog (A. Pak 2010), and politics (I. Maks 2012) are some of the application fields of SA. An overview of SA Techniques is depicted in Figure 2.

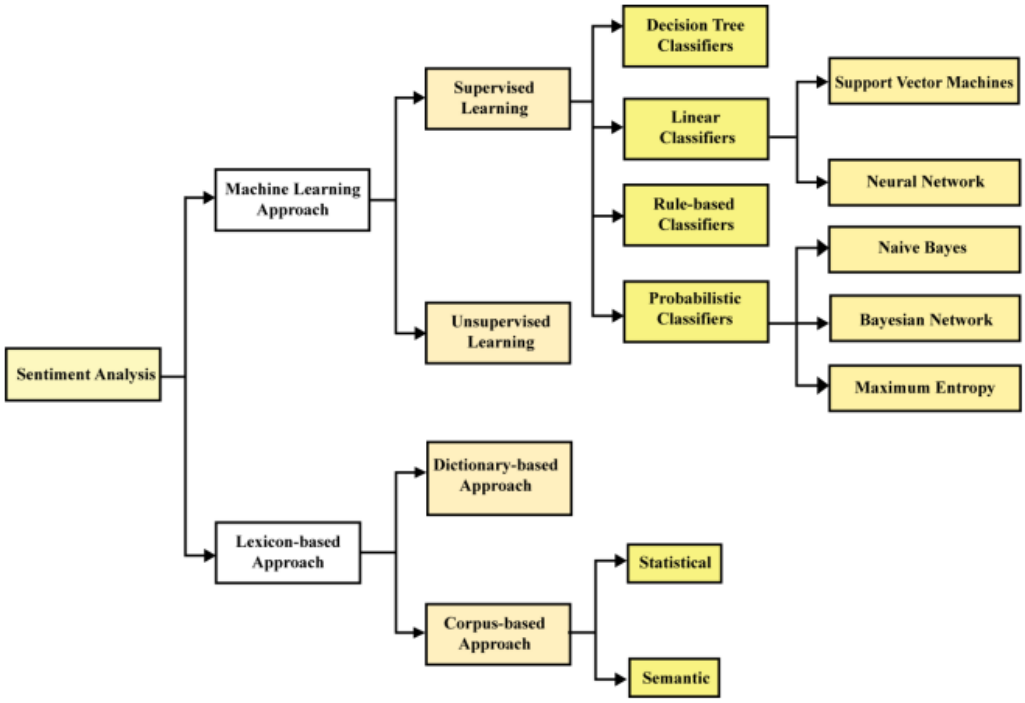

Figure 2 Overview of SA Techniques

HYBRID APPROACHES FOR SA

Due to the complexity of SA tasks, researchers have proposed hybrid SA approaches that integrate various SA methods to reap the benefits of each tactic; lexicon-based ML methodology has emerged as a key player in this regard. CNN or LSTM extracted abstract qualities fed advanced classifiers, while sentiment lexicons offered decision rules or high-quality features for ML like SVM. Proven techniques like HMMs (Hidden Markov Models) predicted sentiment polarities providing unique perspectives for SA. The suggestions of (J. Barnes 2018) offered fresh viewpoints to adaptation problems by recasting them as embedded projection problems. Their fundamental aim was to use joint optimizations and project mono-domain embedding in bi-domain spaces. Extensive testing on 11 domain pairs proved the model's superiority in domain adaptations. In (E. Cambria 2018), the authors presented a collection of AI techniques for SA, both symbolic & non-symbolic. They created SenticNet 5, novel three-level knowledge representations for SA where LSTM identified verb-noun primitives. Their generation procedure greatly expanded SenticNet 4's primary coverage. Using the three granularities of the Chinese language—radicals, characters, & words—H. Peng (2018) presented ATSM-S as a means to explicitly model aspects & effectively conduct sentiment classifications at aspect levels. They also published ATSM-F, a fusion model that performed better than many algorithms on Chinese review datasets.

F. Zhou (2017) proposed hybrid SA methods for mining customer preferences that augmented feature models. The methodology was a cross of sentiment lexicons and rough-set techniques. In the first stage, sentiment vocabulary that included POS tags and valence, arousal, and dominance ratings were created. Subsequently, decision rules or trees with lexicons identified inference relationships between features and their corresponding sentiments. In addition, they improved predefined feature models essential for understanding consumer wants by grouping lexically comparable phrases (such storage & memory) to minimise product features and then utilised SA approaches to analyse customer attitudes in product features. This work was an excellent illustration of combining SA methods and applications. P.M. Sosa (2017) proposed NN, including CNN-LSTM & LSTM-CNN, which ordinarily merged CNN & LSTM. Their experiment findings showed that LSTM-CNN had outperformed standard models by 2.7-8.5%. In recent years, top-tier journals and significant conferences have published some attractive deep learning models (E. Cambri 2016; S. Poria 2015). In a SA framework based on dependency rules & ML (S. Poria 2014), semantic parsers were utilised to parse text into ideas & integrate these concepts into vector spaces. Semantic patterns built from dependence rules were applied for ideas found in SenticNet. ELM classifiers were used when the concepts were not found in SenticNet and the suggested framework outperformed other methods while being executed on movie review datasets. J. Märkle-Huß (2017) presented a SA strategy based on Rhetoric Structure Theory. Despite its rarity, the text's proposed semantic structure offered a powerful level to improve traditional SA. To identify EDUs (Elementary Discourse Units) and name their type of connections, the technique used HILDA parsers (H. Hernault 2010), followed by Henry's finance-specific vocabulary (E. Henry 2008). The work generated EDU sentiments and presented two options for averaging sentiment scores overall EDUs. At the document level, the first method used a weighting approach called node weighting with grid search. The second one transformed Rhetorical Structure data into characteristics for use in classification by means of RFs (Random Forests). Predicting stock returns was a breeze with the help of the presented method. Researchers should investigate approaches that take into account all levels of Rhetoric Structure trees for SA, as suggested by the authors of this paper.

PRIVACY RISK IN PUBLISHED SOCIAL NETWORK DATA

Data anonymization models, de-anonymization attacks, & anonymization process for social networks were discussed, along with the privacy risks associated with users' publicly available social network data. In this section, we also looked at the social network data model that an attacker would utilise to launch assaults on the privacy of users' public social media profiles. An adversary can utilise data

from social networking sites that have been anonymized and made available to third-party users to violate users' privacy. De-anonymization attacks were conducted by an adversary in order to re-identify users in anonymised social networks.

Publishing of social network data

Social media platforms capture user profile & activity data (H. Li, Q. Chen 2020). User-generated content, such as posts, photographs, videos, or comments, might reveal private information about social media users; as a result, social networking companies store this information on their servers for profile maintenance. That's why all the information people post and share on social media platforms is accessible to those people. In order to facilitate data analysis and research, social networking companies have made the personal information of their users available to interested third parties. Social networking sites do not published real social network datasets. Before publishing user's data, social networking sites owners anonymized social networks data using conventional anonymized processes (like; k-anonymity, l-diversity, t-closeness. Anonymized social network data was made available by social media platforms to researchers in fields as diverse as epidemiology (J. Fang 2019), sociology (N. Desai 2021), and criminal justice (2021).

Social network data

Social media users share and post different type of contents on online social networking site and that data shared between multiple social media users. The data, that contain information of social media users and relationship between them, is social network data. Social network data that is released by social networking sites may have a social media group users comment, post or image like or views. A social networking service has issued a modified version of their anonymised user database, which still retains personally identifiable information. In published social network data user identity replaced by random attributes so that user can not be identified in published social networks (J. Ma, Y. Qiao 2018). Social networking sites owner anonymized social network datasets by using conventional anonymization techniques before publishing social networks data for the third party consumers.

SOCIAL NETWORK DATA THREAT ANALYSIS MODEL

Users must first make a profile on a social networking site before they can start utilising the site's services. User profiles on social media sites require users to reveal personally identifiable information and that is private and unique information of social media users. Social media users also have personal information (like; relationship status, workplace, political views) and sensitive

information (like; hobbies, comment, post) on social networking sites (X. Ding 2010). Social media user's private and personal information available on social networking sites in user's profile that is generated by user's activity on social media. Users create profiles on social networking sites so they may communicate with one another via instant message, email, and the sharing of media files like photos, videos, & music. Users' profiles on social networking websites include both publicly viewable information (such as their names, email addresses, genders, & current locations) and potentially sensitive information (such as their occupations, places of employment, relationships, religious beliefs, & political leanings). Interactions amongst different users are also providing a rich dataset. This data may be having sensitive information about users (e.g., hobbies, views and comments). Before publishing user's data, Social media service providers anonymized social networks data using conventional anonymized techniques (like; kanonymity, l-diversity, t-closeness), so that social media user's identity cannot disclose to the others. Social media service providers published anonymized social network dataset so third party consumers can download that anonymized social network data. Attackers also obtain anonymised social network data after the publication of the social network data, and they utilise various de-anonymized assaults to re-identify social media users in the anonymous social network data (J. Qian 2019).

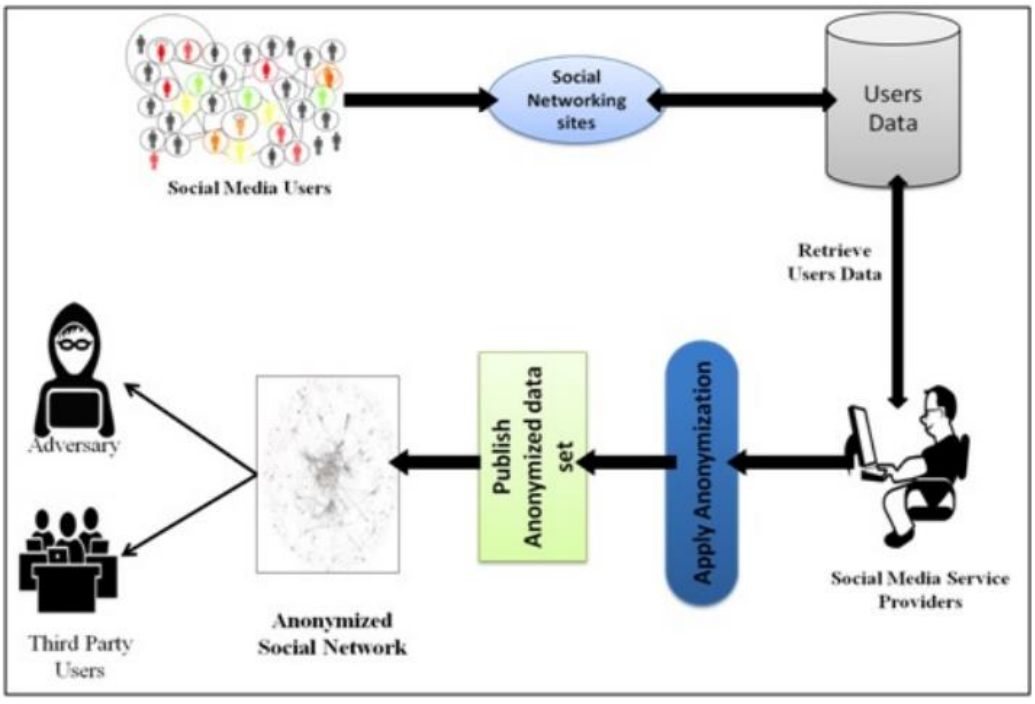

Figure 3: Social Network Threat Analysis Model

In Figure 3, the essential players in an online social network environment are depicted, including social media platforms, operators, users, third-party data recipients, & potential threats. In this

concept a social networking website owner may retrieve social media user's data & perform anonymization method over that data. Third-party researchers & academics can purchase access to this aggregated, anonymized social network data.

SOCIAL MEDIA ANALYTICS APPROACHES

The textual data acquired from social media accounts for a sizeable amount of the unstructured data. Information retrieval from unstructured textual data is accomplished through the use of text mining, NLP, ML, & big data processing techniques. The technology in this chapter focuses on methods & domains of social media analytics.

In their presentation of research on social data analysis, Kursuncu et al. (2019) incorporated real-time social intelligence. A wide range of topics, including political and electoral processes, social movements, crises & disasters, popular culture, and the environment have been the emphasis of their studies. In Spatio-Temporal-Thematic Analysis of Twitter Data, we perform TF-IDF, clustering, & semantic analysis of terms on the Twitter dataset. This method does not involve any sort of predictive analysis. In order to predict influenza epidemics utilising tweets from social media users, Achrekar et al. (2011) created a system called SNEFT. The frequency of flu-related tweets correlates strongly with Centers for Disease Control and Prevention data, with a Pearson correlation coefficient of 0.9847, according to their results. For influenza prediction, they built auto-regression models and evaluated their models with historical data. They found that Twitter data significantly improved their model's accuracy. Using data from Twitter, Gautam et al. (2014) suggested a set of machine learning approaches that incorporate semantic analysis to classify sentences and product evaluations. They used Naive Bayes, Maximum Entropy, and SVM classification algorithms together with the Semantic Orientation based on WordNet, which extracts synonyms and similarities for the content feature, to raise the model's accuracy from 88.2% to 89.9% . Using visual analytics for text-based Twitter data, Brooker et al. (2016) have developed a set of supporting approaches for examining Twitter data as a socio-technical assemblage.

Desai et al. (2015) used a decision tree algorithm to classify social media data and then compared it to Adaboost and Naive Baye. Twitter and UCI machine learning repositories are utilised as a dataset. The decision tree algorithm is more accurate than other algorithms. In addition, the random decision tree (RDT) algorithm performs better than other tree algorithms, such as C 4.5 or ID3, for the categorization of social media data.

LITERATURE REVIEW

Ding et al. (2011) investigated social network data using a d-neighborhood node attack. Here, d is the number of neighbouring nodes away from the desired node, and d is less than 1. With knowledge of the targeted user's neighbours and their relationships, an attacker might potentially utilise social network data to re-identify the victim. This thesis demonstrates the feasibility of the neighbourhood assault in practise and proposes a solution to the challenge of defending against a single such attack. The suggested approach can perform an isomorphic verification with the smallest possible DFS code. However, there is an exponentially growing number of alternative DFS trees for each component as d grows larger.

Yan Liu et al. (2022) The prediction of traffic flow has seen extensive usage of deep learning techniques. They can outperform shallow models by a wide margin. But the majority of deep learning models in use today only pay attention to deterministic data, oblivious to the fact that traffic flow contains a significant amount of uncertain data. In order to estimate short-term traffic flows, this research suggests a unique hybrid model called FGRU that combines a fuzzy inference system (FIS) & gated recurrent unit (GRU) neural network. The FIS compensates for the drawbacks of deep learning by reducing the impact of uncertain data, while the GRU model is used to capture the temporal relationships within traffic flow data. Additionally, a mechanism for temporal feature augmentation is suggested to determine the optimal time intervals for model inputs. Comparative experiments are used to investigate the best model structure & parameters. Furthermore, the simulation results demonstrate that FGRU's mean absolute error is 7.75% lower than ARIMA's & 3.05% lower than the most recent deep learning-based traffic flow prediction model than ARIMA.

Yuankai Wu et al. (2016) In the realm of artificial intelligence, deep learning techniques have attained a celebrity status, & CNN & Recurrent Networks have been key to their success. The CNN consistently performs dominantly on visual tasks by taking advantage of the basic spatial features of images and videos. Furthermore, the RNN, in particular LSTM, exhibit higher performance for time series tasks by accurately characterizing the temporal correlation. Data on traffic flow have several properties in the temporal and spatial domains. However, there are few applications of CNN & LSTM techniques to traffic flow. To predict future traffic flow, we provide a revolutionary deep architecture that combines CNN & LSTM in this paper (CLTFP). Two LSTMs are used to mine the short-term variability & periodicities of traffic flow, while a 1-dimension CNN is used to capture spatial aspects of traffic flow. Short-term forecasting is accomplished through feature-level fusion based on those significant features. On available datasets, the proposed CLTFP is contrasted with various well-known forecasting techniques. According to experimental findings, the CLTFP has

significant advantages in anticipating traffic flow. In particular, the suggested CLTFP is examined from the perspective of Granger Causality, and a number of intriguing CLTFP characteristics are found and explored.

Chencheng Ma et al. (2019) Anomaly detection for network traffic is the primary method to address the growing security concerns of large-scale local area networks. However, it can be difficult to extract precise and useful traffic data for anomaly identification. The current study is meant to address this problem by designing & analyzing various network flow aspects. These features, which precisely characterize the properties of network flows, comprise sequence packet attributes, general statistical features, and environmental factors. Additionally, a method using a hybrid neural network is developed to analyze these properties and find anomalies. While deep neural networks are used to learn the properties of high-dimensional feature vectors, comprising general statistical features & environmental information, ID CNN are employed to assess the sequence features in the hybrid neural network. The technique can do a thorough analysis to find network anomalies. To assess how well the suggested approach and other related algorithms work, two datasets ISCX-IDS-2012 & CIC-IDS are used. The current analysis demonstrates that the proposed method's overall performances are superior to those of other methods. The proposed method can be used for anomaly detection applications with acceptable performance, it is determined.

Syed Rizvi et al. (2022) Network intrusion detection systems (IDS) scan network packets for security breaches and notify system administrators & investigators when a breach is detected. These reports become unmanageable in large networks. There are various obstacles to be addressed in order to develop a flexible & effective IDS for unanticipated attacks. Deep learning approaches have been used extensively in IDS, although frequently significant computational resources & processing time are needed. In this study, an IDS based on a 1D-Dilated Causal Neural Network (1D-DCNN) is used for binary classification. In order to compensate for the maximum pooling layer & stop the information loss caused by pooling & downsampling, dilated convolution with a dilation rate of two is implemented. Additionally, the dilated convolution might widen its receptive field to gather more contextual information. Experiments were carried out on the CIC-IDS2017 and CSE-CIC-IDS2018 datasets, two well-known publically available datasets, to evaluate the effectiveness of the proposed method. According to simulation results, the 1D-DCNN based method is more accurate than several other deep learning techniques currently in use. With a malicious attack detection rate accuracy of 99.7% for CIC-IDS2017 & 99.98% for CSE-CIC-IDS2018, the suggested model reached a high level of precision.

CONCLUSION

In conclusion, the technology underscores the significance of leveraging machine learning hybrid models for social media threat detection and prediction. The rapid expansion and diversification of social media threats demand sophisticated and adaptable solutions, and the synthesis of various machine learning techniques in hybrid models proves to be a promising avenue. future research should focus on refining existing hybrid models, addressing identified gaps, and exploring innovative avenues for improving the efficacy of social media threat detection systems. The interdisciplinary nature of this field invites collaboration between experts in machine learning, cybersecurity, and social sciences to develop holistic solutions that not only detect threats accurately but also contribute to the creation of safer and more inclusive online spaces.

REFERENCES

1. A.I. Kim Boes, Dimitrios Buhalis, P.F. Wilkinson, An Empirical Study on the Relationship between Twitter Sentiment and Influence in the Tourism Domain, Ann. Tour. Res. 28 (2012) 1070–1072. doi:10.1016/S0160-7383(01)00012-3.
2. Achrekar, A. Gandhe, R. Lazarus, S. H. Yu, and B. Liu, "Predicting flu trends using twitter data," In IEEE conference on computer communications workshops,2011, pp. 702-707.
3. Barnes, R. Klinger, S.S. im Walde, Projecting Embeddings for Domain Adaptation: Joint Modeling of Sentiment Analysis in Diverse Domains, (2018).
4. Brooker, J. Barnett, and T. Cribbin, "Doing social media analytics," Big Data & Society, vol. 3, no. 2, pp. 1-12, 2016.
5. Cambria, S. Poria, D. Hazarika, K. Kwok, SenticNet 5: Discovering Conceptual Primitives for Sentiment Analysis by Means of Context Embeddings, Aaai. (2018) 1795–1802.
6. Desai and M. L. Das, “Desan: De-anonymization against background knowledge in social networks,” in 2021 12th International Conference on Information and Communication Systems (ICICS), 2021, pp. 99–105.
7. Ding, L. Zhang, Z. Wan, and M. Gu, “A brief survey on de-anonymization attacks in online social networks,” in 2010 International Conference on Computational Aspects of Social Networks, 2010, pp. 611–615.
8. E. Henry, Are investors influenced by how earnings press releases are written?, J. Bus. Commun. 45 (2008) 363–407. doi:10.1177/0021943608319388.
9. E.Cambria, Affective Computing and Sentiment Analysis, IEEE Intell. Syst. 31 (2016) 102–107. doi:doi: 10.1109/MIS.2016.31

10. Fang, A. Li, Q. Jiang, S. Li, and W. Han, “A structure-based de-anonymization attack on graph data using weighted neighbor match,” in 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC), 2019, pp. 480– 486.
11. Gautam, and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," In Seventh EEE international conference on contemporary computing, 2014, pp. 437-442.
12. H. Peng, Y. Ma, Y. Li, E. Cambria, Learning multi-grained aspect target sequence for Chinese sentiment analysis, Knowledge-Based Syst. 148 (2018) 55–65. doi: 10.1016/j.knosys.2018.02.034.
13. Hernault, H. Prendinger, D.A. DuVerle, M. Ishizuka, HILDA: A discourse parser using Support Vector Machine classification, Dialogue & Discourse. 1 (2010) 1–33. doi:10.5087/dad.2010.003.
14. J. Ma, Y. Qiao, G. Hu, Y. Huang, A. K. Sangaiah, C. Zhang, Y. Wang, and R. Zhang, “De-anonymizing social networks with random forest classifier,” IEEE Access, vol. 6, pp. 10 139–10 150, 2018.
15. J. Märkle-Huß, S. Feuerriegel, H. Prendinger, Improving Sentiment Analysis with Document-Level Semantic Relationships from Rhetoric Discourse Structures, Proc. 50th Hawaii Int. Conf. Syst. Sci. (2017) 1142–1151.
16. J. Qian, X.-Y. Li, C. Zhang, L. Chen, T. Jung, and J. Han, “Social network de-anonymization and privacy inference with knowledge graph model,” IEEE Transactions on Dependable and Secure Computing, vol. 16, no. 4, pp. 679–692, 2019.
17. Kursuncu, M. Gaur, U. Lokala, K. Thirunarayan, A. Sheth, and I.B. Arpinar, “Predictive analysis on Twitter: Techniques and applications,” In Emerging research challenges and opportunities in computational social network analysis and mining, 2019, pp. 67-104.
18. Liu, Sentiment Analysis and Opinion Mining, Synth. Lect. Hum. Lang. Technol. 5 (2012) 1–167. doi:10.2200/S00416ED1V01Y201204HLT016.
19. Liu, Y., Wang, X. K., Hou, W. H., Liu, H., & Wang, J. Q. (2022). A novel hybrid model combining a fuzzy inference system and a deep learning method for short-term traffic flow prediction. Knowledge-Based Systems, 255, 109760.
20. M. Hagenau, M. Liebmann, D. Neumann, Automated news reading: Stock price prediction based on financial news using context-capturing features, Decis Support Syst. 55 (2013) 685–697. doi: 10.1016/j.dss.2013.02.006.
21. Ma, C., Du, X., & Cao, L. (2019). Analysis of multi-types of flow features based on hybrid neural network for improving network anomaly detection. IEEE Access, 7, 148363-148380.

22. Maks, P. Vossen, A lexicon model for deep sentiment analysis and opinion mining applications, in: Decis. Support Syst., 2012: pp. 680–688. doi: 10.1016/j.dss.2012.05.025.
23. P.M. Sosa, Twitter Sentiment Analysis using combined LSTM-CNN Models, (2017) 1–9.
24. Pak, P. Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining, Proc. Seventh Conf. Int. Lang. Resour. Eval. (2010) 1320–1326. doi: 10.1371/journal.pone.0026624.
25. Rizvi, S., Scanlon, M., McGibney, J., & Sheppard, J. (2022). Deep learning based network intrusion detection system for resource-constrained environments. In Springer (pp. 1-7).
26. S. Poria, E. Cambria, A. Gelbukh, Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis, Proc. 2015 Conf. Empir. Methods Nat. Lang. Process. (2015) 2539–2544.
27. S. Poria, E. Cambria, G. Winterstein, G. Bin Huang, Sentic patterns: Dependency-based rules for concept-level sentiment analysis, KnowledgeBased Syst. 69 (2014) 45–63. doi:10.1016/j.knosys.2014.05.005.
28. Wu, Y., & Tan, H. (2016). Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv preprint arXiv:1612.01022.
29. Y. Yu, H. Lin, J. Meng, Z. Zhao, Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks, Algorithms. 9 (2016) 41. doi:10.3390/a9020041.
30. Z. Zhao, S. Zhang, Z. Xu, K. Bellisario, N. Dai, H. Omrani, and B. C. Pijanowski, "Automated bird acoustic event detection and robust species classification." Ecological Informatics, vol.39, pp. 99-108, 2017
31. Zhang, T. Yoshida, X. Tang, T.B. Ho, Improving effectiveness of mutual information for substantival multiword expression extraction, Expert Syst. Appl. 36 (2009) 10919–10930. doi:10.1016/j.eswa.2009.02.026.