Dark Side of the Web: Dark Web Classification Based on TextCNN and Topic Modelling Weight

Ms. V. Pavithra [1], Sriram Gnana Brahma Sanath Kumar [2], Vuribindi Sasidhar Reddy [3], Seela Vishnu [4], Ramanaboina Sravan Kumar [5],

Assistant Professor [1], Student [2][3][4][5], Department of Computer Science and Engineering, Geethanjali Institute of Science and Technology, Nellore, Andhra Pradesh-524137

Abstract: The Dark Web, a concealed internet domain ensuring user anonymity, has become a hub for illegal activities and a source of cyberattack information due to its untraceable nature. Utilizing the "DARK NET: Light in the Dark" dataset, the research focuses on classifying Dark Web services related to attacks. A comparative analysis was conducted using K-Nearest Neighbors (KNN), Random Forest, and a Text Convolutional Neural Network (TextCNN) integrated with Latent Dirichlet Allocation (LDA) topic modeling. Topic modeling weights, which assign higher significance to frequently co-occurring terms, were employed to enhance the feature set and improve classification accuracy. Additionally, a hybrid model combining LDA topics with Hybrid TextCNN was developed, significantly outperforming other methods. The hybrid approach achieved an accuracy of 96.1%, demonstrating its effectiveness in accurately predicting Dark Web services associated with cyber threats. The study emphasizes the potential of deep learning and topic modeling in addressing challenges posed by the Dark Web while contributing to proactive cybersecurity measures.

"Index Terms – *Dark web, dark web analysis, text classification, topic modeling, model explanation".*

1. INTRODUCTION

The global adoption of internet services, combined with the widespread availability of affordable devices, has led to an explosion of data exchange worldwide. While the surface web appears to be an extensive resource, it represents only a small fraction of the entire internet. The surface web makes up less than 10% of the total web, with the majority of the internet residing in the Hidden Web, also referred to as the Deep Web. The Deep Web consists of a vast and diverse collection of data, including academic databases, private communication platforms, and sensitive financial information, all of which are not indexed by traditional search engines [1]. Among the Deep Web, the Dark Web is a distinct subset that has gained notoriety due to its association with illegal activities. This part of the internet is intentionally concealed from the general public and can only be accessed through specialized browsers like Tor, I2P, and Freenet, which anonymize users' online presence, offering a cloak of invisibility to those seeking to hide their identities [2].

While anonymity was once considered a positive feature of the Dark Web, it has increasingly become a shield for illegal operations. With encrypted communication and concealed identities, the Dark Web has become a haven for various unlawful activities. Research has shown that over 57% of malevolent activities online, such as drug

trafficking, illegal pornography, identity theft, and hacking, can be traced back to the Dark Web [3]. The consequences of these activities extend beyond the digital realm, affecting real-world security and safety. For instance, malware like ransomware and trojans, often disseminated through phishing emails, is regularly traded on the Dark Web. These malicious tools are often used to facilitate attacks on unsuspecting individuals or organizations, leading to significant financial and reputational damage [4].

The anonymity provided by the Dark Web emboldens users to engage in illicit activities without the fear of being traced. Criminals use this platform to buy and sell stolen personal data, illicit goods, and even services related to cybercrime. Furthermore, cryptocurrencies, which offer an additional layer of privacy, are frequently used as payment methods for transactions conducted on the Dark Web [5]. This secrecy ensures that users remain largely unaccountable for their actions, making it a thriving marketplace for cybercriminals. Despite ongoing efforts by law enforcement agencies to monitor and shut down illicit activities on the Dark Web, the platform continues to serve as a breeding ground for harmful practices, posing significant challenges to global security [6][7]. As such, understanding the scope of activities on the Dark Web and developing effective detection and prevention mechanisms has become critical in the fight against online crime.

2. RELATED WORK

The growing prevalence of illicit activities on the Dark Web has garnered significant attention in recent years, prompting research efforts to investigate, understand, and mitigate the threats originating from this hidden part of the internet. Several studies have contributed to the understanding of Dark Web dynamics, ranging from investigating cyber threats to developing techniques for identifying and characterizing illicit activities.

Basheer and Alkhatib [8] provide a comprehensive review of research related to Dark Web investigations, specifically focusing on cyber threat intelligence. They explore various methods for analyzing and monitoring malicious activities, providing valuable insights into the challenges and strategies associated with Dark Web threat detection. Their work emphasizes the need for improved tools and approaches to monitor the rapidly evolving landscape of cybercrime on the Dark Web. They also highlight the role of Dark Web marketplaces in facilitating criminal activities, which has led to the development of techniques aimed at understanding the topological properties of Dark Web platforms.

Alharbi et al. [9] focus on the topological characteristics of the Tor network, which is a primary access point to the Dark Web. Their study delves into the structure of the Tor network and its implications for Dark Web investigations. They analyze the interaction patterns of users and services within Tor, offering an understanding of how the Dark Web operates at a network level. This analysis contributes to developing more effective strategies for tracking and identifying malicious actors within this anonymized network. By understanding these topological properties, their work lays the groundwork for future advancements in Dark Web monitoring.

Hayes et al. [10] propose a framework for improving Dark Web marketplace investigations. Their approach outlines a systematic method for collecting and analyzing data from Dark Web marketplaces, which are notorious for facilitating illegal transactions. The framework they introduce is designed to enhance the efficiency and effectiveness of investigations into these marketplaces by using

automated data collection techniques and advanced analysis tools. This work emphasizes the importance of having a structured approach to Dark Web investigations, as manual methods are often time-consuming and inadequate for dealing with the scale of activity on the Dark Web.

Sabbah et al. [11] introduce a hybridized term-weighting method for classifying Dark Web content. This method aims to enhance the accuracy of content classification on the Dark Web, where distinguishing between legitimate and malicious content is challenging due to the volume and anonymity of data. Their hybridized approach combines multiple term-weighting techniques to improve the effectiveness of classification models. This work contributes to the growing body of research focused on automated classification of Dark Web content, providing more accurate and reliable identification of illegal activities.

He et al. [12] focus on the classification of illegal activities on the Dark Web. They propose a classification system designed to categorize Dark Web content based on the nature of the illicit activities associated with it, such as drug trafficking, human trafficking, and financial fraud. Their work contributes to understanding the types of criminal enterprises that thrive on the Dark Web and provides a basis for developing detection systems that can automatically identify such activities. The classification system they propose offers a structured approach to categorizing and analyzing Dark Web content, a critical step for law enforcement agencies and cybersecurity professionals.

Nazah et al. [13] introduce an unsupervised model for identifying and characterizing Dark Web forums. Their model is designed to detect and analyze forums on the Dark Web, which serve as hubs for illegal discussions and exchanges. The unsupervised nature of their approach allows it to operate without needing labeled data, making it more flexible and applicable to a wider range of forums. By characterizing these forums, their work provides a deeper understanding of the social dynamics and structures within the Dark Web, which is crucial for identifying and investigating criminal networks operating in these spaces.

Alnabulsi and Islam [14] focus on identifying illegal forum activities within the Dark Net. They propose a method for detecting and analyzing illegal activities that occur within Dark Web forums, including the exchange of illicit goods and services. Their work highlights the importance of monitoring these forums, as they are often used to coordinate and facilitate criminal enterprises. The techniques they introduce aim to enhance the detection of illegal activities, providing valuable insights for law enforcement agencies and cybersecurity researchers working to combat crime on the Dark Web.

Finally, Cascavilla et al. [15] take a unique approach by combining software quality metrics with Dark Web threat intelligence. Their research explores how software quality metrics can be used to detect and analyze threats within Dark Web code, such as malware and ransomware. By examining the quality of software used for illicit activities on the Dark Web, their work provides a novel perspective on how to understand and mitigate the threats originating from this space. Their approach offers a valuable tool for detecting malicious code and preventing its spread, contributing to a broader effort to secure the Dark Web.

3. MATERIALS AND METHODS

The proposed system aims to classify Dark Web services associated with cyber threats by leveraging advanced machine learning and deep learning

techniques. Using the "DARK NET: Light in the Dark" dataset, the system employs K-Nearest Neighbors (KNN), Random Forest, Text Convolutional Neural Network (TextCNN), and a hybrid model combining Latent Dirichlet Allocation (LDA) topics with TextCNN. Topic modeling weights are incorporated to enhance feature representation by assigning higher importance to frequently co-occurring terms. These weights improve the data's contextual understanding, enabling more accurate classification. The integration of LDA topics with Hybrid TextCNN architecture enhances the system's ability to extract meaningful features, leading to better identification and classification of Dark Web services associated with potential cyber threats.

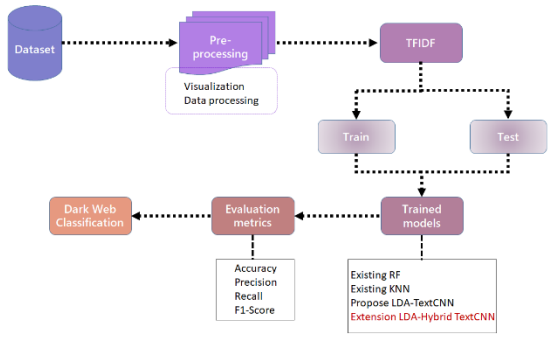


Fig.1 Proposed Architecture

The system architecture (Fig.1) depicted in the image architecture begins with a dataset that undergoes pre-processing steps like visualization and data processing. This data is then converted into TF-IDF representation. This processed data is then split into training and testing sets. Various models are trained on the training data: existing RF, existing KNN, proposed LDA-TextCNN, and an LDA-Hybrid TextCNN. These trained models are then used to classify dark web content. The performance of these models is evaluated using metrics such as accuracy, precision, recall, and F1-score.

i) Dataset Collection:

The dataset "DARK NET: Light in the Dark" contains information about various Dark Web transactions, including details on vendors, categories, items, prices, and ratings. It comprises 15052 rows and 9 columns, with data on product descriptions, origins, destinations, and the associated cryptocurrency value. The dataset provides valuable insights into illegal activities, with entries covering various categories like drugs, electronics, and other illicit goods, mostly originating from the United States and distributed globally.

	Vendor	Category	Item	Item_Description	Price	Origin	Destination	Rating
1027	lonelyimage	Services/Money	List of Cardable sites	List of cardable sites for those people who wa...	0.022200487096774184 BTC	Internet	NaN	~4/5
1028	hansberger	Services/Money	Hacked Bitcoin Mining Cloudmining 1This Anonymous	Hacked Bitcoin Cloud Mining 0 5BTC/Week for 2...	6.2999999999999945 BTC	NaN	NaN	[0 deals]
1029	Currency	Services/Money	20dollarscash£/$DC1x 20 bills)	2/24/2014 Welcome to Cryptonaut currency serv...	0.037490000000000016 BTC	Canada	NaN	[0 deals]
1030	lonelyimage	Services/Money	* Legit * Tesco Online Codes - ÂŁ100	Tesco ONLINE CODES - ÂŁ100 - 100% WORKING Gre...	0.14169786107142865 BTC	Internet	NaN	~4/5
1031	HellMaker	Services/Money	custom listing for 3-5k wells	custom listing for 3-5k wells	0.26207431687499994 BTC	NaN	NaN	4.43/5
...	...	...	...	...	...	...	...	...
109604	beige4you	Drugs/Opioids	30mg BlueBerries (Roxy/Oxycodone) x5	*****PGP ONLY***** ****FE ONLY***** We are ...	0.17323803 BTC	United	States	~5/5
109605	beige4you	Drugs/Opioids	1/4g (250mg) RAW UNCUT East Coast #4	*****PGP ONLY.***** ****FE ONLY.***** We are ...	0.093282015 BTC	United	States	~5/5
109606	AfghanApothecary	Drugs/Opioids	1/2 GRAM GOOD BROWN HEROIN	UK ONLY 500MG OR 1/2 GRAM BAGS OF HEROIN GOO...	0.079956015 BTC	NaN	NaN	[0 deals]
109607	AfghanApothecary	Drugs/Opioids	100 MCG/HR FENTANYL PATCH	TRANSDERMAL FENTANYL PATCH HIGHEST DOSE RAPID ...	0.16657503 BTC	NaN	NaN	[0 deals]
109608	AfghanApothecary	Drugs/Opioids	USA ONLY...1/2G TOP QUALITY #3 BROWN HEROIN	A SPECIAL TREAT FOR THE USA 500MG OF TOP GRAD...	0.131927425 BTC	USA	NaN	[0 deals]

15052 rows × 9 columns

Fig.2 Dataset Collection Table

ii) Pre-Processing:

Preprocessing involves data cleaning and transformation, including handling missing values and filtering irrelevant entries. Visualization techniques are used to explore patterns, followed by feature extraction using TF-IDF (Term Frequency-Inverse Document Frequency) to convert textual data into numerical form for analysis.

a) Visualization: The visualization displays the distribution of class labels in the dataset, showing the count of occurrences for each label. A bar chart is used to represent the frequency of different Dark Web categories, with each label plotted along the x-

axis and its corresponding count along the y-axis. The chart is designed to provide insights into the class distribution, helping to identify the prevalence of specific categories within the dataset, with labels rotated for clarity.

b) Data processing: Data processing involves loading and cleaning the service descriptions from the dataset. Each description is stripped of unnecessary characters and converted to lowercase. Descriptions longer than 100 characters are processed further to remove irrelevant elements. The cleaned text data is then paired with corresponding labels based on the category and stored for further use. This process ensures that the dataset is refined and ready for modeling, providing a structured input for subsequent analysis.

c) TFIDF: TF-IDF (Term Frequency-Inverse Document Frequency) is used to convert the text data into numerical form for analysis. The text is transformed into a matrix, where each word's importance in the dataset is measured. This transformation helps capture the relevance of terms within the documents while reducing the impact of frequently occurring but less informative words. The transformed data is then saved for future use in the model, ensuring efficient processing and analysis.

D) Training and Testing: The dataset is split into training and testing subsets, with 80% of the data allocated for training the algorithms and 20% reserved for testing. This division ensures that the model is trained on a substantial amount of data while still maintaining a separate set for evaluating its performance. The training set contains 10,112 samples, while the testing set includes 2,528 samples, providing a balanced approach for model validation.

iii) Algorithms:

K-Nearest Neighbors (KNN) algorithm classifies Dark Web services by measuring feature proximity to labeled training data. It predicts service categories based on nearest neighbors, serving as a baseline for evaluating performance improvements of advanced models like LDA-TextCNN.

Random Forest leverages multiple decision trees to classify Dark Web services, effectively capturing complex patterns while reducing overfitting. This ensemble method provides robust predictions, acting as a benchmark for comparing advanced techniques like LDA-based deep learning models.

LDA-TextCNN model integrates Latent Dirichlet Allocation (LDA) for topic modeling with a Convolutional Neural Network (CNN). Topic weights enhance feature representation, allowing TextCNN to classify Dark Web services with improved accuracy and semantic understanding.

LDA-Hybrid TextCNN enhances the LDA-TextCNN by introducing an additional CNN2D layer for refined feature extraction. Dropout layers mitigate overfitting, resulting in a more robust and accurate classification of Dark Web services compared to earlier algorithms.

4. RESULTS & DISCUSSION

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} (1)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the

ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (2)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

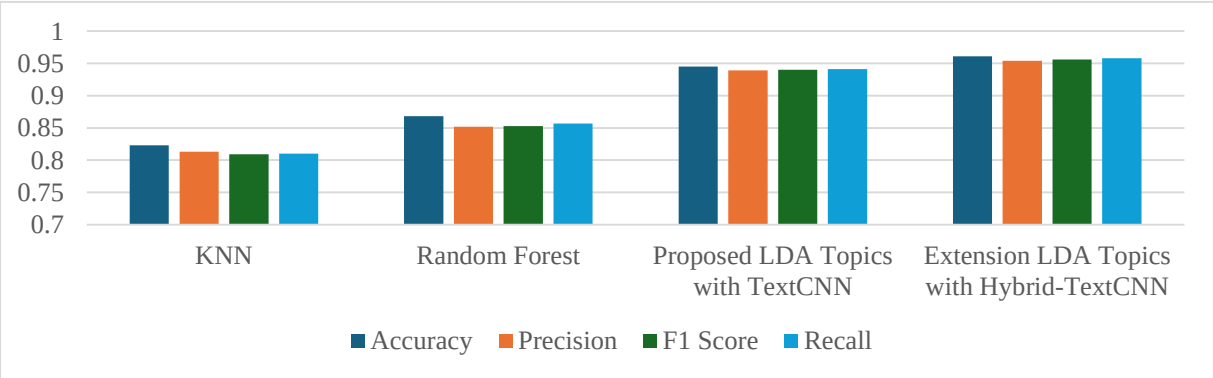
$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100 (1)$$

Table (1) evaluate the performance metrics—Accuracy, precision, recall and F1-score —for each algorithm. The LDA Topics with Hybrid-Text CNN consistently outperforms compared to all other algorithms. The tables also offer a comparative analysis of the metrics for the other algorithm.

Table.1 Performance Evaluation Table

ML Model	Accuracy	Precision	F1 Score	Recall
KNN	0.823	0.813	0.809	0.810
Random Forest	0.868	0.852	0.853	0.857
Proposed LDA Topics with TextCNN	0.945	0.939	0.940	0.941
Extension LDA Topics with Hybrid-TextCNN	**0.961**	**0.954**	**0.956**	**0.958**

Graph.1 Comparison Graphs



Accuracy in blue, precision is represented in orange, recall in green and F1-Score in Sky blue in ***Graph (1)***. In comparison to the other models, the LDA Topics with Hybrid-TextCNN Model shows superior performance, achieving the highest values. The graphs above visually illustrate these findings.

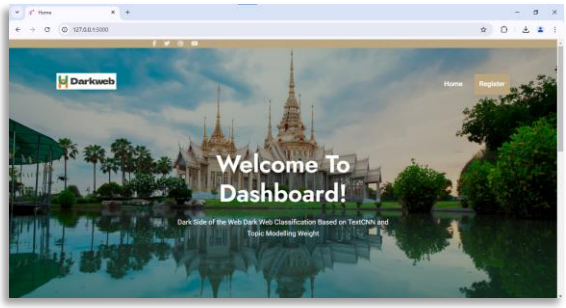


Figure.3 User interface

In the above figure 3, this is a user interface dashboard for a welcome message.

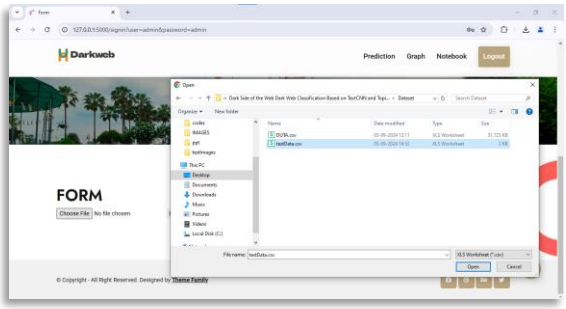


Figure.4: Upload data for checking

In the above figure 4, this is a user input screen, used to upload data for testing.

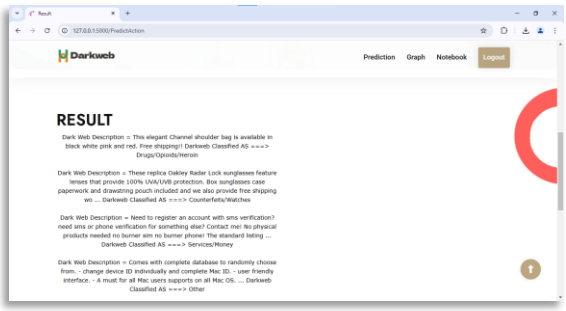


Figure.5: Prediction result

In the above figure 5, this is a result screen, in this user get prediction for uploaded data.

5. CONCLUSION

The exploration of Dark Web service classification emphasized the application of advanced deep learning techniques to identify potential threats with high precision. The study utilized the "DARK NET: Light in the Dark" dataset and incorporated topic modeling weights to enhance the feature representation. The LDA-Hybrid TextCNN model emerged as the most effective approach, achieving a remarkable accuracy of 96%. This model leveraged the strengths of topic modeling to highlight significant co-occurring terms while integrating them into a robust convolutional neural network framework, enabling superior feature extraction and classification capabilities. The impressive performance of the LDA-Hybrid TextCNN underscores its potential as a powerful tool for analyzing and classifying complex and anonymous Dark Web services, which are often associated with cyber threats. By achieving this high level of accuracy, the model showcases its ability to contribute to proactive threat identification and cybersecurity measures, offering a significant step toward mitigating risks posed by activities within the Dark Web.

The promising results of the LDA-Hybrid TextCNN model open avenues for further exploration in enhancing Dark Web service classification. Future efforts could focus on expanding the dataset to include more diverse and emerging threat patterns, integrating real-time analysis capabilities, and refining topic modeling techniques for better context understanding. Additionally, exploring ensemble approaches and incorporating advanced natural language processing methods could further improve detection accuracy and adaptability to evolving threats.

REFERENCES

[1] C. A. S. Murty and P. H. Rughani, "Dark Web text classification by learning through SVM optimization," J. Adv. Inf. Technol., vol. 13, no. 6, 2022, doi: 10.12720/jait.13.6.624-631.

[2] A. H. M. Alaidi, R. M. Al Airaji, H. T. S. Alrikabi, I. A. Aljazaery, and S. H. Abbood, "Dark Web illegal activities crawling and classifying using data mining techniques," Int. J. Interact. Mobile Technol., vol. 16, no. 10, pp. 122–139, May 2022, doi: 10.3991/ijim.v16i10.30209.

[3] N. Deguara, J. Arshad, A. Paracha, and M. A. Azad, "Threat miner—A text analysis engine for threat identification using dark Web data," in Proc. IEEE Int. Conf. Big Data, Dec. 2022, pp. 3043–3052, doi: 10.1109/Big Data55660.2022.10020397.

[4] Y. Jin, E. Jang, Y. Lee, S. Shin, and J.-W. Chung, "Shedding new light on the language of the dark Web," 2022, arXiv:2204.06885.

[5] H. Ma, J. Cao, B. Mi, D. Huang, Y. Liu, and Z. Zhang, "Dark Web traffic detection method based on deep learning," in Proc. IEEE 10th Data Driven Control Learn. Syst. Conf. (DDCLS), May 2021, pp. 842–847, doi: 10.1109/DDCLS52934.2021.9455619.

[6] K. N. Singh, S. D. Devi, H. M. Devi, and A. K. Mahanta, "A novel approach for dimension reduction using word embedding: An enhanced text classification approach," Int. J. Inf. Manage. Data Insights, vol. 2, no. 1, Apr. 2022, Art. no. 100061, doi: 10.1016/j.jjimei.2022.100061.

[7] W. Alkhatib, C. Rensing, and J. Silberbauer, "Multi-label text classification using semantic features and dimensionality reduction with autoencoders," in Proc. 1st Int. Conf. Lang., Data, Knowl., 2017, pp. 380–394, doi: 10.1007/978-3-319-59888-8_32.

[8] R. Basheer and B. Alkhatib, "Threats from the dark: A review over dark Web investigation research for cyber threat intelligence," J. Comput. Netw. Commun., vol. 2021, pp. 1–21, Dec. 2021, doi: 10.1155/2021/1302999.

[9] A. Alharbi, M. Faizan, W. Alosaimi, H. Alyami, A. Agrawal, R. Kumar, and R. A. Khan, "Exploring the topological properties of the tor dark Web," IEEE Access, vol. 9, pp. 21746–21758, 2021, doi: 10.1109/ACCESS.2021.3055532.

[10] D. Hayes, F. Cappa, and J. Cardon, "A framework for more effective dark Web marketplace investigations," Information, vol. 9, no. 8, p. 186, Jul. 2018, doi: 10.3390/info9080186.

[11] T. Sabbah, A. Selamat, M. H. Selamat, R. Ibrahim, and H. Fujita, "Hybridized term-weighting method for dark Web classification," Neurocomputing, vol. 173, pp. 1908–1926, Jan. 2016, doi: 10.1016/j.neucom.2015.09.063.

[12] S. He, Y. He, and M. Li, "Classification of illegal activities on the dark Web," in Proc. 2nd Int. Conf. Inf. Sci. Syst., Mar. 2019, pp. 73–78, doi: 10.1145/3322645.3322691.

[13] S. Nazah, S. Huda, J. H. Abawajy, and M. M. Hassan, "An unsupervised model for identifying and characterizing dark Web forums," IEEE Access, vol. 9, pp. 112871–112892, 2021, doi: 10.1109/ACCESS.2021.3103319.

[14] H. Alnabulsi and R. Islam, "Identification of illegal forum activities inside the dark net," in Proc. Int. Conf. Mach. Learn. Data Eng. (iCMLDE), Dec. 2018, pp. 22–29.

[15] G. Cascavilla, G. Catolino, F. Ebert, D. A. Tamburri, and W. J. van den Heuvel, "'When the code becomes a crime scene' towards dark threat

intelligence with software quality metrics," in Proc. IEEE Int. Conf. Softw. Maintenance Evol. (ICSME), Oct. 2022, pp. 439–443, doi: 10.1109/ICSME55016.2022.00055.