



LDA-Driven TextCNN with CNN2D Enhancement for Accurate Dark Web Service Classification

K.Pavani¹, M.Priyanka², P.Manohar Reddy³

#1 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract: The Dark Web is an internet domain that ensures user anonymity and has increasingly become a focal point for illegal activities and a repository for information on cyberattacks due to the challenges in tracking its users. This paper employs a deep learning algorithm to predict Dark Web services primarily used for attacks, enabling precautionary measures. To enhance prediction accuracy, we utilize Topic Modelling weights as training features, assigning higher weights to frequently co-occurring words. Traditional algorithms like TF-IDF, Document Matrix, and Latent Semantic Analysis often fail to exclude irrelevant data, which hampers machine learning accuracy. Our approach involves data collection dataset, preprocessing to clean the data, and applying Latent Dirichlet Allocation (LDA) to extract 90 topic weights. The LDA weights are then input into a proposed TextCNN algorithm for classification. We compare the performance of this algorithm against KNN, Random Forest, and others, achieving a prediction accuracy of 95%. A hybrid model utilizing features from TextCNN and trained with a CNN2D algorithm further enhances accuracy to 96%, showcasing the effectiveness of the proposed methods for Dark Web classification.

Index terms - Dark Web Classification, TextCNN, Topic Modeling, Latent Dirichlet Allocation (LDA), Deep Learning, TF-IDF, CNN2D, Dropout Layer, Dark Web Services Detection, Cybersecurity.

1. INTRODUCTION

Because of its secrecy, the Dark Web acts as a hub for both legitimate and illicit activity. Cybercrimes such as unlawful trade, hacking, and data breaches also exploit it, despite its use for private correspondence and privacy protection. Finding and classifying Dark Web services is crucial for cybersecurity because it makes it easier to monitor and halt undesired behavior. However, standard machine learning approaches struggle to accurately categorize Dark Web content due to its complex linguistic patterns and irrelevant data.

This work proposes a DL classification model based on TextCNN and Topic Modelling weights to address these problems. Unlike conventional approaches like TF-IDF, Document Matrix, and Latent Semantic Analysis, which are unable to effectively eliminate unnecessary phrases, our method employs Latent Dirichlet Allocation (LDA) to assign frequently recurring words higher weights. The proposed



technique improves classification accuracy by combining TextCNN with topic weights obtained by LDA.

This study includes many phases, including data collection, preprocessing, feature extraction using LDA, and classification using TextCNN. A Kaggle dataset including over 10,000 Dark Web services is used for training and evaluation. Experimental results show that selecting the top 90 topic features significantly improves classification performance. A Dropout layer is added to an expanded CNN2D model to better optimize feature selection and increase overall accuracy. Comparing the suggested method with other models, such as KNN and Random Forest, shows how effective it is at correctly detecting Dark Web services.

2. LITERATURE SURVEY

a) Dark Web Text Classification by Learning through SVM Optimization:

The Darkweb keeps the most forbidden stuff because of its isolation and secrecy. Because of its secrecy and lack of control, the dark web is ideal for illicit activity, which boosts traffic and the growth of onion-based URLs. Cybercrime investigators worldwide must classify dark web data in order to monitor and classify URLs containing unlawful activity by feature engineering, given the increasing number of persons using the dark web. In this work, optimization approaches were used to classify dark web data using Support Vector Machines (SVM). Feature engineering was used to extract text-based keywords, and support vector machines (SVMs) were used to evaluate the system's performance. Good results of 0.83 for Precision, 0.90 for Recall, and 0.96 for F-measure were obtained using 1800 URLs.

b) Dark Web Illegal Activities Crawling and Classifying Using Data Mining Techniques:

The phrase "dark web" describes the whole of the Internet, where unidentified people or organizations engage in illicit activities that are difficult to locate. The illicit content on the dark web is ever-evolving. It takes a lot of time and work to compile and classify illegal conduct. Both academia and industry have recently recognized this topic's basic significance. This study suggests an automated way to collect, clean, and store dark web pages in a document database. Web pages are automatically categorized into five groups by the crawler. Numerous classifiers, including SVC, NB, and TF-IDF, were used to categorize the pages. The results of the trials indicated that the accuracy was 92% for SVC and 81% for NB.

c) Threat Miner - A Text Analysis Engine for Threat Identification Using Dark Web Data:

Complex cyberattacks on computer systems demand sophisticated defenses. Many state-of-the-art cyber security solutions keep an eye on network or system activity. When hostile actors use the dark web to disseminate vulnerabilities, breaches, and data dumps, it becomes appealing to use this knowledge to create defenses. This article explains our text mining engine, Threat Miner, and how it analyzes articles from dark web forums to provide useful data. Threat Miner uses state-of-the-art machine learning and a customized threat vocabulary to collect valuable information from dark web forums and convert it into STIX format for threat intelligence systems. We also provide the findings of a comprehensive assessment of our approach, which was done using the CrimeBB



dataset [1] to ascertain its viability and effectiveness in enhancing cyber defenses.

d) **Shedding New Light on the Language of the Dark Web:**

The Dark Web is infamously hard to analyze, especially in terms of language, because it is anonymous and does not have easily accessible data. Because the terminology used on the Dark Web is different from that on the Surface Web, deep neural networks may have difficulty categorizing writings on the Dark Web. However, there hasn't been much study done on Dark Web languages. In this paper, we compare the language of the Surface Web with the Dark Web using CoDA. Additionally, many techniques for classifying Dark Web pages are assessed. Lastly, we evaluate CoDA's suitability for different use cases by contrasting it with a publicly available Dark Web dataset.

e) **Dark web traffic detection method based on deep learning:**

Network security and the well-known topic of network traffic detection are closely connected. Since many crimes have been perpetrated on the dark web and traffic identification has become more difficult due to encryption technologies, this study focused on dark web traffic detection. In this study, we used publicly available data sets to evaluate a deep learning-based dark web traffic identification method. We were able to obtain a 95.47% detection precision by analyzing the study outcomes.

3. METHODOLOGY

i) Proposed Work:

To improve the categorization accuracy of Dark Web services, this study offers an extended deep learning model based on the best-performing LDA-TextCNN method. The proposed method uses Latent Dirichlet Allocation (LDA) to obtain topic modeling features before training a TextCNN model for classification. To further increase performance, we extend this model by utilizing CNN2D, which enhances classification accuracy and fine-tunes feature selection.

In this extended approach, the best features from the LDA-TextCNN model are given to the CNN2D network. By using convolutional layers to capture better spatial correlations among topic attributes, CNN2D improves the model's ability to discern between relevant and irrelevant input. A Dropout layer is also incorporated to remove noisy and redundant features, ensuring that only the most significant topic weights contribute to classification.

By combining LDA-TextCNN with CNN2D and Dropout, the extended model effectively reduces misclassification rates and achieves higher accuracy when compared to existing methods like KNN and Random Forest. Experimental results show that this approach significantly enhances Dark Web service detection, making it a viable choice for cybersecurity applications.

ii) System Architecture:

To improve the categorization accuracy of Dark Web services, this study offers an extended deep learning model based on the best-performing LDA-TextCNN method. The proposed method uses Latent Dirichlet Allocation (LDA) to obtain topic modeling features before training a TextCNN model for classification. To further increase performance, we extend this



model by utilizing CNN2D, which enhances classification accuracy and fine-tunes feature selection.

In this extended approach, the best features from the LDA-TextCNN model are given to the CNN2D network. By using convolutional layers to capture better spatial correlations among topic attributes, CNN2D improves the model's ability to discern between relevant and irrelevant input. A Dropout layer is also incorporated to remove noisy and redundant features, ensuring that only the most significant topic weights contribute to classification.

By combining LDA-TextCNN with CNN2D and Dropout, the extended model effectively reduces misclassification rates and achieves higher accuracy when compared to existing methods like KNN and Random Forest. Experimental results show that this approach significantly enhances Dark Web service detection, making it a viable choice for cybersecurity applications.

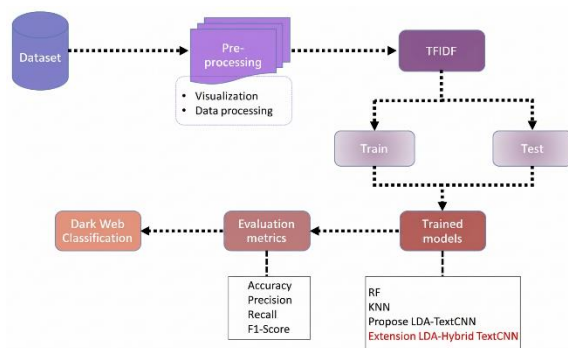


Fig.1. Proposed Architecture

iii) MODULES:

i. Data Loading:

- Imports the dataset for training and testing.
- Uses a Kaggle dataset containing over 10,000 Dark Web service descriptions.

ii. Visualization:

- Generates graphs to display different Dark Web services.
- X-axis represents service names, and Y-axis indicates their counts.

iii. Data Preprocessing:

- Removes digits, special symbols, and stop words.
- Applies stemming and lemmatization to standardize words.

iv. TF-IDF Transformation:

- Converts processed text into numerical vectors.
- Assigns word frequency scores for better feature extraction.

v. LDA Topic Modeling:

- Extracts the most relevant words from each sentence.
- Selects the top 90 topic weights for classification.

vi. Data Splitting:

- Divides the dataset into training and testing sets.
- Ensures an optimal split for model evaluation.

vii. Model Generation:

- *Existing Models:* KNN and Random Forest for baseline comparison.
- *Proposed LDA-TextCNN:* Uses LDA topic weights with TextCNN for classification.
- *Extension LDA-Hybrid TextCNN:*
 - Extracts best features from LDA-TextCNN.
 - Trains a CNN2D model on extracted features.
 - Uses a Dropout layer to remove irrelevant features.
 - Enhances classification accuracy.

viii. Admin Login:

- Allows admin authentication and access to system functionalities.

ix. Dark Web Classification:

- Enables users to upload test data.



- Classifies Dark Web services based on the trained models.

x. Prediction:

- Displays final classification results.
- Shows accuracy improvements using the extension model.

iv) ALGORITHMS:

a) KNN

The K-Nearest Neighbors (KNN) algorithm classifies Dark Web services based on how closely their attributes resemble labeled training data points. KNN uses feature space distance measurements to forecast the class label of a service. This study uses KNN as a baseline classification model so that its accuracy and effectiveness may be compared to more complex algorithms.

This helps assess the overall performance improvements resulting from the use of LDA-TextCNN and other recommended models.

b) Random Forest:

The Random Forest ensemble learning approach is used to build numerous decision trees during training, and it outputs the mode of the trees' predictions for classification problems. It minimizes overfitting and well handles high-dimensional data, making it appropriate for the Dark Web classification attempt. Reliable predictions for the categories of Dark Web services are produced by the model's usage of Random Forest to find complex patterns in the data. This algorithm's performance serves as a benchmark for evaluating proposed methods.

c) LDA-TextCNN:

The proposed LDA-TextCNN model combines a Convolutional Neural Network (CNN) with Latent

Dirichlet Allocation (LDA) topic modeling to classify Dark Web services. Through the extraction of topic weights from the processed text, LDA enhances feature representation. These improved features are then used by the TextCNN to conduct classification, boosting its accuracy and resilience.

This approach aims to better capture the semantic connections found in the data, which will ultimately result in better classification outcomes when compared to more traditional methods like KNN and Random Forest.

d) Extension LDA-Hybrid Texting:

The extension LDA-Hybrid TextCNN model builds upon the proposed LDA-TextCNN by including a second CNN2D layer that was trained using the best features from the previous model. This hybrid approach aims to increase classification accuracy by enhancing feature representation and reducing the impact of irrelevant data.

The use of dropout layers in this model lessens overfitting and results in a more precise classification of Dark Web services. This enhanced model aims to surpass earlier algorithms.

4. EXPERIMENTAL RESULTS

The proposed system was evaluated using a Kaggle dataset containing more than 10,000 Dark Web service descriptions. Data preprocessing techniques such as stop word removal, stemming, and lemmatization were applied to improve text quality. TF-IDF was used for initial feature representation, followed by Latent Dirichlet Allocation (LDA) to extract meaningful topic features. Through experimentation, selecting the top 90 topics provided optimal performance. The dataset was split into



training and testing sets to ensure unbiased evaluation, and multiple models including KNN, Random Forest, LDA-TextCNN, and the extended LDA-Hybrid TextCNN (CNN2D with Dropout) were implemented for comparison.

The results demonstrate that the proposed models significantly outperform traditional approaches. KNN and Random Forest achieved accuracies of 83.70% and 87.46%, respectively, while the LDA-TextCNN model improved accuracy to 95.05%. Further enhancement using the LDA-Hybrid TextCNN (CNN2D with Dropout) achieved the highest accuracy of 96.09%, along with superior precision, recall, and F1-score. The integration of CNN2D and Dropout effectively reduced noise and overfitting, leading to better generalization. These findings confirm that the proposed hybrid deep learning model provides a highly efficient and reliable solution for Dark Web content classification.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Accuracy} = \frac{(TN + TP)}{T}$$

Test Accuracy: 0.9895

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

$$\text{Precision} = \frac{\text{True positives}}{(\text{True positives} + \text{False positives})} = \frac{TP}{(TP + FP)}$$

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

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

mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

AP_k = the AP of class k
 n = the number of classes

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(\text{Recall} \cdot \text{Precision})}{(\text{Recall} + \text{Precision})}$$



Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Existing KNN	83.7	82.76	82.92	82.75
Existing Random Forest	87.46	86.74	86.36	86.41
Proposed LDA-TextCNN	95.05	94.93	94.19	94.34
Extension LDA-Hybrid TextCNN	96.09	95.83	95.94	95.55

Fig.7. Comparison table of performance evaluation metrics of all algorithms

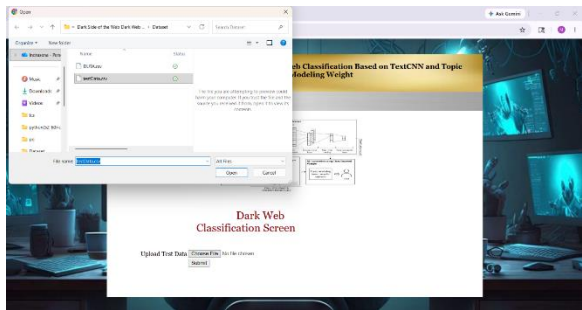


Fig.8. dataset upload page

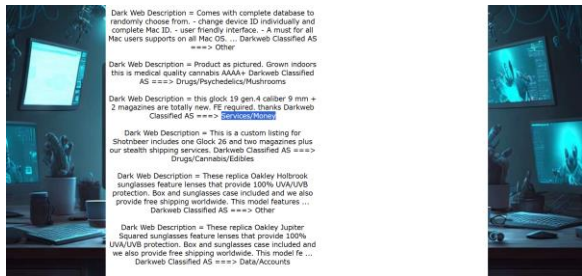


Fig.9. classification service page

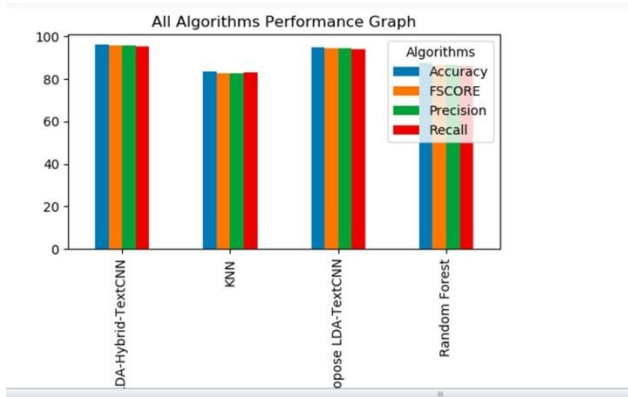


Fig.10. Accuracy Graph

5. CONCLUSION

This paper presented an efficient deep learning-based approach for Dark Web content classification using LDA-weighted TextCNN and an extended CNN2D model. The integration of Latent Dirichlet Allocation (LDA) with TextCNN improved feature representation by capturing meaningful topic-based information, overcoming the limitations of traditional methods like TF-IDF and LSA. Further enhancement using CNN2D with a Dropout layer enabled better feature refinement and reduced overfitting, leading to improved classification performance.

Experimental results demonstrated that the proposed hybrid model significantly outperforms conventional machine learning algorithms such as KNN and Random Forest in terms of accuracy, precision, recall, and F1-score. With an accuracy of 96.09%, the system proves to be highly effective in detecting and classifying Dark Web services. Overall, the proposed approach provides a robust and scalable solution for cybersecurity applications, enabling efficient monitoring and identification of potentially malicious activities on the Dark Web.

6. FUTURE SCOPE

The proposed system can be further enhanced by incorporating advanced deep learning models such as Transformers and attention-based architectures (e.g., BERT) to capture deeper contextual relationships in Dark Web text data. Additionally, integrating real-time data collection from live Dark Web sources and multilingual text analysis can improve the system's applicability and robustness across diverse environments.



Future work can also focus on improving scalability and deployment by implementing the model in a real-time cybersecurity monitoring system. Furthermore, combining text-based analysis with other modalities such as network traffic data or image-based content can lead to a more comprehensive Dark Web detection framework, enabling more accurate identification of complex and evolving cyber 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.

[45] 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.jjime.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] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, “Fake news stance detection using deep learning architecture (CNN-LSTM),” *IEEE Access*, vol. 8, pp. 156695–156706, 2020, doi: 10.1109/ACCESS.2020.3019735.

[9] 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.

[10] 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.

[11] L. Ouyang and Y. Zhang, “Phishing Web page detection with HTML level graph neural network,” in *Proc. IEEE 20th Int.*



Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom), Oct. 2021, pp. 952–958, doi: 10.1109/TrustCom53373.2021.00133.

[12] 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.

[13] 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.

[14] M. W. Al-Nabki, E. Fidalgo, E. Alegre, and L. Fernández-Robles, “ToRank: Identifying the most influential suspicious domains in the tor network,” *Expert Syst. Appl.*, vol. 123, pp. 212–226, Jun. 2019, doi: 10.1016/j.eswa.2019.01.029.

[15] M. W. Al Nabki, E. Fidalgo, E. Alegre, and D. Chaves, “Supervised ranking approach to identify influential websites in the darknet,” *Int. J. Speech Technol.*, vol. 53, no. 19, pp. 22952–22968, Oct. 2023, doi: 10.1007/s10489-023-04671-9.

[16] 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.

[17] 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.

[18] 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.

[19] R. Li, S. Chen, J. Yang, and E. Luo, “Edge-based detection and classification of malicious contents in tor darknet using machine learning,” *Mobile Inf. Syst.*, vol. 2021, pp. 1–13, Nov. 2021, doi: 10.1155/2021/8072779.

[20] 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.

[21] T. Guo and B. Cui, “web page classification based on graph neural network,” in *Innovative Mobile and Internet Services in Ubiquitous Computing*. Cham, Switzerland: Springer, 2022, pp. 188–198, doi: 10.1007/978-3-030-79728-7_19.

[22] M. Ball, R. Broadhurst, A. Niven, and H. Trivedi, “Data capture and analysis of darknet markets,” *SSRN Electron. J.*, to be published, doi: 10.2139/ssrn.3344936.

[23] Z. Ursani, C. Peersman, M. Edwards, C. Chen, and A. Rashid, “The impact of adverse events in darknet markets: An anomaly detection approach,” in Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW), Sep. 2021, pp. 227–238, doi: 10.1109/EuroSPW54576.2021.00030.



[24] P. W. Foltz, “Latent semantic analysis for text-based research,” *Behav. Res. Methods, Instrum., Comput.*, vol. 28, no. 2, pp. 197–202, Jun. 1996, doi: 10.3758/BF03204765.

[25] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[26] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, “Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora,” in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, vol. 1, 2009, pp. 248–256, doi: 10.3115/1699510.1699543.

[27] J. Mcauliffe and D. Blei, “Supervised topic models,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 20, 2007.

Author Profiles



Ms.M.Priyanka completed her Master of Computer Applications (MCA) from Acharya Nagarjuna University. She is currently working as an Assistant Professor. Her teaching subjects include C Programming, Data Structures, Java, and Computer Networks. Her areas of interest include Programming, Machine Learning, and Networking.



Ms.K.Pavani Working as Assistant & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with MCA ,M. Tech in Computer Science .She has 10 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.



Mr.P.Manohar Reddy is an MCA Student in the Department of Computer Application (MCA) at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in B.Sc.(Computers) from P.B.Siddhartha College of Arts & Science, Vijayawada. His area of interest are DBMS and Machine Learning with Python.