International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Dark Web Classification: A Dynamic Crawler and Robust Classification Framework

Sasirekha Devarajan[1*], Pakutharivu Panneerselvam[2], Aditya Mudigonda[3], Perichetla Kandaswamy Hemalatha[4]

Submitted: 25/09/2023 **Revised:** 15/11/2023 **Accepted:** 28/11/2023

Abstract: The dark web presents significant challenges for law enforcement agencies due to its anonymous and constantly evolving nature, making it difficult to trace and monitor illegal activities. To combat this issue, a proposed system collects and cleans dark web pages, focusing on studying the market that specializes in selling illegal and harmful products through three key modules. The Crawler module accesses the market through the Tor Network, gathers data, and extracts crucial information about the products, sellers, and prices. The Pre-Processing module cleans and organizes the extracted data, ensuring its integrity and transforming it into a mineable format. The Data Mining module extracts insights and knowledge from the processed data, using techniques like clustering, classification, and association rule mining to identify patterns and trends. These modules provide valuable insights to law enforcement agencies and security researchers to combat illicit activities on the dark web market. The system's efficiency is evaluated through metrics like throughput and speedup, demonstrating its capability to handle large datasets and improve performance through parallel processing. Additionally, the proposed Support Vector Machine (SVM) with Neural Network (NN) outperforms other methodologies, highlighting its accuracy in predicting dark web links and establishing a robust classification framework. The research contributes to a comprehensive understanding of the dark web landscape and fosters advancements in cybersecurity and law enforcement practices. By integrating these solutions, this study aims to enhance the accuracy, adaptability, and effectiveness of dark web classification. The research contributes to a comprehensive understanding of the dark web landscape and fosters advancements in cybersecurity and law enforcement practices.

Keywords: *Data Mining module, Dark Web Classification, Neural Network, Support Vector Machine, Dark Web Links*

1. Introduction

The internet can be categorized into three main types based on their public or private access and accountability or anonymity, as outlined by Lautenschlager in 2016: the surface web, the deep web, and the dark web [1]. The surface web is accessible to everyone without authentication or payment and is indexed by search engines. Those who use it can be identified and held accountable by law enforcement. The deep web, on the other hand, is private and not indexable by search engines. It requires authentication or access to internal networks. However, accountability is still present. The dark web - also known as darknets or hidden services - is a subset of the internet that requires special software for access. It is not indexed by search engines and contains both public and private elements. Only those with the correct credentials and software can access it.

The Dark Web is a hidden part of the internet that poses a challenge for detecting and monitoring illegal activities due to its encryption and anonymous nature. Unlike the regular World Wide Web (WWW), it operates within a network known as The Tor Onion Routing [2], which offers users anonymity by redirecting their traffic through multiple nodes. This makes it difficult to track and attribute actions to individuals, making it a safe haven for criminal activities such as illegal information sharing, contraband trading, and unlawful recruitment across legal jurisdictions.

In recent years, law enforcement agencies from various countries have collaborated to take down illegal activities on the Dark Web. "Operation SaboTor" in April 2018 [3] was one such operation that led to the shutdown of several illicit marketplaces and the arrest of many individuals involved in criminal activities. The Wall Street Market, one of the largest darknet marketplaces, was shut down in May 2019, and several administrators and vendors associated with the marketplace were arrested [4]. In July 2020, the darknet forum Dread was taken down [5], and in November 2020, the alleged operator of Hydra was arrested [6]. Most recently, DarkMarket, one of the largest darknet marketplaces, was taken down in January 2021 [7], leading to the arrest of the marketplace's operator and the seizure of its infrastructure. These international law enforcement operations have successfully targeted illegal activities on the Dark Web.

[1,2] *Department of Computer Science, Anna Adarsh College for Women, Chennai-600040, Tamil Nadu, India. Email: d.sasirekha@annaadarsh.edu.in*, Email: pakutharivu.p@annaadarsh.edu.in*
[3] *JNIAS School of Planning and Architecture, Hyderabad, Telangana 500034, India.Email: aditya_ouce@outlook.com*
[4] *Department of Mathematics, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Avadi.600062, Chennai, Tamil Nadu, India. Email: pkhemalathamsc@gmail.com[4]*

Classifying content on the Dark Web is a challenging task due to its hidden and constantly changing nature. To ensure accurate classification, diverse and comprehensive datasets are crucial. A diverse dataset helps improve accuracy and generalization, avoids bias towards specific categories, enables more nuanced decisions, and reduces bias by representing different demographics and regions. The Dark Web constantly evolves, making adaptability essential for effective classification [8]. An adaptable system can keep its knowledge up to date, detect emerging threats and trends, and adjust its strategies to handle website churn and evolving cybercriminal techniques. With an adaptable and diverse dataset, meaningful insights can be gained despite limited attribution. The objective of this research includes:

- Develop a dynamic crawler to efficiently explore and collect data from the Dark Web.
- Create a robust classification framework using machine learning techniques to accurately categorize Dark Web content.
- Analyze the significance of efficient Dark Web classification for cybersecurity and law enforcement efforts.
- Address existing challenges related to identifying and monitoring illicit activities within the Dark Web
- Contribute to broader efforts in combating cybercrime and enhancing digital ecosystem security by enhancing Dark Web classification capabilities.

2. Literature Review

The internet has created a global digital society that transcends legal jurisdictions, nationalities, religions and race. Despite its unstructured nature, this society is made up of individuals who are bound by the laws of their country. Online identities, primarily IP addresses, are linked to websites or individuals that own them, making it easy to attribute most online activities, particularly for law enforcement who can monitor traffic or request logs from the owners [9]. As a result, the concept of digital anonymity (the lack of connection between a physical identity and a digital one) is crucial.

There are various ways in which the dark web is used. It is used for the market trading of illegal drugs [10], trading credit card information, trading malware and exploits, identities, and stolen information, trading child abuse media, and trading weapons. It also serves as a communication platform for forums and chat for real-time communication. The dark web is also used as a tool for cybercrime [11], including malware-as-a-service, command-and-control servers deployed as hidden services, and terrorism operations. It is a source of threat intelligence and is used for anonymous financial transactions [12], such as using Bitcoin over Tor for anonymity and money laundering of cryptocurrencies via tumbling services. The dark web is also used as a proxy to the surface web to avoid censorship and protection from persecution by local authorities due to browsing anonymity [13].

There are various methods used for classifying the Dark Web, including crawling techniques, machine learning, and natural language processing (NLP) [14]. Crawling involves exploring hidden services on the Tor network to collect and index data. Machine learning techniques, such as supervised and unsupervised learning, are used for classification by feeding features extracted from collected data into models. NLP is used to extract meaningful information from textual content on the Dark Web, identifying keywords, patterns, and linguistic features associated with illicit activities.

The classification of the Dark Web poses several challenges that researchers must overcome. Firstly, obtaining reliable labeled data for machine learning models is difficult due to the sensitive and illegal nature of the content [15]. Secondly, the Dark Web is constantly changing, with new websites and services emerging while others disappear, making it challenging to keep up with these changes [16]. Additionally, the anonymity and encryption mechanisms on the Dark Web make it hard to attribute content to specific individuals or entities, which poses challenges for content classification and monitoring [17]. Finally, some websites on the Dark Web may serve multiple purposes or have content that blurs the lines between legal and illegal activities, making accurate classification more challenging. Figure 1 shows the typical system architecture of a Tor network.

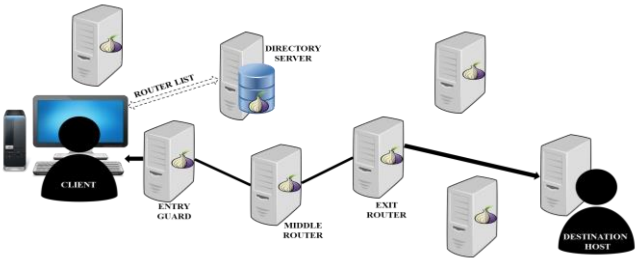

Fig. 1. Tor network system architecture

To classify content on the Dark Web, feature extraction techniques like Bag-of-Words, TF-IDF, N-grams, and Word Embeddings are used [18]. These techniques help capture keyword-based patterns, emphasize important words, capture local context, and represent words as dense vectors in a continuous space to enhance the classification model's representation power. Various algorithms are used for Dark Web Classification. These include SVM [19], Random Forest (RF), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long

Short-Term Memory (LSTM). These algorithms are suitable for text and image classification tasks and can handle non-linear decision boundaries and model dependencies between words in a sequence. Evaluation metrics used in Dark Web classification include accuracy, precision, recall, and F1-score. While accuracy is commonly used for balanced datasets, precision, recall, and F1-score are more suitable for imbalanced datasets. Area Under Curve – Region of Convergence (AUC-ROC) is useful for providing insights into the classifier's ability to distinguish between classes, and the confusion matrix aids in understanding the types of misclassifications made by the model.

There are several limitations of current systems used for classification of content on the Dark Web [20-22]. These include

- False positives can impact the effectiveness of the system by mistakenly identifying legitimate content as illegal.
- False negatives can also pose a problem as illegal content may not be recognized.
- Scalability may be a concern with some techniques struggling to handle the vast and diverse content on the Dark Web, leading to performance issues.
- Ethical and legal concerns must be considered when accessing and analyzing illegal content on the Dark Web.
- Classification models trained on one language or region may not be applicable to content in other languages or regions, thereby limiting the system's usefulness.

3. Materials

The aim of this research is to develop and implement a system for analysing and extracting insights from the dark web market that specializes in selling illegal and harmful products. The study adopts a multi-module approach involving a Crawler module, Pre-Processing module, and Data Mining module. The research is designed as a technical implementation study. The setting involves the connection to the Tor Network, a secure and anonymous network, to access the dark web market for data collection and analysis. The participants in this study are not human subjects but rather digital entities, including the dark web market website and the data extracted from it. No human participants, data, or tissue are involved in this study.

3.1 Processes and Methodologies

The Crawler module is responsible for accessing the dark web market via the Tor Network, navigating through its pages, and collecting relevant data about products for sale. This includes details like product names, descriptions, prices, and seller information. Privacy 4 via Tor software is used to establish a connection with the Tor Network. The Crawler simulates a user login with previously used credentials to access the market. It collects structured data directly from the website pages, eliminating the need for separate data files. This module gathers data on various categories of illegal products.

The collected data is passed to the Pre-Processing module, which involves several steps to clean, organize, and structure the data. The Cleaner module reads the data from JSON files, extracts the "Title" element, and processes it. The processing includes removing non-alphabetic characters and common stop words. The resulting cleaned data is stored in a new JSON file and further transformed into an XLSX file format using the Panda library. This module operates on both vendors' data and product data, producing separate spreadsheets for each.

The pre-processed data is then handed over to the Data Mining module. This module employs various data mining techniques, including data clustering, classification, and association rule mining. The aim is to extract knowledge and insights from the data. Association rules are used to identify common word combinations used by vendors to increase product visibility. Classification is performed using SVM algorithm, enhanced with a neural network for improved accuracy. This module enables clustering of data from both vendors and products lists, helping to uncover patterns, correlations, and trends within the collected data.

3.2 Statistical Analysis

Statistical analysis in this research involves measuring the performance and efficiency of the system. Efficiency is evaluated using speedup metrics as the number of nodes increases. Throughput is calculated by dividing data processing size by response time. Performance comparison of various algorithms, including SVM with NN, SVM, and others, is done using metrics like accuracy, MCC, and ROC area. Performance is also assessed across different datasets using metrics such as precision, recall, F1 score, and percentage of dark web link prediction.

4. System Architecture

There are three basic aspects of the proposed system modules. To study the market, a connection with the Tor Network is established by registering with the service. This market specializes in selling and promoting illegal and harmful products in various categories.

- The first module of the system is the Crawler module, which is responsible for accessing the dark web market and gathering data. It requires a connection to the Tor network, which provides anonymity and privacy. The crawler module will navigate the market website, crawl through its pages, and collect information about the products being sold. This includes details such as the product name, description, price, seller information, and any other relevant data.

- Once the crawler module has collected the necessary data, it is passed to the pre-processing module. This module is responsible for cleaning, organizing, and structuring the extracted data. It involves removing any unnecessary information, formatting the data into a standardized format, and eliminating any duplicates or inconsistencies. The pre-processing module also performs tasks like data normalization and data validation to ensure the quality and integrity of the data.
- After the data has been pre-processed, it is handed over to the data mining module. This module is responsible for extracting knowledge and insights from the processed data. It uses various techniques like data clustering, classification, and association rule mining to discover patterns, relationships, and trends within the data. The data mining module can help identify popular products, analyze pricing trends, detect correlations between different product categories, and uncover other valuable insights.

Overall, these three modules work together to crawl the dark web market, process the extracted data, and mine useful information from it. This system can provide valuable insights to law enforcement agencies, security researchers, and other relevant stakeholders to better understand and combat illicit and malicious activities on the dark web market.

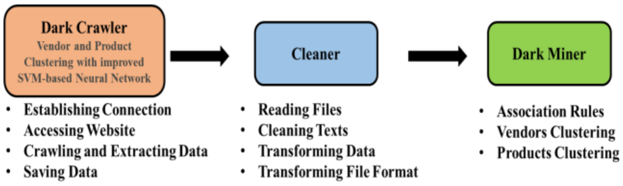

Fig. 2. System Modules

5. Design Implementation

5.1 The Crawler

The first involves establishing a connection to the Tor network using Privacy 4 via Tor software. Once the connection is established a user login is simulated by using the credentials that were previously used in the dark web market. The next process involves crawling through all the pages of the websites, thereby getting structured data directly without the need to download or save data in a separate json file. Table 1.1 gives a brief overview of the different vendors and their products extracted from the Dark Web.

Table 1. Dark Web Vendors and Products Extracted

Vendors	Products
Weapons	7
Drugs and Chemicals	2096
Guides and Tutorials	1439
Fraud	650
Security and Hosting	75
Gold	24
Carded Items	53
Digital Products	3564
Counterfeit Products	362

5.2 Data Pre-Processing

The data pre-processing stage helps to transform data into a more readable and mineable format. The following are the processes involved in the pre-processing stage.

1. The Cleaner module reads data from json files. Specifically, it focuses on retrieving the "Title" element from the json files.
2. Processing the "Title" element: For the "Title" element, the Cleaner performs two main tasks. Firstly, it removes non-alphabetic or meaningless characters present in the title. This helps to clean up the data and make it more meaningful. Secondly, the Cleaner removes the presence of stopwords such as "a", "the", "is", etc.
3. After processing the elements, the cleaner module adds them to a new json file with the processed and clean data, ready for further analysis.
4. The final step performed by the cleaner module is to transform the json file to a new file format, namely xlsx. This format is commonly used for mining and analyzing data, as it can be easily imported into tools like spreadsheets or data mining software. The module is designed taking into consideration the two versions namely vendors' data and products data.

Since we have two different types of data - products data and vendors' data - the cleaner module needs to be designed in two versions.

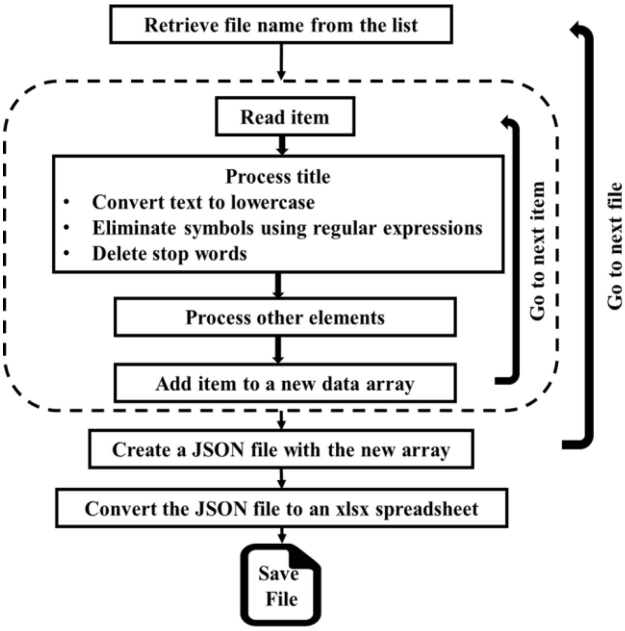
Retrieve file name from the list
Read item
Process title
• Convert text to lowercase
• Eliminate symbols using regular expressions
• Delete stop words
Process other elements
Add item to a new data array
Go to next item
Go to next file
Create a JSON file with the new array
Convert the JSON file to an xlsx spreadsheet
Save File

Fig. 3. The cleaner workflow

Fig.3 shows the working of Cleaner with products data. The process involved is as follows:

1. The file names which hold the products are chosen.
2. The sign name is red to identify the elements.
3. Every letter in the title is transformed to lowercase and non-alphabetical characters are also removed. An expression is defined which holds all the characters that have been removed and were used in the titles. This includes strange shapes like ♛, ✯, ✪, ❥ and common characters such as (,), [,],”, etc. The vendors commonly use these to draw attention to the product. However, characters that might distort the actual meaning of the title such as % and - have been excluded. Moreover, nltk.corpus library is used to track the stopwords which should be removed by the Cleaner.
4. A new data array is created with the processed elements which is then moved to the next file.
5. This is further transmitted to a new json file.
6. This data is then transformed into an xlsx using panda library.
7. Using a similar methodology, the vendor’s data file is also transformed.

Thus, a spreadsheet is created separately for the vendors and another for the products.

5.3 Data Pre-Processing, Clustering and Classification with Neural Network

Now that the vendors and the products file has been created, the next step involves studying the website content. This involves employing a methodology to study and analyze the different words based on their frequency along with illicit product titles. Association rules which are commonly used for experimenting with shopping traffic is hence adopted into this system. Here, the purpose of it is to identify and narrow down the words that are used by the vendors when increasing the visibility of the product. This is carried out by taking into account the words as items and the Title as the transaction. Based on the analysis carried out, the association rules enable capturing the most used combination of words in titling products. Classification of the data can be clustered using several algorithms. SVM is identified to be the optimal mechanism when compared to other similar algorithms such as Multilayer perceptron (MLP), LSTM, and RF. Initially, SVMs were recognized for their ability to investigate and control the volume of neural networks along with the ability to resolve overfitting complications. SVMs are capable of making decisions by minimizing classification errors and reducing operational risks. The SVM algorithm utilizes the Maximum Margin Classifier (MMC) to tackle the particle problem. In this study, an enhanced SVM algorithm incorporating neural networks is introduced.

Neural Network is the perception of classifying data, learning, and performing analysis. Using this technique, it is possible to cluster the data from the Vendor and Product list extracted from darkweb. The proposed methodology focuses on determining an accurate and resourceful framework which can be used to uniquely identify the pre-processed data. The proposed classifier model enhances the performance of the existing classification technique. This model is built upon the fusion of expanded backpropagation neural network (BPNN) with a neural network classifier. In our approach, labeled data is utilized for binary classification, where the probability of each dataset in the classification class is determined and converted into weights. These weights are then distributed among the specific classes and used for neural network training. During the classification process, testing data undergoes a similar progression, and the weights are adjusted based on the essential labels of the dataset. The output of this testing phase is used to predict the performance of the system. The data clustering technique improves upon previous distributed or parallel works by introducing a new training neural network and incorporating additional feature engineering methods. This approach addresses the challenges of web application classification and enhances the accuracy of big data. The results of this research are crucial for training datasets in neural network-based classification problems, particularly for information recommendation based on user interests. The proposed model minimizes the classified data web structure, allows conversion of unlabeled instances into labeled instances, and improves accuracy and efficiency through fusion-based feature selection algorithms. Through simulation on real-time datasets, we have evolved the mapreduced model's efficiency and resulted in effective results.

6. Results and Discussion

An important metric that needs to be evaluated is efficiency. Efficiency measures how well the system performs after adding additional nodes. It is determined by the speedup achieved, which is the ratio of the number of nodes used. As more nodes are added, the efficiency increases. However, it decreases when the number of nodes is further increased, for complete dark web dataset sizes. On the other hand, throughput refers to the amount of data processed per unit of time. In this study, throughput is calculated by dividing the size of the handled data by the response time, as defined by the equation:

Throughput = data processing size (DPS) / response time (T, R).

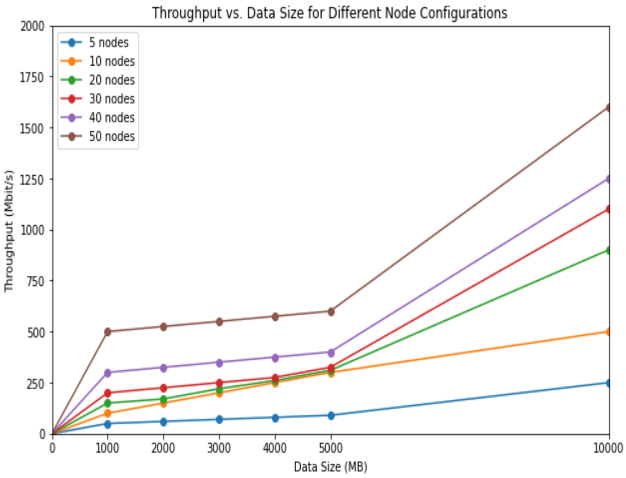

Fig.4. Average Throughput

Increasing the number of nodes initially leads to improved performance and speedup. However, at a saturation point of around 2000 MB for 10 nodes, the performance becomes saturated, indicating that adding more nodes does not provide significant benefits. For 30 and 40 nodes, the speedup achieved is excellent, indicating efficient parallel processing. The shared bandwidth among these nodes is effectively utilized, resulting in significant performance gains compared to sequential processing. However, the speedup of 50 nodes is lower than that of 30 and 40 nodes. This anomaly suggests that the shared bandwidth among nodes undergoes significant changes, leading to a decrease in performance.

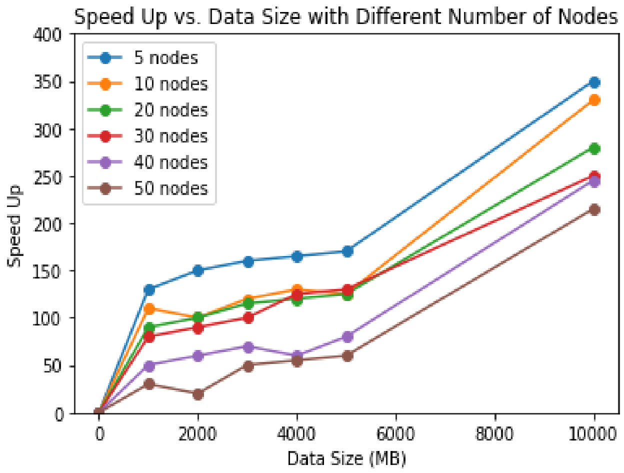

Fig. 5. Speed-Up vs. data size

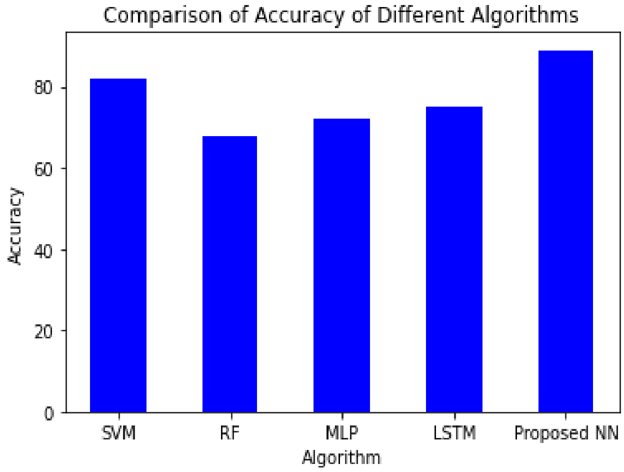

Fig.6 Accuracy of various algorithm

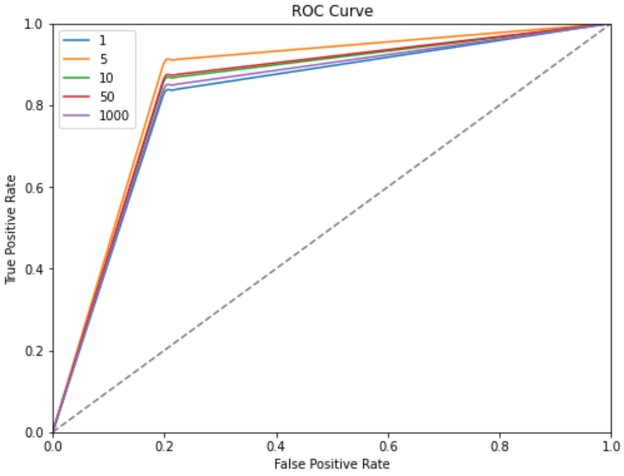

Fig. 7. ROC Curve

Fig.6 shows the accuracy of the proposed optimized SVM with Neural Network is found to be higher than SVM and other similar methodologies. Similarly, the ROC curve is represented in Fig. 7. Similarly, to determine the efficacy of the various algorithms, Precision, Recall, F1 Score and percentage of Dark Web Link prediction is evaluated for several databases such as Android botnet dataset, DDoS evaluation dataset, Intrusion detection evaluation dataset and CIRA-CIC-DoHBrw-2020 which indicate a strong positivity with the proposed methodology as represented in table 2.

Table 2. Performance comparison of different datasets

Dataset	Accuracy (%)	MCC	ROC area
Android botnet	96	0.48	0.97
DDoS evaluation	94	0.15	0.96
Intrusion detection evaluation	88	0.30	0.95
CIRA-CIC-DoHBrw-2020	82	0.56	0.85

A graphical comparison of various parameters namely precision recall, F1 score and percentage of dark web link prediction for SVM, RF, MLP, LSTM and the proposed SVM-NN algorithm is represented in Fig.8.

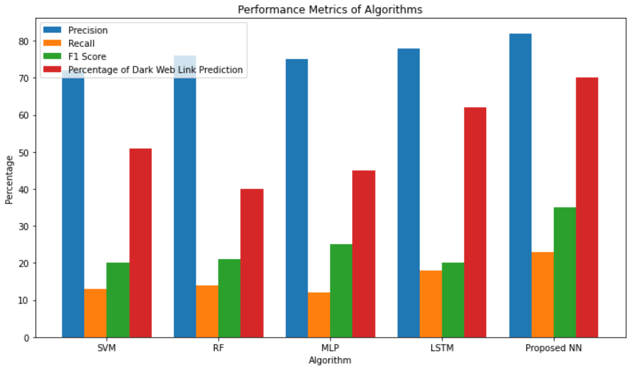

Fig.8. Performance of Models

7. Conclusion

A comprehensive solution is presented in this paper for analyzing the dark web market that specializes in selling illegal products. The system integrates three essential modules - Crawler, Pre-Processing, and Data Mining - to collect and process relevant data efficiently. This empowers law enforcement agencies and security researchers with valuable insights to combat illicit activities. The Crawler module connects with the Tor Network to ensure anonymity and privacy while navigating the dark web market and gathering product-related information. The Pre-Processing module plays a vital role in cleaning and organizing the data, making it readable and mineable for further analysis. Lastly, the Data Mining module leverages advanced techniques such as clustering, classification, and association rule mining to discover patterns and relationships within the data. The product titles of top vendors are then clustered to identify common shared characteristics, using the SVM-based Neural Network Algorithm. Optimal results in terms of accuracy, precision, recall and F1 score are achieved with 50 nodes. Even with an increased data size of up to 10,000 MB, a throughput of 1600 Mbps is achieved. The design of the crawler and data mining operations can vary based on the dark website and data quality. Future work can be directed toward implementing a real-time scenario for criminal activity tracking using an artificial neural network.

List of abbreviations

AUC – Area Under Curve

CNN – Convolutional Neural Networks

DPS – Data Processing Size

LSTM – Long Short-Term Memory

MLP – Multi-Layer Perceptron

MMC – Maximum Margin Classifier

NLP - Natural Language Processing

NN – Neural Network

RF – Random Forest

RNN – Recurrent Neural Networks

ROC – Region of Convergence

SVM – Support Vector Machine

WWW- World Wide Web

Declaration Statement

Availability of data and material

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Funding Details

No funding was received to assist with the preparation of this manuscript.

Authors Contributions

The author Sasirekha Devarajan Contributed and put effort on paper to Organize the Paper. Also technically contributed to data analysis. Pakutharivu Panneerselvam technically contributed and made English Corrections and grammar checking. Also involved and helped to derive the mathematical equation. The author Aditya Mudigonda involved in the Background study of the Paper and helped the mathematical derivations. Perichetla Kandaswamy Hemalatha technically involved and provided a factual review and helped edit the manuscript.

Acknowledgements

I am grateful to all of those with whom I have had the pleasure to work during this and other related Research Work. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general.

References

[1] Lautenschlager, S. (2016). Surface web, deep web, dark web—What's the difference. Cambia research [Online] Available at: https://www.cambiaresearch.com/articles/85/surface-web-deep-web-dark-web-whats-the-difference.

[2] Dingledine, R., Mathewson, N., & Syverson, P. F. (2004, August). Tor: The second-generation Ahmad, Maynard, A., S., & Gupta, A. (2019). The dark web as a phenomenon: A review and research agenda. In 30th Australasian Conference on Information Systems. Perth, Australia. Onion router USENIX security symposium, 4.

[3] Baney, L. L., & Lewis, N. M. (2018). Internet pharmacies: Trends, opportunities, and risks. Health Law, 31, 1.

[4] Magnúsdóttir, H. (2019). Darknet drug markets in a Swedish context: A descriptive analysis of Wall Street market and flugsvamp 3.0.

[5] Meland, P. H., Bayoumy, Y. F. F., & Sindre, G. (2020). The Ransomware-as-a-Service economy within the darknet. Computers and Security, 92, 101762. https://doi.org/10.1016/j.cose.2020.101762

[6] Bracci, A., Nadini, M., Aliapoulios, M., McCoy, D., Gray, I., Teytelboym, A., Gallo, A., & Baronchelli, A. (2022). Vaccines and more: The response of Dark Web marketplaces to the ongoing COVID-19 pandemic. PLOS ONE, 17(11), e0275288. https://doi.org/10.1371/journal.pone.0275288

[7] Abdel Samad, Y. (2021). Case study: Dark web markets. Dark web investigation (pp. 237–247).

[8] Rajawat, A. S., Bedi, P., Goyal, S. B., Kautish, S., Xihua, Z., Aljuaid, H., & Mohamed, A. W. (2022). Dark web data classification using neural network. Computational Intelligence and Neuroscience, 2022, 8393318. https://doi.org/10.1155/2022/8393318

[9] Ahmad, A., Maynard, S., & Gupta, A. (2019). The dark web as a phenomenon: A review and research agenda. In 30th Australasian Conference on Information Systems. Perth, Australia.

[10] Rhumorbarbe, D., Werner, D., Gilliéron, Q., Staehli, L., Broséus, J., & Rossy, Q. (2018). Characterizing the online weapons trafficking on crypto markets. Forensic Science International, 283, 16–20. https://doi.org/10.1016/j.forsciint.2017.12.008

[11] Liggett, R., Lee, J. R., Roddy, A. L., & Wallin, M. A. (2020). The dark web as a platform for crime: An exploration of illicit drug, firearm, CSAM, and cybercrime markets. The Palgrave handbook of international cybercrime and cyberdeviance (pp. 91–116).

[12] Nazah, S., Huda, S., Abawajy, J., & Hassan, M. M. (2020). Evolution of dark web threat analysis and detection: A systematic approach. IEEE Access, 8, 171796–171819. https://doi.org/10.1109/ACCESS.2020.3024198

[13] Gulati, H., Saxena, A., Pawar, N., Tanwar, P., & Sharma, S. (2022, January). Dark web in modern world theoretical perspective: A survey. In International Conference on Computer Communication and Informatics (ICCCI), 2022 (pp. 1–10). IEEE Publications. https://doi.org/10.1109/ICCCI54379.2022.9740785

[14] Dalvi, A., Paranjpe, S., Amale, R., Kurumkar, S., Kazi, F., & Bhirud, S. G. (2021, May). SpyDark: Surface and dark web crawler. In 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), 2021 (pp. 45–49). IEEE Publications. https://doi.org/10.1109/ICSCCC51823.2021.9478098

[15] Kadoguchi, M., Hayashi, S., Hashimoto, M., & Otsuka, A. (2019, July). Exploring the dark web for cyber threat intelligence using machine leaning. In IEEE International Conference on Intelligence and Security Informatics (ISI), 2019 (pp. 200–202). IEEE Publications. https://doi.org/10.1109/ISI.2019.8823360

[16] Sasnouskaya, T. (2023). Unveiling the dark web and the impact of REvil's cyberattacks.

[17] Saleem, J., Islam, R., & Kabir, M. A. (2022). The anonymity of the dark web: A survey. IEEE Access, 10, 33628–33660. https://doi.org/10.1109/ACCESS.2022.3161547

[18] Qasem, A. E., & Sajid, M. (2022, October). Exploring the effect of N-grams with BOW and TF-IDF representations on detecting fake news. In International Conference on Data Analytics for Business and Industry (ICDABI), 2022 (pp. 741–746). IEEE Publications. https://doi.org/10.1109/ICDABI56818.2022.10041537

[19] Nezhad, S. Z. (2023). Dark web traffic detection using supervised machine learning.

[20] Basheer, R., & Alkhatib, B. (2021). Threats from the dark: A review over dark web investigation research for cyber threat intelligence. Journal of Computer Networks and Communications, 2021, 1–21. https://doi.org/10.1155/2021/1302999

[21] Prabha, C., & Mittal, A. (2023, February). Dark Web: A review on the deeper side of the Web. In O. International (Ed.) Technology Conference on Emerging Technologies for Sustainable Development (OTCON), 2022 (pp. 1–6). IEEE Publications. https://doi.org/10.1109/OTCON56053.2023.10113989

[22] Alaidi, A. H. M. (2022). Dark web illegal activities crawling and classifying using data mining techniques. iJIM. Roa'a, M. ALRikabi, H.T.S., Aljazaery, I. A, and Abbood, S.H., 16(10), 123.