Science Journal of University of Zakho
DOI: https://doi.org/10.25271/sjuoz

Original article

APPLICATION OF ARTIFICIAL INTELLIGENCE FOR MONITORING DARK WEB ACTIVITIES

Abimbola G. Akintola[1], Yusuf O. Olatunde[2,*], Kolawole Y. Obiwusi[3], Ganiyat K. Afolabi-Yusuf[3], Muhammed A. Adebisi[3], Ayobami A. Tewogbade[2], Lawrence O. Omotosho[4], Olajide Y. Adebayo[4] and Aminat A. Oladipo-Tanimowo[4]

[1]Department of Computer Science, University of Ilorin, Ilorin, 240001, Nigeria.
[2]Department of Cyber Security Osun State University, Osogbo, 230103, Nigeria.
[3]Department of Computer Science, Summit University Offa, Offa, 250101, Nigeria.
[4]Department of Computer Science, Osun State University, Osogbo,230103, Nigeria.

[*] Corresponding author. E-mail: yusuf.olatunde@uniosun.edu.ng (Tel.: +234-8062174417)

Received: 15, Jun, 2025

Accepted: 29, Nov, 2026

Published: 17, Jan,2026

ABSTRACT

The Dark Web is a hidden part of the internet that attracts perpetrators due to its anonymity. It is used for cybercrime, illegal trading, and by terrorist groups to exchange stolen data and personal information. Current techniques for combating and identifying these concerns are weak. The primary focus of this study is on the application of artificial intelligence to monitor the Dark Web. The solution mitigates monitoring deficiencies to facilitate the timely identification of prospective threats, including data breaches and cybercrimes. The six forms of activities prioritised include cyberterrorism, terrorist activities, weapons trading, drug trafficking, human trafficking, and regular activities. The tasks carried out in this study include crawling relevant data, training machine learning models using an ensemble voting approach, and implementing a JavaScript-powered crawling engine. The AI model is tested and evaluated using data collected from various Onion sites that include both legal and illegal content. The approach used in this study gives 97% accuracy and a macro average precision of 98%. The macro average for recall and f1-score is 97%. The precision, recall and F1-Score all have the same weighted average of 97%.

KEYWORDS: Dark web, Web Crawler, Cyber Crime, Machine Learning, TOR Browser.

1. INTRODUCTION

The content on the internet is more than the simple websites we frequently visit. It also includes a hidden section known as the Dark Web (DkW), which is inaccessible by traditional search engines. As seen in Figure 1, the internet is categorized into Dark Web, Deep Web, and the Surface Web (Alshammery and Aljuboori, 2022; Ali and Qasim, 2023). The easily navigable portion of the web may be accessed using a conventional search engine. This is known as the Surface Web. While "Deep Web" represents the portion of the internet that contains websites that are not indexed or discoverable by web browsers, some examples of deep web content are email, Netflix, online banking, dynamic pages, databases, paywalls, and anything else that needs authorization and authentication. The Dark Web, on the other hand, is composed of an unseen network that relies on specialized programs and protocols (Alshammery and Aljuboori, 2022), with 96% of the Internet comprising the Deep Web (Kaur and Randhawa, 2020).

Access this article online

https://doi.org/10.25271/sjuoz.2026.14.1.1626

Printed ISSN 2663-628X;
Electronic ISSN 2663-6298

Science Journal of University of Zakho
Vol. 14, No. 01, pp. 112 –123 January-2026

Figure 1: Overview of the Surface Web, Deep Web, and Dark Web (Mahmood *et al*., 2022).

As illustrated in Figure 1, access to the Dark Web requires specialized tools such as TOR (The Onion Router) and cannot be reached through conventional search engines. Its encrypted networks allow users to carry out operations while maintaining their confidentiality. Even though the DkW has its good parts, like Privacy and Anonymity, Whistleblowing and Journalism, Research and security, and Legitimate Marketplace(Yang *et al*., 2020). The DkW anonymity fosters an environment where illegal activities can thrive, posing significant challenges for law enforcement and cybersecurity. The covert nature of this location supports criminality such as phishing, malware, identity theft, and stalking (Basheer and Alkhatib, 2021).

Traditional methods of monitoring and safeguarding against these harmful behaviours are ineffective (Wiafe *et al*., 2020). As a result, criminal activities like cybercrime and unlawful business operations thrive on the DkW because it is less monitored.

Therefore, this study aims to design and evaluate an AI-based framework for monitoring and detecting illicit activities on the DkW by integrating web crawling, data preprocessing, and ensemble-based machine learning. To achieve this aim, the following research questions are addressed:

i. How can automated crawling techniques be adapted to effectively acquire and preprocess unstructured DkW data for crime detection?
ii. To what extent can ensemble machine learning methods improve classification accuracy compared to individual models?
iii. What are the limitations and implications of deploying such a system in real-world scenarios, particularly in terms of false positives and false negatives?

The novelty of this study lies in its integrated pipeline that not only crawls and processes DkW content but also applies a voting ensemble to enhance classification accuracy and reliability, a gap not sufficiently addressed in existing works.

At the end, the key contributions of this study are summarized as follows:

i. We developed a crawling framework powered by Puppeteer and TOR integration to systematically collect and structure data from DkW onion sites.
ii. We created a curated dataset of 2,767 terms spanning six categories of Dark Web activities (cyberterrorism, terrorist activities, weapons trading, drug trafficking, human trafficking, and normal activities).
iii. We designed a voting ensemble model combining Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB), achieving a classification accuracy of 97%, which outperforms existing single-model approaches.
iv. We developed a user-friendly interface to visualize detection results in real time, making the system practical for cybersecurity monitoring and law enforcement use.
v. We provide a comparative evaluation of our method against previous work (He *et al*., 2019; Zenebe *et al*., 2019; Ebrahimi *et al*., 2020), demonstrating superior accuracy, precision, recall, and F1-score.

These contributions highlight the novelty of integrating DkW crawling, AI-driven classification, and practical system deployment for real-world monitoring of illicit online activities.

Dark web and how it works:

The term ‘dark web’ is often associated with illicit activities carried out by anonymous individuals or organizations, which makes investigation and monitoring challenging. (Alaidi *et al*., 2022). It employs the TOR Hidden Service (HS) protocol, which makes it difficult to trace the sites’ internet addresses; thus, these sites may engage in malicious and unlawful content propagation while being immune to police intervention and shutdowns (Ghosh *et al*., 2017). TOR anonymizes the internet data by forwarding it through a circuit of at least three nodes called relay nodes (Schafer *et al*., 2019). TOR is used in onion routing, which helps to encrypt data to layers where each layer is decrypted by a node given instruction for the node behind it. This technique makes sure that no node receives the whole path and content of the data. As stated by Mahmood *et al*., (2022), websites presented on the dark web are briefly hidden and can be accessed only with the help of the addresses containing the “*.onion*” suffix and are not visible in the regular search engine results.

Ways of accessing the dark web:

The frequently utilized resources for internet access include Google Chrome, Wikipedia, Firefox, Safari, and more. The dark web's obscurity makes it impossible for it to be indexed. To examine the dark web, two conditions must be met: (i) using dedicated software and (ii) a proxy server (Alaidi *et al*., 2022; Alshammery and Aljuboori, 2022). The software, as shown in Table 1, includes TOR Browser, Freenet, and I2P (Invisible Internet Project). Other common software includes Tails (The Amnesic Incognito Live System), Subgraph OS, Whonix, and VPN (Virtual Private Network), each providing different ways for ensuring anonymity and privacy while browsing the internet.

To anonymize browsing habits, the TOR Browser distributes traffic through a network of servers run by volunteers, and I2P makes it easier to browse websites and communicate in a way that are resistant to censorship. Freenet functions as an anonymous peer-to-peer, decentralized network for file sharing and online browsing. With integrated support for TOR, Subgraph OS is an operating system with a strong security focus that offers a haven for anonymous browsing. For increased anonymity, Tails is a live operating system that channels all internet traffic via TOR, which is similar to Whonix, that runs on a virtual computer. Through IP address concealment and internet traffic encryption, VPNs provide an additional layer of privacy and anonymity. Users are encouraged to use these tools responsibly and legally while taking precautions to protect their privacy and security

Table 1: Tools for accessing the Dark Web and their key features

S/N	Tools (URL)	Features	Open Source
1	Tor Browser (*https://www.torproject.org*)	i. Encrypts internet traffic and routes it through multiple nodes for anonymity. ii. Access to *onion* websites. **iii.** Free and open-source.	√
2	I2P (Invisible Internet Project; *https://geti2p.net*)	i. Peer-to-peer anonymity network. ii. Supports anonymous browsing, email, and file sharing. iii. Designed for secure communication.	√
3	Tails OS (*https://tails.boum.org*)	i. A portable operating system running from a USB or DVD. ii. Preconfigured with Tor for anonymous browsing. iii. Leaves no trace on the host system.	√
4	Whonix (*https://www.whonix.org*)	i. Privacy-focused virtual machine. ii. Routes all internet traffic through Tor. iii. Provides strong isolation for online activities.	√
5	FreeNet (*https://freenetproject.org*)	i. Decentralized network for anonymous communication. ii. Enables file sharing, forums, and uncensored publishing.	√
6	Brave Browser (Tor integrated; *https://brave.com*)	i. Built-in Tor tab for easy anonymous browsing. ii. User-friendly interface with enhanced privacy settings.	√
7	OnionShare (*https://onionshare.org*)	i. Securely share files over the Tor network. ii. Host anonymous websites.	√
8	DuckDuckGo (onion version; *http://3g2upl4pq6kufc4m.onion*)	i. Privacy-focused search engine. ii. Available within the Tor network.	**X**
9	Qubes OS (*https://www.qubes-os.org*)	. Secure operating system with compartmentalization. ii. Supports Tor for anonymous browsing.	√
10	ProtonMail (Tor Version; *https://protonmail.com/tor*)	i. Encrypted email service. ii. Accessible via Tor for added privacy.	**X**

Existing dark web monitoring approaches:

The landscape of dark web monitoring has evolved with the growing recognition of its significance for cybersecurity and law enforcement. Some methods were used to monitor and obtain intelligence from the internet's dark corners. One of them is BlackWidow, a highly automated modular system that monitors Dark Web services and consolidates the collected data into a single analytics platform. BlackWidow (Schafer *et al*., 2019) uses a Docker-based microservice architecture that allows for the mixing of pre-existing and customised machine-learning technologies.

Also, Saini and Bansal (2019) developed an automatic classifier to detect the procurement of modern weapons, including drones, in dark web discussion forums. Additionally, a system dubbed "HackerRank" was created to automatically identify prominent hackers. HackerRank leverages content and social network analysis (Huang *et al*., 2021). Also, in (Mahmood *et al*., 2022), Crime

detection techniques involving instruments such as Black Widow, HoneyPot, and Dark Crawler were discussed. Conventional methods encountered obstacles such as the vast amount of content on the dark web, the ever-changing nature of cyber threats, and the requirement for real-time monitoring capabilities despite these efforts. These made the need for more sophisticated and flexible monitoring systems necessary.

Related Work:

The survey of Mahmood *et al*. (2022) on monitoring DkW and detection of corresponding threats reveals that tools such as BelkaSoft, RegShot, and Wireshark are being used and criminal detection techniques like BlackWidow, HoneyPot and Dark Crawler are also well employed.

While interpreting content on the dark web becomes central in balancing cybercrimes and gaining a glance into the criminal's mind fuels the review conducted by Basheer and Alkhatib, (2021). It was also observed that hackers share information and learn from each other, and such discussions from forums may contain data that is crucial to the discovery of cyber threats.

Underground forums (Fang *et al*., 2019) were examined, and the report claims that buying and selling as well as bartering personal data breaches trends in the forums. The authors then followed a solution to construct a system that could automatically determine and categorise the forum threads that concern data breaches. This makes the structure of the system comprise a collector and preprocessor, a feature selector, and a classifier. For collecting the dataset, six forums were considered: This is Nulled, Breach Forums, HackThisSite, Hellboundhackers, Hidden Answers and Brotherhood. The data were collected in two ways, where the study used crawling and the other used the publicly leaked database dump. To achieve improved results, five supervised learning algorithms were evaluated: support vector machine, Naive Bayes, K-nearest neighbours, decision tree, and Random Forest and the latter was used to train the classifier. Finally, the system was able to identify 92% of data breach threads.

Similarly, Schafer *et al*., (2019) proposed to design a highly automated modular system for monitoring DkW services, collecting data, and providing a unified analytics framework for cybersecurity intelligence called BlackWidow, capable of real-time data collection in the Deep and Dark Web and the integration of external translation capabilities in a scalable way.

Wiafe *et al*., (2020) adopted systematic mapping to select and review 131 articles that focused on AI techniques to mitigate cybercrimes with DkW. It was then found that the most dominant approach is SVM. The authors also stress that researchers should not remain passive and should start using and incorporating new AI applications, as the applications of AI exhibit potential for cybersecurity.

While several works have applied machine learning for DkW intelligence, there is an emerging interest in semi-supervised and deep learning approaches. For instance, Ebrahimi *et al*., (2020) proposed a semi-supervised cyber threat identification system in Dark Net Markets (DNMs) using transudative learning and deep models. In a Semi-supervised learning approach, a model is trained on a small amount of labelled data along with a larger pool of unlabelled data. Transductive learning is a form of semi-supervised learning that aims to directly assign labels to the given unlabelled data instead of generalising to unseen future samples. Their method reduced the cost of manual labelling and improved F1-score performance by 3–5%.

Similarly, Zenebe *et al*. (2019) explored Cyber Threat Intelligence (CTI) discovery from unstructured Dark Web data by applying traditional classifier methods, showing that Random Forest could outperform Random Tree and Naïve Bayes models in certain contexts. Harnessing the strength of deep learning, Li *et al*. (2025) present a novel framework that converts dark web traffic into multi-channel images for classification using a 3D-CNN. The approach is innovative as it simultaneously captures spatial and temporal traffic features, offering a richer representation than traditional models. Their method achieved strong results, with accuracy, precision, recall, and F1-scores of 0.957, 0.931, 0.940, and 0.935, respectively, demonstrating the potential of deep learning in cyber threat detection. These studies highlight the potential of semi-supervised and deep learning methods when labelled data is scarce.

However, despite their advantages, deep learning-based models often require large annotated datasets and high computational resources, which are difficult to obtain for Dark Web data due to its covert, volatile, and often multilingual nature. In contrast, ensemble methods remain efficient, interpretable, and effective in low-resource scenarios. For example, Fang *et al*. (2019) used Random Forest classifiers to identify data breach trends in underground forums, achieving high performance without relying on deep architectures. Similarly, Saini and Bansal (2019) applied SVM and boosting techniques to detect weapon procurement activities.

Building upon these works, this study presents a voting ensemble of SVM, Random Forest, and Naïve Bayes to balance interpretability, computational efficiency, and robustness. Unlike prior studies that focused on a single illegal activity (such as data breaches or weapons procurement), our approach classifies six distinct categories of Dark Web activities while integrating an automated crawling pipeline and a real-time monitoring interface. This positions our work as a practical, scalable solution that bridges the gap between traditional ML and resource-intensive deep learning approaches.

2. METHODOLOGY

The method adopted in this project involves several steps, which include Data crawling, Data preprocessing, AI model building and training and Analysis of dark web content for crime investigation, as shown in Figure 2. The data crawling is fully discussed in section 3.1, the data preprocessing step is also presented in section 3.2, and the development of the Machine learning model is discussed in section 3.3

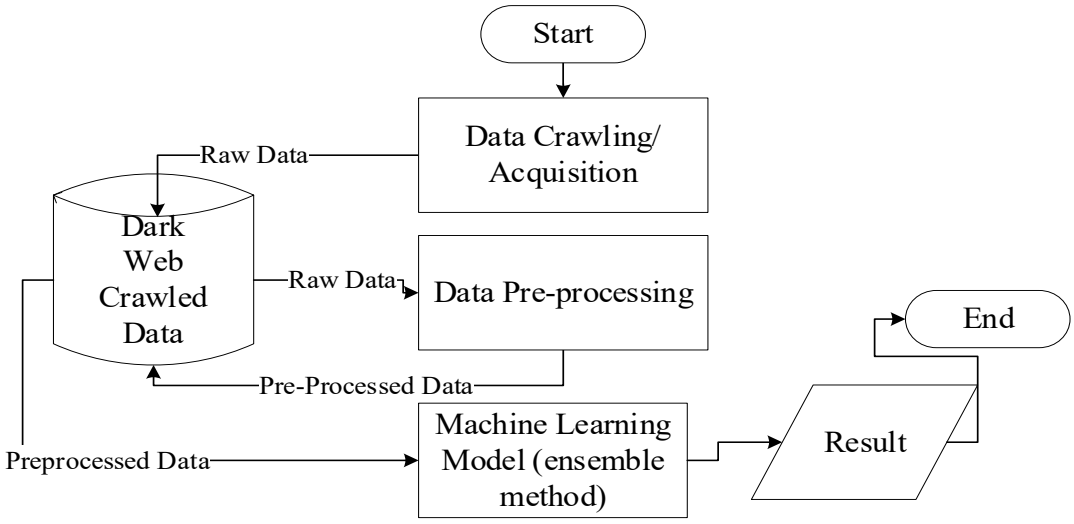

Figure 2: Conceptual design of the proposed methodology.

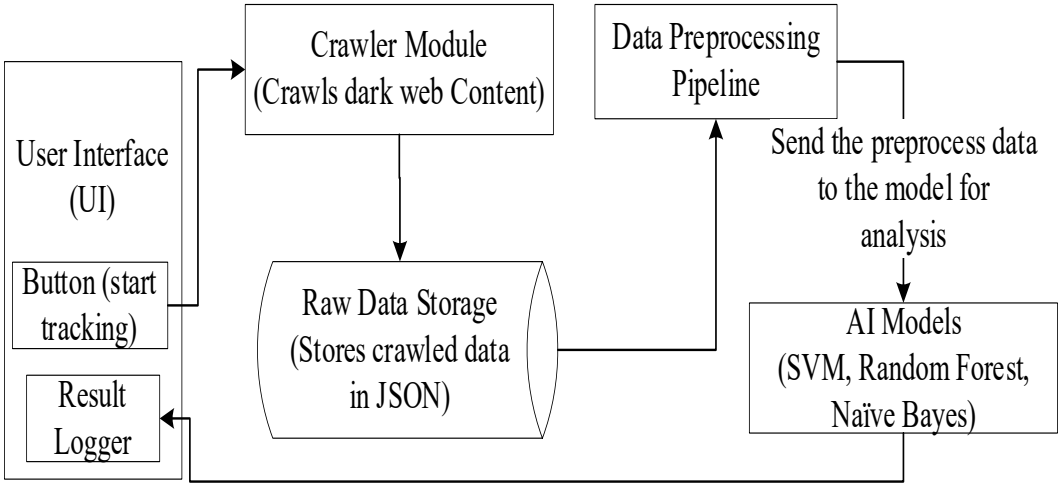

Figure 3: Architectural diagram of the proposed system.

The architectural diagram in Figure 3 shows how different modules/components of the system interact. The end user will have access to a graphics user interface containing a start button. When the "Start Tracking" button is clicked, the crawler module is triggered, and the output of the crawling operation is stored in the database. Concurrently, the data pre-processing pipeline fetches data from the database to prepare it for analysis by the machine learning model. The result of the model analysis is interpreted and displayed on the user interface under the "Result logger" section.

Crawling and data acquisition:

To build dataset for training the AI models, Data crawling becomes the first stage in this project. This entails harvesting data from the dark web using web scraping techniques. This study uses Puppeteer, a Node.js software that provides a high-level application interface for managing headless Chrome or Chromium browsers. As shown in Pseudocode 1, the headless browser was launched using the default SOCKS proxy address for TOR to successfully crawl the dark web. This will crawl the whole body of the onion site, and the extracted data is stored in JSON format for further processing. Each crawled onion site's content is saved as a JSON object, enabling structured data storage and easy retrieval for subsequent stages

Pseudocode 1: Data Crawling Pseudocode

```
Start
Initialise Puppeteer for web scraping
Define a list of URLs to crawl
For each URL in the list:
Navigate to the URL
Extract content
Store content in JSON format
Save all extracted data to a single JSON file
End
```

Data pre-processing:

Following a thorough crawl of the dark web and the extraction of raw data, attention turned to organizing and cleaning the data so that it could be processed and analyses within the AI-enhanced monitoring system. Python, which is well-regarded for its adaptability, was selected as the main language for data processing because of its strong libraries and features. Using the Python Natural Language Tool Kit (NLTK) package, we developed a text preprocessing function, as seen in Pseudocode 2, that will carry out a number of operations to clean and standardize the text data gathered from the dark web. In this process, newlines, non-alphabetic characters, and certain undesired characters will be eliminated. Tokenizing the data that was crawled and eliminating stop words from each sentence will also be performed.

Pseudocode 2: Data Pre-Processing Pseudocode

```
Start
Load JSON file with raw data
Define a function to preprocess the text:
    Replace newline characters with spaces
    Remove specific unwanted characters
    Tokenise text into sentences
    Initialise the pre-processed sentences' empty list
    For every sentence in the list:
      Perform Tokenisation of sentence into words
      Eliminate stop words
      Remove non-alphanumeric characters and lowercase all words
      Join words back into a sentence
Append the pre-processed sentence to list
Join pre-processed sentences into a single string
Return pre-processed text
Apply the preprocess function to each item in the dataset
Save pre-processed data to file
End
```

In total, the curated dataset consisted of 2,767 terms categorized into six distinct activities: Cyberattacks, Terrorist activities, Weapons trading, Drug trafficking, Human trafficking, and Normal/legal activities.

Each category was manually verified by domain experts to ensure accurate labelling. The dataset proportions varied across categories: normal/legal activities (18.3%), weapons trading (21.5%), counterfeit money (3.8%), human trafficking (11.5%), cyberattacks (14.5%), drug trafficking (14.6%), and terrorism (15.8%). This distribution is also illustrated in Figure 4.

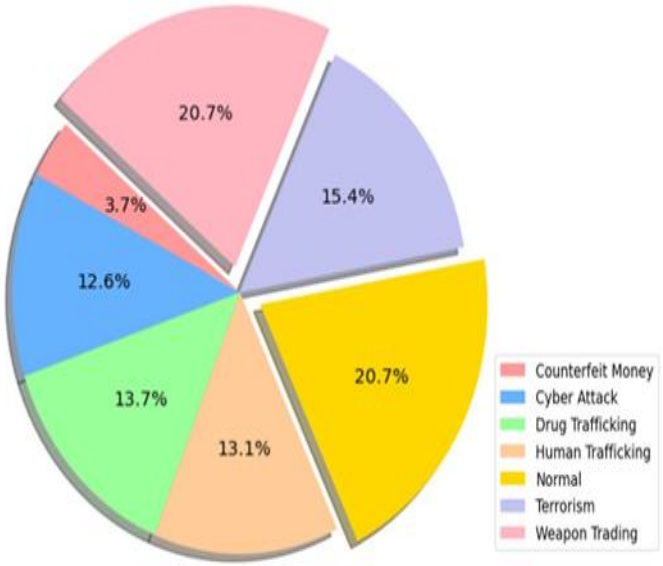

Figure 4: Dataset size distribution across dark web activities

A. Feature extraction and normalisation:

After data preprocessing with the Natural Language Toolkit (NLTK), which included stop word removal, tokenisation, and lemmatisation, the text was transformed into numerical representations using Term Frequency–Inverse Document Frequency (TF-IDF). To prevent longer documents from dominating the feature space, L2 normalisation was applied to the TF-IDF vectors.

B. Model building and training:

The processed dataset was used to train multiple machine learning classifiers, including SVM, RF, and NB. Each model was trained using its default parameters in scikit-learn, and their performances were evaluated and compared.

The choice of these algorithms was motivated by their suitability for text classification tasks:

i. SVM is effective in high-dimensional spaces such as TF-IDF vectors and is known for strong performance in text categorisation problems.
ii. Random Forest provides robustness by combining multiple decision trees, handling both nonlinear relationships and noise in the dataset.
iii. Naïve Bayes is widely used for text data due to its probabilistic foundation and efficiency, particularly with word frequency features.

Term Frequency-Inverse Document Frequency (TF-IDF) was adopted for text vectorisation, as it remains a simple, interpretable, and computationally efficient baseline widely used in text classification tasks.

To improve robustness, this study uses a voting ensemble strategy where Naive Bayes, Random Forest and SVM form the base classifiers. Pseudocode 3 shows that each model contributes to the overall prediction based on its unique skills. With this strategy, the system makes use of the advantages of each particular model, resulting in a more precise and trustworthy classification of content found on the dark web.

Pseudocode 3: Involvement of the machine learning model in the developed system

```
Start
Extract sentences that resemble these illicit activities, as well as normal activities, from various sources and store them in a CSV format
Load the curated dataset
Split data into training and testing sets
Vectorise text data using TF-IDF
Initialise SVM, Naive Bayes, and Random Forest models
Train each model in line 6 on the training data
Define voting ensemble method:
  Collect predictions from each model
  Combine predictions using majority voting
  Train the ensemble method using the training data
Evaluate the ensemble method on the testing data
Save trained models and vectorizer
End
```

4. RESULT AND DISCUSSION

The developed system was tested on real-world data from four DkW sites. An interface was created to achieve user friendliness and present findings. The dark web websites considered during testing are:
1.http://p53lf57qovyuvwsc6xnrppyply3vtqm7l6pcobkmyqsiofyeznfu5uqd.onion
2.http://wbz2lrxhw4dd7h5t2wnoczmcz5snjpym4pr7dzjmah4vi6yywn37bdyd.onion
3.http://7sk2kov2xwx6cbc32phynrifegg6pklmzs7luwcggtzrnlsolxxuyfyd.onion and
4.http://prjd5pmbug2cnfs67s3y65ods27vamswdaw2lnwf45ys3pjl55h2gwqd.onion
These websites are recognized as “onion site 1”, “onion site 2”, “onion site 3” and “onion site 4” respectively.

Analysis of experiment result:

Four models were trained on the dataset: Random Forrest, Naive Bayes, SVM, and a Voting Classifier that combines the predictions of the three models. Each model was trained using an 80-20 train-test split. The model evaluation metrics considered are accuracy, precision, recall, and F1-scores.

For the Random Forest algorithm, Figure 5 shows the performance, as it gives 96% accuracy. The macro average of F1-score and recall is 96% each, while the precision is 97%. The weighted average of precision, recall, and F1-score is 96% each. As shown in Figure 5, the Random Forest model achieved 96% accuracy, with a macro-average F1-score and recall of 96% each, and a precision of 97%.

Random Forest Report:	Precision	Recall	F1-score	Support
Counterfeit Money	1.00	1.00	1.00	21
Cyber Attack	0.99	0.97	0.98	68
Drug Trafficking	0.94	0.89	0.92	76
Human Trafficking	1.00	0.92	0.96	73
Normal	0.97	1.00	0.99	112
Terrorism	0.99	0.93	0.96	91
Weapon Trading	0.88	0.98	0.93	113
Accuracy			0.96	554
Macro Average	0.97	0.96	0.96	554
Weighted Average	0.96	0.96	0.96	554

Figure 5: Performance metrics of Random Forest classifier.

Figure 6 shows that SVM gives 97% for accuracy, weighted average of precision, recall and F1-score. Macro average of recall and F1-score. The macro average of precision is 98%.

SVM Report:	Precision	Recall	F1-score	Support
Counterfeit Money	1.00	0.95	0.98	21
Cyber Attack	0.96	1.00	0.98	68
Drug Trafficking	0.96	0.95	0.95	76
Human Trafficking	0.99	0.95	0.97	73
Normal	0.99	1.00	1.00	112
Terrorism	0.98	0.93	0.96	91
Weapon Trading	0.96	1.00	0.98	113
Accuracy			0.97	554
Macro Average	0.98	0.97	0.97	554
Weighted Average	0.77	0.97	0.97	554

Figure 6: Performance metrics of Support Vector Machine Classifier the Naïve Bayes algorithm is shown in Figure 7. In this study, Naïve Bayes gives 86% accuracy. The macro average of F1-score and recall is 85% and 83% respectively, while the macro average precision is 87%. The weighted average of precision and F1-score is 87% each, and recall is 86%.

Report:	Precision	Recall	F1-score	Support
Counterfeit Money	1.00	0.62	0.76	21
Cyber Attack	0.89	0.94	0.91	68
Drug Trafficking	0.69	0.72	0.71	76
Human Trafficking	0.77	0.88	0.82	73
Normal	0.98	0.96	0.97	112
Terrorism	0.87	0.85	0.86	91
Weapon Trading	0.92	0.88	0.90	113
Accuracy			0.86	554
Macro Average	0.87	0.83	0.85	554
Weighted Average	0.87	0.86	0.87	554

Figure 7: Performance metrics of Naïve Bayes classifier.

The ensemble voting classifier achieved an overall accuracy of 97%, with macro-average recall and F1-score both at 97%, and macro-average precision at 98%. The weighted average of precision, recall, and F1-score was 97% each (Figure 8).

The voting ensemble classifier, which is the central focus of this study, achieved strong results as shown in Figure 8. The classifier's performance closely aligns with that of the SVM model but provides slightly more balanced metrics.

Specifically, the ensemble model attained an overall accuracy of 97%, a macro average recall and F1-score of 97%, and a macro average precision of 98%. The weighted averages of precision, recall, and F1-score were all 97%, demonstrating the robustness of the proposed approach.

Ensemble Report:	Precision	Recall	F1-score	Support
Counterfeit Money	1.00	0.95	0.98	21
Cyber Attack	0.96	1.00	0.98	68
Drug Trafficking	0.96	0.97	0.97	76
Human Trafficking	0.97	0.97	0.97	73
Normal	1.00	1.00	1.00	112
Terrorism	0.98	0.91	0.94	91
Weapon Trading	0.97	0.99	0.98	113
Accuracy			0.97	554
Macro Average	0.98	0.97	0.97	554
Weighted Average	0.97	0.97	0.97	554

Figure 8: Performance metrics of the ensemble voting classifier.

Error analysis and confusion matrix:

While overall performance metrics (accuracy, precision, recall, and F1-score) demonstrate the robustness of the ensemble model, a closer examination of the confusion matrices provides further insights. Figure 9 illustrates the confusion matrix for the ensemble classifier, highlighting the system's ability to distinguish between cyberterrorism, weapons trading, drug trafficking, human trafficking, terrorism, and normal/legal activities. The majority of misclassifications occur between cyberattack and terrorism, which share overlapping linguistic patterns, and between drug trafficking and human trafficking, where slang and coded terms sometimes resemble each other.

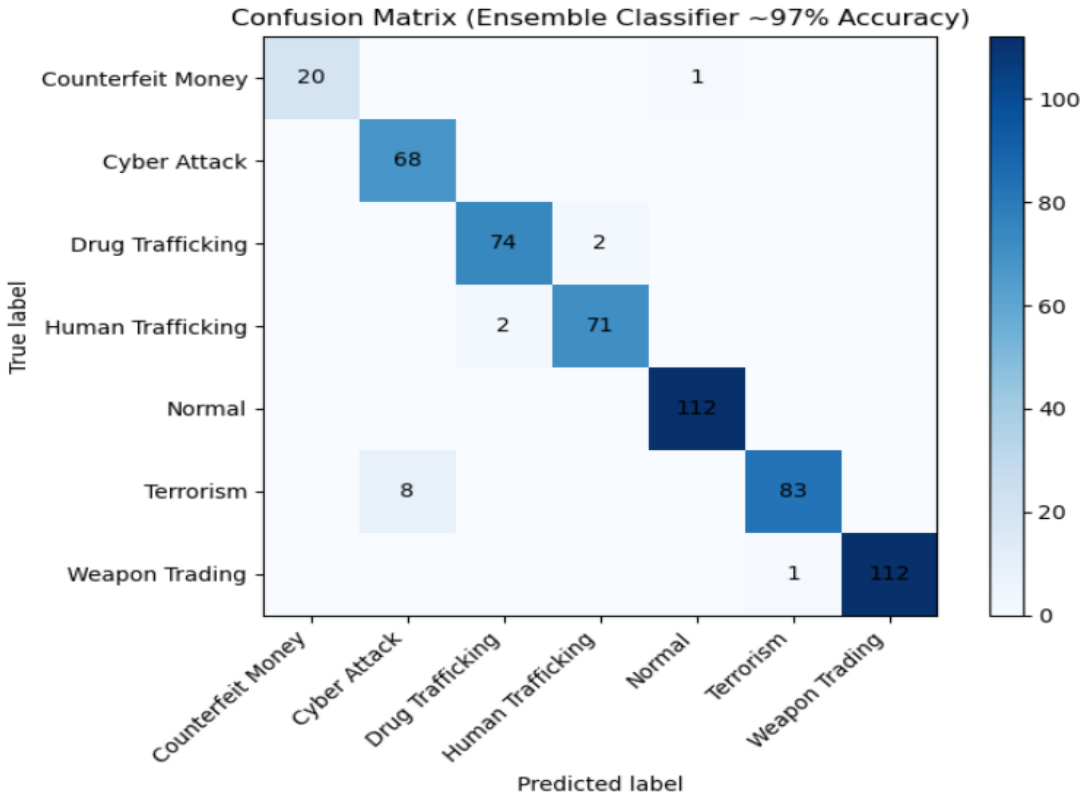

Figure 9: Confusion matrix for the ensemble classifier.

Also, False Positives (FP, flagging normal content as illegal) are particularly critical in real-world deployments. For instance, a false positive on a legitimate forum could lead to unnecessary investigations, wasting resources. Conversely, False Negatives (FN, failing to detect illegal activity) pose a greater threat, as they allow criminal activities to go unnoticed. In our results, false negatives were comparatively fewer, but their impact emphasizes the importance of integrating human-in-the-loop validation for high-risk categories such as terrorism and trafficking.

Table 2 summarizes the FP, FN, True Positive (TP) and True Negative (TN) distribution per category to complement the confusion matrix.

Table 2: Ensemble classifier TP, FN, FP and TN distribution per category

Class	TP	FN	FP	TN
Counterfeit Money	20	1	0	533
Cyber Attack	68	0	0	486
Drug Trafficking	74	2	0	478
Human Trafficking	71	2	2	479
Normal	112	0	0	442
Terrorism	83	8	0	463
Weapon Trading	112	1	0	441

In Table 3, the results obtained during the experiment of this study are presented and compared with the study of He *et al*. (2019). Similar to this study, He *et al*. (2019) also worked on classifying illegal activities that occur within the dark web. The comparison shows that this study achieved a better result compared to the work of He *et al*. (2019).

Table 3: Comparison of the proposed study results with previous studies

Author	Classifier	Accuracy	Precision	Recall	F1-Score
This Study	**Ensemble (voting)**	**0.97**	**0.97**	**0.97**	**0.97**
	Random Forest	0.96	0.97	0.976	0.96
	NB	0.86	0.87	0.86	0.87
	SVM	0.973	0.973	0.973	0.973
He *et al*., (2019)	Multinominal NB	0.935	0.938	0.935	0.935
	SGD	0.880	0.891	0.880	0.881
	SVM	0.687	0.792	0.687	0.627
Zenebe *et al*., (2019)	Random Forest	0.783	*	*	*
	Random Tree	0.675	*	*	*
	Naïve Bayes	0.528	*	*	*
Ebrahimi *et al*., (2020)	TSVM+LSTM	0.947	0.833	96.77	89.55
	SVM	0.916	0.763	0.935	0.840
	Random Forest	0.878	0.674	0.935	0.783

In addition to the comparison with He *et al*. (2019), we extended our analysis to include results reported in Ebrahimi *et al*. (2020) and Zenebe *et al*. (2019). While Ebrahimi *et al*. (2020) investigated semi-supervised learning and deep learning approaches for Dark Web threat detection, our ensemble method achieved comparable or better results without requiring the large-scale labelled datasets and computational resources typically needed for deep neural architectures. For instance, Ebrahimi *et al*. (2020) reported an F1-score improvement of 89.55% using transudative learning with semi-supervised labelling, whereas our voting ensemble achieved an F1-score of 97% on a smaller but domain-focused dataset. Similarly, Zenebe *et al*. (2019) highlighted the effectiveness of Random Forest for dark web intelligence extraction, yet their reported accuracy (around 78%) falls below the 97–98% achieved by our ensemble model.

The observed improvement arises from the complementary strengths of the base classifiers. Naïve Bayes contributes robustness on sparse features, Random Forest captures non-linear dependencies, while SVM provides strong generalization in high-dimensional spaces. By combining these through majority voting, our approach reduces the bias–variance trade-off inherent in single models and mitigates misclassification errors, especially across underrepresented classes such as human trafficking and counterfeit money. This explains why the ensemble consistently outperformed individual classifiers on our dataset.

Furthermore, unlike prior work, our contribution extends beyond algorithm selection to an integrated pipeline involving (i) automated crawling of onion sites, (ii) domain-specific preprocessing of Dark Web text, and (iii) development of a practical monitoring interface. Hence, the novelty of this work lies in applying a tailored ensemble learning approach within a full-stack Dark Web monitoring system, rather than proposing an entirely new algorithm.

Interface development:

An interface to facilitate the ease of use of the model on crawled data was developed. The interface shown in Figure 10 allows users to initiate the analysis with the click of a button, “Start Tracking”, which makes the status of the system change from “Pending” to “Tracking”. The necessary operation is performed under the hood, and the status changes to “Done” at the end of the process. The result is indicated under “Alert” and “Attack”. The “Alert” changes from between None, Red and Green, where “None” stands for no operation performed, “Red” stands for “illegal activity found” and “Green” stands for “normal or legal activities found”. The “Attack” sections indicate the specific type of illegal activities found; otherwise, it indicates “normal”. Figure 10 also shows the result for the four onion sites investigated, with only “onion site 2”

being normal. The "onion site 1", "onion site 3" and "onion site 4" have traces of illegal activities, which are Cyber Attack, Counterfeit Money and Human Trafficking, respectively. This user-friendly interface enhances the accessibility and usability of the system for monitoring dark web activities.

Figure 10: User interface of the developed dark web monitoring system.

Summary and Conclusion:

The internet consists of three layers: the surface web, the deep web, and the dark web. The latter is the most obscure and impossible to find using regular search engines; it can only be accessed via programs such as TOR. Because of its anonymity, it fosters the spread of illegal activities such as identity theft, cyberstalking, phishing, cyberterrorism, weapon sales, human trafficking, and malware distribution. While current cybersecurity solutions struggle to handle the complex threats emerging from the Dark Web, developing an AI-enables system capable of analyzing and monitoring dark web content and detecting risks at an early stage is critical to combating this issue. Our program aims to overcome this gap by establishing an intelligent system that proactively identifies and stops potential cyber threats using machine learning, natural language processing, and anomaly detection. This innovative approach attempts to improve internet safety by protecting user data and cultivating a more secure online community by stopping cybercrimes before they worsen.

In conclusion, this project developed an AI model capable of monitoring and analyzing material on the dark web for criminal activities and preventing attacks from occurring. The model was trained with a self-acquired dataset of 2767 terms related to six different dark web activities: cyberterrorism, weapon trading, drug trafficking, human trafficking, terrorist activities, and regular/legal activities. This work used machine learning models, including Support Vector Machine, Naïve Bayes, and Random Forest, to create an ensemble voting technique. Only one of the four onion sites used for real-time testing of the system was not involved in any unlawful activities. The remaining three (3) sites were flagged by the developed system for the creation of counterfeit money, cybercrime, and human trafficking, respectively. The ensemble model gives 97% accuracy and a macro average precision of 98%. The macro average for recall and F1-score is 97%.

The precision, recall, and F1-Score all have the same weighted average of 97%. This presented a better result as compared to the work of He *et al*. (2019), Zenebe *et al*. (2019), and Ebrahimi et al. (2020).

These findings indicate that while the model performs well overall, it is limited to the use of TF-IDF which does not capture semantic relationships between words. Future work should focus on reducing overlap-related errors by incorporating context-aware embeddings such as Word2Vec, GloVe, and contextual embeddings like BERT and expanding the dataset with more domain-specific terminology. This will enhance reliability in operational environments, enhance semantic understanding and improve model generalization.

Acknowledgment:

We would like to express our gratitude to Professor Musa A. Aibinu (Vice Chancellor of Summit University Offa) for sharing his experience and inspiring this research work.

Ethical statement:

The study did not require ethical approval as it involved neither human nor animal subjects.

Authors Contributions:

All authors contributed to the conception and design of the study. A.G.A. and Y.O.O. led the conceptualization, methodology design, and overall supervision of the

research. K.Y.O., G.K.A.Y., and M.A.A., contributed to data collection, analysis, and model development. A.A.T., L.O.O., and O.Y.A., supported experimentation, validation, and interpretation of results. A.A.O.T., contributed to the literature review and manuscript drafting. All authors reviewed, revised, and approved the final version of the manuscript.

REFERENCES

Alaidi, A. H. M., Al_airaji, R. M., Alrikabi, H. T. S., Aljazaery, I. A., & Abbood, S. H. (2022). Dark Web Illegal Activities Crawling and Classifying Using Data Mining Techniques. *International Journal of Interactive Mobile Technologies (IJIM)*, *16*(10), 122–139. https://doi.org/10.3991/ijim.v16i10.30209

Ali, A., & Qasim, M. (2023). Terrorist Acts on the Surface and Dark Web. In *Dark World* (pp. 84–101). CRC Press. https://doi.org/10.1201/9781003404330-7

Alshammery, M. K., & Aljuboori, A. F. (2022). Crawling and Mining the Dark Web: A Survey on Existing and New Approaches. *Iraqi Journal of Science*, 1339–1348. https://doi.org/10.24996/ijs.2022.63.3.36

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

Ebrahimi, M., Nunamaker, J. F., & Chen, H. (2020). Semi-Supervised Cyber Threat Identification in Dark Net Markets: A Transductive and Deep Learning Approach. *Journal of Management Information Systems*, *37*(3), 694–722. https://doi.org/10.1080/07421222.2020.1790186

Fang, Y., Guo, Y., Huang, C., & Liu, L. (2019). Analyzing and Identifying Data Breaches in Underground Forums. *IEEE Access*, *7*, 48770–48777. https://doi.org/10.1109/access.2019.2910229

Ghosh, S., Das, A., Porras, P., Yegneswaran, V., & Gehani, A. (2017). Automated Categorization of Onion Sites for Analyzing the Darkweb Ecosystem. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1793–1802. https://doi.org/10.1145/3097983.3098193

He, S., He, Y., & Li, M. (2019). Classification of Illegal Activities on the Dark Web. In *Proceedings of the 2019 2nd International Conference on Information Science and Systems* (Vol. 54, pp. 73–78). ACM. https://doi.org/10.1145/3322645.3322691

Huang, C., Guo, Y., Guo, W., & Li, Y. (2021). HackerRank: Identifying key hackers in underground forums. *International Journal of Distributed Sensor Networks*, *17*(5), 155014772110151. https://doi.org/10.1177/15501477211015145

Kaur, S., & Randhawa, S. (2020). Dark Web: A Web of Crimes. *Wireless Personal Communications*, *112*(4), 2131–2158. https://doi.org/10.1007/s11277-020-07143-2

Li, J., Pan, Z., & Jiang, K. (2025). A Three-Dimensional Convolutional Neural Network for Dark Web Traffic Classification Based on Multi-Channel Image Deep Learning. Computers, 14(8), 295. https://doi.org/10.3390/computers14080295

Mahmood, I., Rahman, M. A., Kabir, M. A., & Shahriar, M. (2022). *A Survey on Dark Web Monitoring and Corresponding Threat Detection*. *October*. https://www.researchgate.net/publication/364898748

Saini, J. K., & Bansal, D. (2019). A Comparative Study and Automated Detection of Illegal Weapon Procurement over Dark Web. *Cybernetics and Systems*, *50*(5), 405–416. https://doi.org/10.1080/01969722.2018.1553591

Schafer, M., Fuchs, M., Strohmeier, M., Engel, M., Liechti, M., & Lenders, V. (2019). BlackWidow: Monitoring the Dark Web for Cyber Security Information. In *2019 11th International Conference on Cyber Conflict (CyCon)*. IEEE. https://doi.org/10.23919/cycon.2019.8756845

Wiafe, I., Koranteng, F. N., Obeng, E. N., Assyne, N., Wiafe, A., & Gulliver, S. R. (2020). Artificial Intelligence for Cybersecurity: A Systematic Mapping of Literature. *IEEE Access*, *8*, 146598–146612. https://doi.org/10.1109/access.2020.3013145

yang, Y., Zhu, G., Yang, L., & yu, H. (2020). Crawling and Analysis of Dark Network Data. In *Proceedings of 2020 6th International Conference on Computing and Data Engineering* (Vol. 24, pp. 116–120). ACM. https://doi.org/10.1145/3379247.3379272

Zenebe, A., Shumba, M., Carillo, A., & Cuenca, S. (2019). Cyber Threat Discovery from Dark Web. *EPiC Series in Computing*, 174–163. https://doi.org/10.29007/nkfk