International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

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

Generating Voice Text of Cyber Crime in Explainable AI Using Large Language Model

C. Syamsundar Reddy[1], Prof. G. Anjan Babu[2]

Submitted: 28/12/2023 **Revised**: 04/02/2024 **Accepted**: 12/02/2024

Abstract: Machine learning is an AI application that mimics human intelligence's ability to learn from experience, particularly in pattern recognition. This is crucial in criminal justice, as AI aims to replicate human capabilities in software algorithms and hardware, such as identifying people, performing complex tasks, and making predictions. Data leaks are increasing due to work-life balance changes and remote working, with workers accounting for 22% and attackers 23%. Cybercriminals use social engineering techniques like phishing to trick employees into providing sensitive information. Regular cybersecurity awareness training and fostering a positive workplace culture are essential for preventing data leaks. This study examines how cybercriminals impersonate real-world emails, voice and IT help desks in order to fool workers into disclosing critical information. They do this by using phishing and one-on-one social engineering techniques. Through the use of a large language model to generate speech text for electronic crime using explainable artificial intelligence. The study demonstrated strong predictive performance with an imbalanced cybercrime dataset, but identified limitations in the ROC AUC metric, which compares True Positive Rate to False Positive Rate.

Keywords: *Cyber Crime, Explainable Artificial Intelligence, Generative Voice Text, Large Language Model etc.*

1. Introduction: Generative AI, a rapidly growing technology, produces high-quality content like text, images, and audio in seconds through user-friendly interfaces. Generative AI, introduced in the 1960s, was first used in chatbots but only in 2014 with the introduction of GANs, enabling the creation of authentic human images, videos, and audio. New capabilities in film dubbing and educational content offer opportunities, but also raise concerns about deepfakes and cybersecurity attacks on businesses. Transformers and breakthrough language models have significantly contributed to the mainstream adoption of generative AI. Transformers enable researchers to train large models without labelling data in advance, resulting in more detailed answers. They also unlock attention, enabling models to track connections across pages, chapters, and books. Large language models (LLMs) have revolutionized AI, enabling generative models to write engaging text, paint photorealistic images, and create sitcoms. Multimodal AI innovations enable content generation across multiple media types, enabling tools like Dall-E. Generative AI uses inputs like text, images, videos, and designs to process and generate new content. Initially, it required API submission and Python language. Pioneers now offer plain language user experiences and allow customization of content style, tone, and other elements. This technology can be used for essays, problem-solving, or realistic fakes. Generative AI models use various algorithms to represent and process content, such as text and images. Natural language processing techniques transform raw characters into sentences and visual elements. Developers apply neural networks like GANs and VAEs to generate new content, such as realistic human faces or synthetic data for AI training.

Generative AI can improve business workflows by automating content creation, reducing email responses, improving technical queries, and creating realistic representations. It can also simplify content style and narrative, making it easier to understand and respond to complex information. Generative AI has limitations due to its specific use cases and the difficulty in identifying content sources. It may not always identify the source, assess bias, and identify

Research Scholar,
e-mail: cssreddi@gmail.com
e-mail: gabsvu@gmail.com
Department of Computer Science,
Sri Venkateswara University College of Commerce, Management and Computer Science,
Sri Venkateswara University, Tirupati

inaccurate information. Realistic-sounding content can make it harder to identify inaccurate information. Tuning for new circumstances can be challenging, and results may gloss over bias, prejudice, and hatred. Generative AI raises concerns about its quality, potential misuse, and disruption of business models. Issues include providing inaccurate information, making trust difficult, promoting plagiarism, disrupting search engine optimization and advertising, enabling fake news generation, and potentially impersonating people for social engineering cyber-attacks. These concerns highlight the need for careful consideration and regulation. Generative AI technologies, often compared to general-purpose technologies, can significantly impact various industries. However, implementing these technologies requires decades of effort to optimize workflows, rather than just speeding up small portions. Generative AI is a creative tool that creates original content, chat responses, designs, synthetic data, or deepfakes. It uses neural network techniques like transformers, GANs, and VAEs. Traditional AI algorithms follow predefined rules to process data. Generative AI is better for tasks involving Natural Language Processing (NLP) and content creation, while traditional algorithms are more effective for rule-based processing and predetermined outcomes.

New research indicates that employees are increasing the frequency of data leaks, leading to a close race between employee actions and cyberattacks. The study on IT Security Economics found that workers account for 22% of data leaks, while attackers account for 23%. Post-pandemic work-life balance changes and remote working have increased this percentage. Employee-triggered leaks often involve ignoring cybersecurity policies, but over 36% of them are deliberate acts of sabotage or espionage, according to security managers. Cybercriminals use social engineering techniques, such as phishing, to trick employees into providing sensitive information. This sophisticated form of phishing mimics real-life emails and emails and is more targeted and personalized. Cybercriminals can also engage in one-on-one social engineering, pretending to be from the IT support desk, to persuade employees to share login and password details. Cyber hygiene involves implementing everyday practices to prevent cybercriminals from accessing systems, such as using complex passwords, using VPNs, and deleting unnecessary digital data. Cybersecurity awareness training teaches employees to recognize and avoid social engineering tricks. The research shows businesses with strong cyber skills programs report better overall preparedness. Airport's program focuses on long-term behaviour change, targeting employees with mock phishing emails to identify those at risk. Regular training is essential for a company-wide cybersecurity culture. Cybersecurity awareness training should foster 'cyber pride', encouraging positive motivation for good cyber behaviour, rather than instilling fear. Comprehensive cybersecurity education may not prevent 36% of employee-generated data leaks, as disgruntled employees deliberately share customer data and hand access keys to cybercriminals. Regular user access reviews are important, but maintaining a positive workplace culture is more crucial. Leaders should 'temperature check' their organization by involving all staff in anonymous feedback mechanisms about their managers. This can help uncover hidden attitudes and views. The rise of data leaks should prompt businesses to focus on workplace culture and employee wellbeing. Regular cybersecurity education, strong cyber hygiene, and understanding employee feelings should be central to data protection strategies.

The study explores the process of creating voice-uncertain text in cybercrime and its impact on various voice-speaking applications. The format of the paper are as follows: Part 1 contains the introduction. Deep Learning for the Recognition of Speech to text is provided in this article's Part 2. Sections 3 Procedure of Encryption of Voice Speaking and 4 Generating Text of Live Criminal Voice Using Naïve Bayes with Large Language Model of the proposed system. The section 5 denotes Imbalanced Text Generating Criminal Voice Uncertain Text Data. The last one is the conclusion of Section 6, which is cited in Section 7.

2. Voice Speaking to Text Conversion Using NLP: Text-to-speech (TTS) is an essential part of systems for interaction between humans and machines that converts text into spoken sounds. The two steps involved are text processing, which converts input text into phonetic representations, and audio waveform synthesis shown in figure-1. The goal of using novel natural language processing (NLP) approaches to improve automatic voice recognition. By comparing deep learning methods based on quality standards, Multimedia

Tools and Applications examines their platforms, researcher contributions, and text-to-speech system lifecycles.

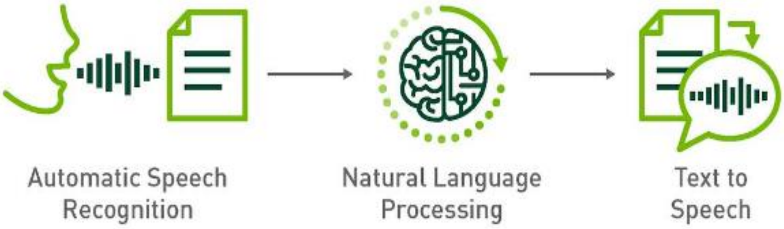

Fig 1: Instinctive Speech Recognition

In the field of cutting-edge technology and dependable infrastructure, the creation of an Automatic Speech Recognition (ASR) system is being investigated. By enabling machines to easily comprehend and translate spoken language, automatic speech recognition is changing accessibility, voice assistants, and a wide range of other uses in technology. A state-of-the-art solution for self-supervised speech representation learning that expands on ASR's capabilities is the Wav2Vec 2.0 framework. The Cloud GPU server from E2E will be utilized to handle the ASR system in an effective and smooth manner. This offers a thorough how-to on constructing a successful ASR system, including component configuration and cloud-based model optimization for speech recognition.

The presents a novel method for generating high-quality text embeddings using synthetic data and less than 1k training steps, utilizing proprietary Large Language Models, fine-tuning LLMs, and achieving strong performance without labeled data. Text embeddings are natural language representations that encode semantic information, widely used in Natural Language Processing (NLP) tasks like information retrieval and question answering. Information retrieval methods include Text Embeddings, synthetic data generation, and Large Language Models (LLMs).

3. Procedure of Encryption of Voice Speaking:

Encryption is a crucial tool in the ever-changing digital landscape, serving as a reliable wingman to protect your data. Data protection from illegal access and manipulation is achieved through the process of encryption. It is a method of encrypting data such that only the right "key" can be used to decrypt it and read it shown in figure-2.

1.Let's say we are working with a team that is both offshore and onshore. In order for us to interact with each other, we would need a network bridge over here. As an illustration, Aadhya receives the communication from offshore and relays it to the person onshore, but she modifies it. Chinna did not attend the summit as a result. Since the message must go through a third party, there is no privacy between the source and the destination.

Golden Rule: We should never trust anyone with data. Always protect your private information.

2. Raj transmits the encrypted communication to Aadhya, which then forwards it to Chinna in order to prevent data manipulation. However, how is the encrypted message decrypted? Raj will provide Chinna with the shared secure key, and Chinna will use the encrypted key it got to decrypt the message. Adhya, the middleman, won't be able to decipher the message.

3. The message is being encrypted by Raj and sent to Chinna in an outer layer. The message is being ciphered or decrypted using the encryption key, and Chinna will use the encryption key to decrypt it so that it can be read.

Golden Rule: To secure our data, we need to make sure that the enclosing layer hides the real data inside of it.

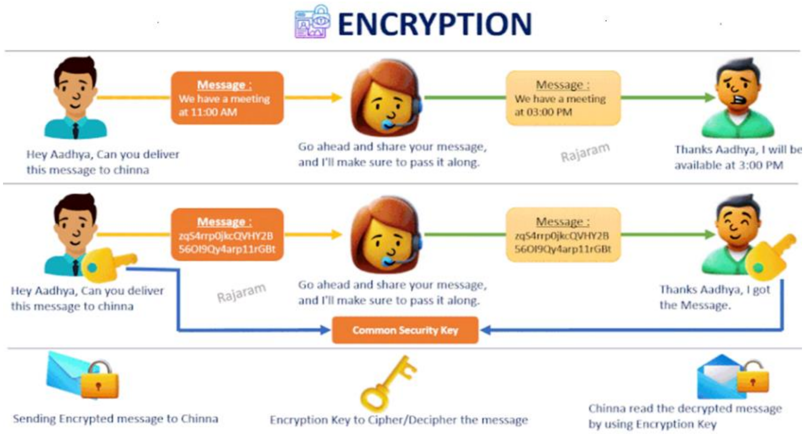
ENCRYPTION
Message :
We have a meeting at 11:00 AM
Message :
We have a meeting at 03:00 PM
Hey Aadhya, Can you deliver this message to chinna
Go ahead and share your message, and I'll make sure to pass it along.
Rajaram
Thanks Aadhya, I will be available at 3:00 PM
Message :
zqS4rrp0jkcQVHY2B 56OI9Qy4arp11rGBt
Message :
zqS4rrp0jkcQVHY2B 56OI9Qy4arp11rGBt
Hey Aadhya, Can you deliver this message to chinna
Rajaram
Go ahead and share your message, and I'll make sure to pass it along.
Thanks Aadhya, I got the Message.
Common Security Key
Sending Encrypted message to Chinna
Encryption Key to Cipher/Decipher the message
Chinna read the decrypted message by using Encryption Key

Fig 2: Encryption Procedure of Voice Message

4. Generating Uncertain Text of Live Criminal Voice Using Naïve Bayes with Large Language Model: LLMs are optimized for predicting the probability of the next token, but how to generate text is a complex task. The naive bayes method is to select the word with the highest probability, auto-regress, and use the probability vector that the model generates. This is the avaricious method, although it often results in extended, repeating statements that become unintelligible. An alternative method would be to sample the words according to the model's generated probability. Adjusting a temperature parameter typically modifies the degree of unpredictability in this process. As a result, phrases can be produced that are more inventive and less repetitious. The goal of generating a sentence is to maximize the probability of the entire output sequence, not just the next token.

$$\mathbf{P}(\textbf{Output sequence }|\textbf{Prompt}) \qquad -- \rightarrow (1)$$

Fortunately, this probability can be expressed as a product of the likelihoods of predicting the following token:

P (token-1,...,token-N | Prompt) = P(token-1| Prompt) x,...,P(token-N|Prompt, token-1,..., token-N -1)

However, it is NP-hard to solve this particular problem. The problem can be approximated by selecting k candidate tokens at each iteration, testing them, and retaining the sequences that maximize the probability of the entire sequence. The sequence with the highest probability should be chosen. This is referred to as beam search generation, and it can be used with the multinomial and greedy approaches.

Contrastive search is a strategy that incorporates additional metrics like variety or fluency. The process involves selecting candidate tokens at each iteration, penalizing the probability using a previous generation of token similarity measure, and selecting the tokens with the highest new score shown in figure-3.

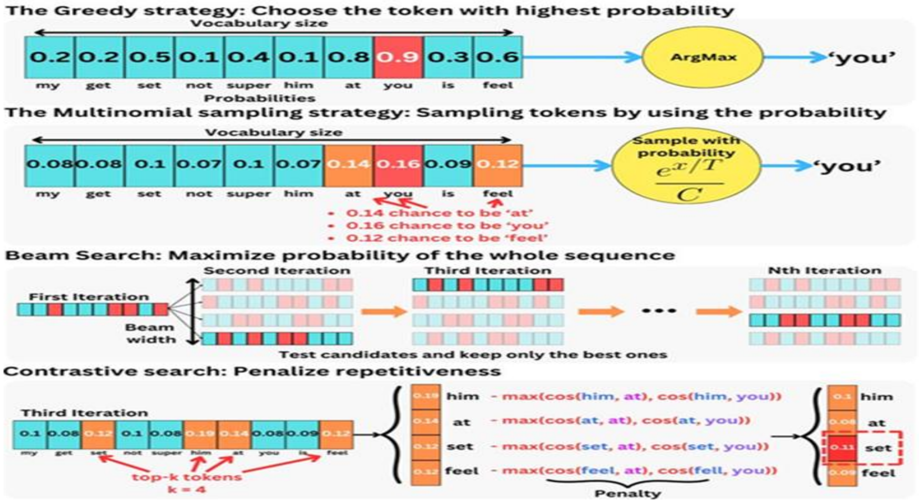
The Greedy strategy: Choose the token with highest probability
Vocabulary size
0.2 0.2 0.5 0.1 0.4 0.1 0.8 0.9 0.3 0.6
my get set not super him at you is feel
Probabilities
ArgMax
'you'
The Multinomial sampling strategy: Sampling tokens by using the probability
Vocabulary size
0.08 0.08 0.1 0.07 0.1 0.07 0.14 0.16 0.09 0.12
my get set not super him at you is feel
0.14 chance to be 'at'
0.16 chance to be 'you'
0.12 chance to be 'feel'
Sample with probability $\frac{e^{x/T}}{C}$
'you'
Beam Search: Maximize probability of the whole sequence
First Iteration
Second Iteration
Third Iteration
Nth Iteration
Beam width
Test candidates and keep only the best ones
Contrastive search: Penalize repetitiveness
Third Iteration
0.1 0.08 0.12 0.1 0.08 0.19 0.14 0.08 0.09 0.12
my get set not super him at you is feel
top-k tokens
k = 4
0.19 him - max(cos(him, at), cos(him, you))
0.14 at - max(cos(at, at), cos(at, you))
0.12 set - max(cos(set, at), cos(set, you))
0.12 feel - max(cos(feel, at), cos(fell, you))
Penalty
0.1 him
0.08 at
0.11 set
0.09 feel

Fig 3: Text Generation of Criminal Voice Using Large Language Model

5.Imbalanced Text Generating Criminal Voice Uncertain Text Data: The study demonstrated strong predictive performance with an imbalanced criminal dataset, but identified limitations in the ROC AUC metric, which compares True Positive Rate to False Positive Rate.

The formulas provided are as follows:

$$TPR = \frac{True\ Positives}{Positive} = \frac{TP}{P} \qquad - - \rightarrow (2)$$

$$FPR = \frac{False\ Positives}{Negative} = \frac{FP}{N} \qquad - - \rightarrow (3)$$

AUC determines maximum separation between negative and positive predictions. An imbalanced class threshold of 0.5 may not achieve maximum separation, but a 0.9 threshold in inference should work fine. The curve is drawn by ranking the probability score of the model and computing those rates for each value of the score used as a threshold to assign the classes by Fawcett, T,2006 [1].

A low threshold (e.g., p=0.1 => 1 if p > 0.1, 0 otherwise) will lead to a high number of TP but also a high number of FP. A high threshold (e.g., p=0.9 => 1 if p > 0.9, 0 otherwise) will lead to a low number of TP but also a low number of FP. In fact, at p=0, we have:

TP = P => TPR = 1

FP = N => FPR = 1

and for p=1, we have

TP = 0 => TPR = 0.

FP = 0 => FPR = 0.

Now, what happens in the case of imbalanced classes? We have N >> P. As a consequence, the probability score distribution will tend to be skewed toward smaller values: most of the scores will be small. Most of the true positive samples will be on the high side of the probability score. Let's consider the following label (c) and resulting score (p) lists:

Move the threshold from left to right

c = [0,0, 0, 0, 0, 1,0]

p = [0.05, 0.06, 0.07, 0.8, 0.15, 0.3, 0.5]

Thresholds typically involve a negative sample transition, with TPR jumping down only when transitioning to a positive sample, and FPR dropping down minimally when transitioning to a negative one.

Therefore, for the majority of criteria, The TPR will be approximately 1(TPR ~ 1), while FPR will progressively decrease to 0 almost linearly, making FPR uninformative. Only at the highest values of the score will the TPR begin to fall. When it starts to fall, we have 1/ N << 1/ P, so when the increments are non-zero, we have D (FPR) << D (TPR), as FPR ~ 1/N and TPR ~ 1/P. Therefore, the ROC AUC looks artificially like the curve of a perfect classifier when the imbalance is high. The Precision-Recall curve is plotted as a function of recall for various thresholds.

$$\boldsymbol{Precision = \frac{True\ Positive}{Predicted\ Positives}} \qquad - - \rightarrow (\mathbf{4})$$

$$\boldsymbol{Recall = TPR = \frac{True\ Positive}{Positive}} \qquad - - \rightarrow (\mathbf{5})$$

Precision-Recall AUC is a dynamic metric for imbalanced data, as it's more invariant to class distribution changes, making it a more actionable metric for highly imbalanced data by Provost, F,2013[2].

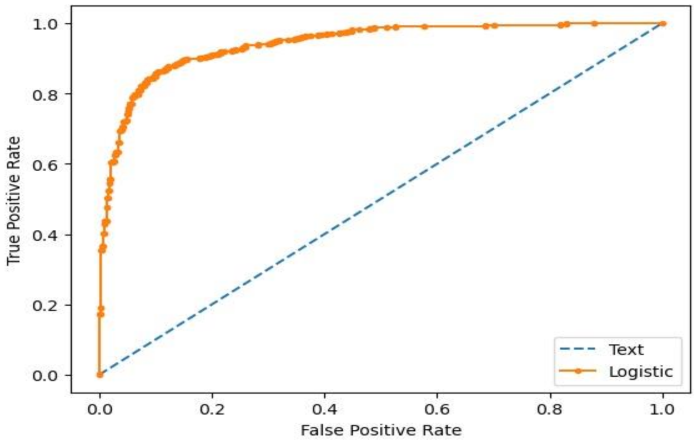

Fig 4: ROCAUC Curve

Text: ROC AUC=0.500

Logistic: ROC AUC=0.941.

6. Experimental Result: Machine learning is an AI application that mimics human intelligence's ability to learn from experience, particularly in pattern recognition. This is crucial in criminal justice, as AI aims to replicate human capabilities in software algorithms and hardware, such as identifying people, performing complex tasks, and making predictions. Accuracy and precision are crucial measures of observational error, indicating the accuracy and precision of a set of measurements and their proximity to a recognized value.

Accuracy: The accuracy of a binary classification test, also known as the "Rand index," is a statistical measure that compares pre- and post-test probability estimates.

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

where FN = False negative, TN = True negative, FP = False positive, and TP = True positive.

Based on 91 accurate predictions made out of 100 crime samples, the model has a 91% accuracy rate in identifying 100 crime samples as uncertainty. However, the samples only correctly identify 1 sample out of the 9 ones, resulting in 8 out of 9 certain going undiagnosed. This suggests that the model is not as effective as a model that always predicts benign. Accuracy alone is insufficient when dealing with a class-imbalanced data set with a significant difference between positive and negative labels.

$$\boldsymbol{Accuracy = \frac{1 + 90}{1 + 90 + 1 + 8} = 0.91}$$

Precision: Precision refers to the percentage of correctly classified occurrences or samples, as determined by a specific formula.

$$\boldsymbol{Precision = \frac{TP}{TP + FP}}$$

where FN = False negative, FP = False positive, and TP = True positive.

Out of the 160 crime samples in Dataset -1, 105 of the predictions made by an Uncertain voice text model are accurate, whereas the remaining 55 are not. Determine this model's precision value. True positives (TP) = 105 and false positives (FP) = 55 are the model's results.

Precision is calculated as follows:

$$\boldsymbol{Precision = \frac{105}{105 + 55} = 0.66}$$

As a result, the model's precision is 0.66.

7. Conclusion: Machine learning, an AI application, is crucial in criminal justice due to its ability to mimic human intelligence. The proposed study addresses the critical issue of data leaks in the context of increasing remote work and shifting work-life balances. The study focuses on cybercriminals' modus operandi such as phishing and one-on-one social engineering, to impersonate legitimate entities like emails, voice, and IT help desks/ support teams. The work demonstrates explainable artificial intelligence and the application of large language model to generate speech to text for electronic crime, showing strong predictive performance in handling imbalanced criminal dataset but, identifying limitations in the ROC AUC metric.

8. References:

[1] A.S. Rajawat, R. Rawat, R.N. Shaw and A. Ghosh: Cyber Physical System Fraud Analysis by Mobile Robot in Machine Learning for Robotics Applications. Studies in Computational Intelligence, Singapore: Springer, vol. 960, 2021.

[2] B. Rajput: Exploring the Phenomenon of Cyber Economic Crime in Cyber Economic Crime in India, Springer, pp. 53-78, 2020.

[3] Christopher Rigano: Using Artificial Intelligence to Address Criminal Justice Needs National Institute of Justice | NIJ.ojp.gov, NIJ Journal / Issue No. 280, January 2019.

[4] Fawcett, T: An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874,2006.

[5] Gunay Abdiyeva-Aliyeva, Jeyhun Aliyev, Ulfat Sadigov: Application of classification algorithms of Machine learning in cybersecurity, https://doi.org/10.1016/j.procs.2022.12.093. Procedia Computer Science, Volume 215, Pages 909-919,2022.

[6] J. Jang-Jaccard and S. Nepal: A survey of emerging threats in cybersecurity, Journal of Computer and System Sciences, vol. 80, no. 5, pp. 973-993, 2014.

[7] K. Veena, K. Meena, Ramya Kuppusamy, Yuvaraja Teekaraman, Ravi V. Angadi, Amruth Ramesh Thelkar: Cybercrime: Identification and Prediction Using Machine Learning Techniques, Compute Intelligence Neuroscience, 2022; 2022: 8237421, Published online 2022 Aug 27, Doi:10.1155 /2022 /8237,2022.

[8] Lisa Quest, Anthony Charrie, and Subas Roy: The Risks and Benefits of Using AI to Detect

Crime, Oliver Wyman Risk Journal Volume 8, Harvard Business Review on August 9, 2018.

[9] M. Kadoguchi, H. Kobayashi, S. Hayashi, A. Otsuka and M. Hashimoto: Deep Self-Supervised Clustering of the Dark Web for Cyber Threat Intelligence, 2020 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 1-6, November 2020.421,2020.

[10] Provost, F., & Fawcett, T: Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media, Inc,2013.

[11] R. Rawat, V. Mahor, S. Chirgaiya, R.N. Shaw and A. Ghosh: Sentiment Analysis at Online Social Network for Cyber-Malicious Post Reviews Using Machine Learning Techniques in Computationally Intelligent Systems and their Applications, Studies in Computational Intelligence, Singapore: Springer, vol. 950, 2021.

[12] R. Rawat, A.S. Rajawat, V. Mahor, R.N. Shaw and A. Ghosh: Surveillance Robot in Cyber Intelligence for Vulnerability Detection in Machine Learning for Robotics Applications, Studies in Computational Intelligence, Singapore: Springer, vol. 960, 2021.

[13] S. Kaur and S. Randhawa: Dark Web: A Web of Crimes, Wireless Personal Communications, vol. 112, no. 4, pp. 2131-2158, 2020.

[14] Vinod Mahor; Romil Rawat; Shrikant Telang; Bhagwati Garg; Debajyoti Mukhopadhyay; Prajyot Pali: Machine Learning based Detection of Cyber Crime Hub Analysis using Twitter Data, 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Date of Conference: 24-26 September 2021, Date Added to IEEE Xplore: 02 November 2021.