Machine Learning–Driven Classification of Text-Based Cybercrime under the Indian IT Act

Submitted: 25th September 2025; accepted: 23rd October 2025

Sukrati Agrawal, Hare Ram Sah, Rajesh Kumar Nagar

DOI: 10.14313/jamris-2026-020

Abstract:
Cybercrimes encompass crime against children, data breaches, and privacy violations. The increased frequency of cybercrimes due to the quick development of technology emphasizes the necessity of complex systems to analyze and categorize these offenses. There are many opportunities to analyze cybercrime data using Machine Learning (ML) techniques because of its enormous accumulation. This study proposes a model that has the potential to automatically analyze text-based reported cybercrime complaints based on the features by use of Random Forest (RF) and Gradient Boosting (GB) algorithms. This model includes a Bag of Words (BoW) approach for feature engineering to analyze reported cybercrime and suggest relevant Indian IT Act sections, such as Section 66E for privacy protection, Section 43A for reported data breach, and Section 72A for disclosure of information, using Natural Language Processing (NLP) for feature extraction and classification. This strategy enhanced the law and enforcement process by timely and accurately categorizing crime. By automating cyber law and providing timely legal answers to various reported cybercrimes, especially those concerning privacy and data protection, the model improves the capabilities of cybercrime units and achieves high accuracy and precision in anticipating pertinent legal sections.

Keywords: *Cybercrime, Machine Learning, NLP, Cyber Law, IT Act, ensemble stacking*

1. Introduction

1.1. Background

The rate of cybercrime has increased as a result of technology globalization, imparting serious risks to both personal safety and national security. In order to defend against these cyberattacks, modern technologies are essential. India stopped 500 million cyberattacks in the first quarter of the year 2023 [1].

As per the data accumulated from the National Crime Records Bureau (NCRB), there has been a persistent rise in reported cybercrime cases in recent years. The number of reported cybercrime cases and ongoing investigations is increasing yearly at an alarming rate, as depicted in the graph shown in Figure 1. This causes an increase in the backlog, which emphasizes how urgently the system must be automated to guarantee that the accused receive just punishment and that victims receive prompt remedies.

Figure 1. Cybercrime complaints reported in India during 2012 and 2022 [2]

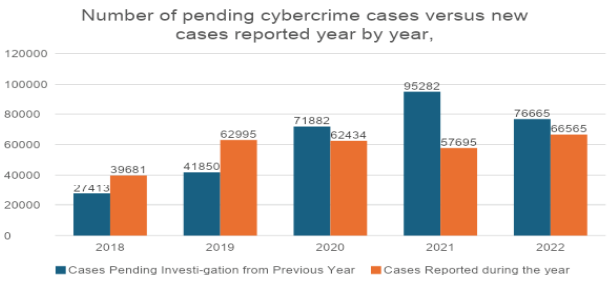

Figure 2. Cybercrime complaints reported and pending during the years 2018–2022 [2]

1.2. Problem Statement

In the present scenario, the manual processing and analysis of reported cybercrime cases delay legal responses, which may cause trauma to victims, as the process is laborious, time-consuming, and subject to human error, as illustrated in Figure 2. Machine learning techniques enable the rapid treatment and categorization of these cybercrime complaints under the applicable Indian IT Act sections. Machine learning also enhances the accuracy of these classifications.

The steady increase in reported cybercrime cases highlights the need for creative and artificial intelligence-based approaches to counter the expanding risks connected to the internet. The drawbacks of conventional case analysis and management techniques result in backlogs and delays in the execution of justice [3, 4].

The sharp increase in reported cybercrime complaints indicates that quick and automated technology is needed to effectively manage these complaints and

respond swiftly. An efficient solution to this problem is to combine the Information Technology (IT) Act with machine learning (ML)-based cybercrime complaint analysis systems to safeguard cyberspace and assist victims [5, 6].

As machine learning algorithms have the potential to analyze large amounts of text-based data and identify trends, they are required for categorizing cybercrime complaints and identifying relevant legal provisions of the Indian IT Act. This technology-driven architecture will enhance the cybercrime response system in terms of time as well as accuracy, ensure efficient classification and legal review, and enable a more effective administration of justice. [7–9].

1.3. Objectives

- To apply RF and GB models for categorizing cybercrime complaints under sections of the Indian IT Act for 66E, 43A, and 72A.
- To evaluate each model's performance in terms of accuracy, precision, recall, and F1-score.

Approach:

In the process of data collection, relevant cybercrime data is gathered from multiple sources like incident reports, legal documents, and cybercrime articles. These sources include a few samples of cybercrime complaints reported at police departments, legal institutions, and online resources. Once the data is gathered, preprocessing is applied to handle missing values and also to convert the data into the required format. Categorical variables are encoded, and outliers in numerical fields are managed to maintain data integrity. Feature engineering includes metrics like complaint duration and resolution status, and text data is cleaned and tokenized. The dataset is split into training, testing and validation sets (70%, 30%) to support thorough model evaluation. Model training utilizes Random Forest and Gradient Boosting algorithms [10, 11], with performance evaluated through precision, recall, F1 score, AUC, and accuracy metrics, ensuring a reliable classification model.

Outline:

The rest of the paper is outlined as follows: Section 2 provides a literature review, including an overview of the relevant provisions for cybercrime, focusing on Sections 66E, 43A, and 72A. It also presents a comprehensive review of legal frameworks for cybercrime and machine learning applications in text classification. Section 3 details the methodology, including data preparation, feature extraction, and the use of Random Forest and Gradient Boosting models. Section 4 presents the results, comparing model performance metrics and classification accuracy. Section 5 discusses the legal implications and limitations of automated classification in cybercrime complaints. It also outlines potential improvements and future research directions. Finally, this section concludes the paper by summarizing the key findings and their significance for enhancing cybercrime management under the IT Act.

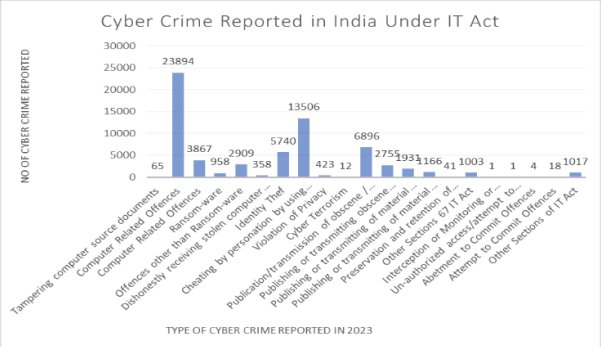

Figure 3. Types of cybercrime registered during the year 2023 under the Indian IT Act [12]

2. Review of the Literature

2.1. Legal Provisions for Cybercrime

The IT Act 2000 addresses various forms of cybercrimes to ensure data protection and privacy, as illustrated in Figure 3. It was amended in 2008 to add stronger provisions, notably Sections 66E, 43A, and 72A (Indian IT Act, 2008). Section 66E criminalizes the unauthorized capture or transmission of private images, focusing on safeguarding personal privacy. Section 43A mandates data controllers to adopt security practices, providing guidelines for reporting data breaches. Section 72A addresses the unauthorized disclosure of personal information by service providers, protecting user confidentiality. The effectiveness of these legal provisions has been widely acknowledged [5, 7].

Geetika Bhardwaj et al. [13] revealed a 46% rise in crimes against women in 2021 compared to 2020, emphasizing the need for proactive action to gather crime data and forecast future trends.

Olena V. et al. [14] highlighted the global increase in cyberattacks, particularly in finance, retail, technology, and communication industries, highlighting the challenges in combating such threats due to geographical disparities.

Gangwar Suraj et al. [15] report a surge in cybercrime due to the Internet of Things, cloud services, and improved connections. Cybercriminals threaten human lives. Targeting industrial control systems, elections, and digital wildfires rank among the top threats worldwide.

The study by Shuai Chenet et al. [16] indicates a positive global correlation between cybercrime and social, economic, and technological variables, with most incidents occurring in urbanized areas with greater infrastructure.

P. Datta et al. [17], in their study, show an increase in fraud cases, primarily affecting 20-29-year-olds, particularly mothers and children, necessitating awareness campaigns to combat cybercrime in India.

The study by Chudasama Dhaval et al. [18] suggests that third-party apps are frequently used by attackers for money transfers, leading to fraud.

Table 1 below summarizes the analysis of already existing algorithms used for cybercrime text detection and related applications.

Table 1. Comparison of existing algorithms

Author(s)	Dataset Used	Algorithm	Efficiency	IT / IPC Detected
Alami & Elbeqqali (2015) [19]	Microblog data	Text mining + SVM	Not detailed	No
Mbaziira & Jones (2016) [20]	Deceptive cybercrime text	Linguistics + ML	Medium	No
Kumari et al. (2018) [21]	Labeled text samples	NLTK, Scikit-learn	Moderate	No
Andleeb et al. (2019) [22]	MySpace bullying texts	Text mining + ML	Not detailed	No
Ch et al. (2020) [23]	State-wise crime stats	SVM, Decision Tree	Good	No
K. veena et al. (2022) [24]	Cybercrime reports	SVM	High	Potential
Pandey et al. (2022) [25]	Custom labeled reports	Ensemble (RF, NB, etc.)	High (noted)	No

Our literature review observed that most prior works focus on classification accuracy without explicitly mapping outputs to IT ACT provisions. This highlights a research gap where existing models are effective in detection but are not efficient in providing legally actionable outcomes, hence motivating the need for more advanced frameworks that provide integration of legal context with efficient machine learning.

2.2. Machine Learning in Legal Text Classification

An automated cybercrime text classification in the legal domain is gaining more attention. This automated cybercrime classification enhances the efficiency of handling a huge number of complaints. For the categorization of text-based reported cybercrime complaints, the most widely used ML models are RF and GB, as they have the potential to process huge datasets and provide strong performance [26]. ML algorithms are proven to be more successful in categorizing and analyzing text-based complaints, especially in cases when it comes to differentiating between several legal categories, as per recent studies [27–29].

ML is updating the classification of legal texts, especially those regarding cybercrime. In order to reduce human error and provide more efficient classification, Ch et al. [23] proposed a sustainable computational framework for classifying cybercrime offenses using ML. Andleeb et al. [22] showed how effective ML is at extracting features, analyzing cybercrime complaints, and classifying offenses through text mining. Patel and Sharma [30] highlighted the wider function of ML in automating legal procedures, especially in cyber law, demonstrating its efficiency in organizing and classifying legal documents. Pandey et al. [25] created a model that outperforms conventional methods by using ensemble learning to increase classification accuracy. Together, these studies highlight how important machine learning is to automate, analyze, and improve the classification of legal texts in cybercrime.

According to our literature review, there are various ways to approach countermeasures and prevent cybercrime. These include analyzing cyber threats, analyzing them with ML, and enhancing law enforcement tactics. The existing background of cybercrime will be examined, along with trends and patterns found in the literature. Proactive cybercrime detection and classification by machine learning is the main emphasis of this work. The research aims to create robust models that identify small irregularities and patterns indicative of cyber threats using various methods and datasets. Because of the increasing complexity of cybercriminals and the expansion of internet connectivity, cybercrime is a growing threat. Anonymity, exponential expansion in digital data and insufficient cybersecurity safeguards are some of the factors that make people and businesses vulnerable.

2.3. Law and enforcement to combat cybercrime

Recent research explores various strategies for combating cybercrime under the Indian IT Act. For instance, ensemble learning can be used to classify cybercrimes under Sections 66 and 67, and models like SVM and Random Forest can be used to improve accuracy, thus offering practical tools for cybercrime cells [26] to analyze cybercrime trends and prevention and providing global insights and statistical data on rising cybercrime patterns [31, 32]. Additionally, study of regional crime patterns using regression can find correlations in cyber offenses such as unauthorized photo sharing and computer theft [29]. Finally, the need to present a computational tool with high accuracy for identifying cybercrime rates at the state level in India underscores the role of machine learning in crime analytics [30].

3. Proposed Methodology

According to recent studies, machine learning techniques are becoming more and more necessary to identify which IT Act section applies to committed cybercrimes. Prior studies have focused on classifying cybercrime, but they have not addressed the crucial part of providing victim justice by determining whether the IT Act is applicable. This facilitates speedier investigations and raises the possibility that perpetrators of cybercrimes may be held accountable. Additionally, machine learning enables proactive enforcement measures to effectively combat cybercrime by keeping law enforcement agencies updated about evolving legal requirements and cyber threats, as presented in Figure 4.

3.1. Dataset

Compiling a Cyber Crime and Law Classification dataset involves gathering relevant and varied information about cybercrime incidents and legal provisions from multiple sources, including FIRs, case studies, news headlines, etc., as presented in Figure 5. Present research includes victim statements, incident reports, court records, and open databases that document cybercrime incidents.

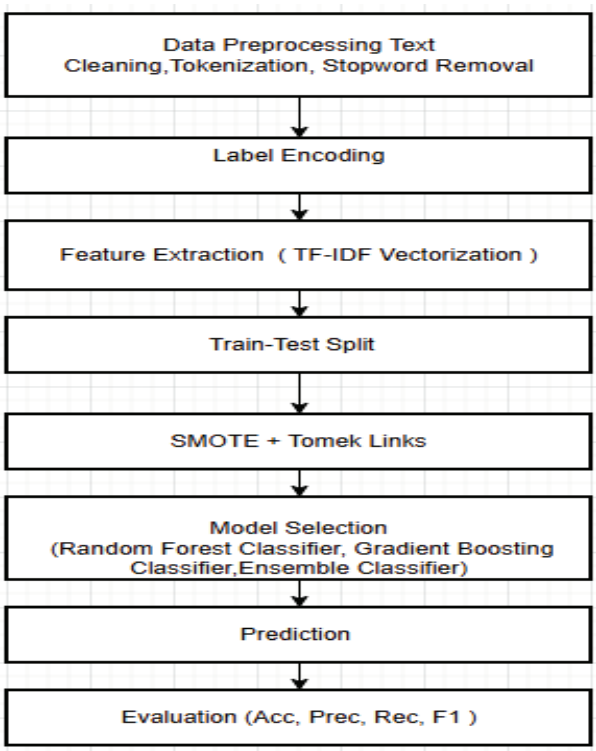

Figure 4. Proposed methodology

Figure 5. Distribution of dataset

3.2. Text Pre-Processing

Some preprocessing steps are applied to the Cyber Crime and Law Classification dataset to ensure the quality of the data. Missing values are addressed, numerical columns are cleaned and converted, and date columns are formatted as a datetime type. Outliers in the numerical data are handled, and categorical features are encoded by label encoder. In text data processing, normalization is done by converting all the text to lowercase. Noise is removed using regular expressions, and stop words are removed. The text is tokenized and lemmatized to ensure consistency.

$$x_i = \text{Clean}(\text{Tokenize}(t_i)) \quad (1)$$

Where:

- t_i = raw text
- x_i = processed token sequence

$$y_i = \text{LabelEncoder}(c_i), y_i \in \{0,1,\ldots,K-1\} \quad (2)$$

Figure 6. Word cloud for Section 43A

Where:

- c_i = category/section name
- y_i = encoded label

3.3. Feature Extraction

Feature extraction is used to change the unprocessed textual input so that the machine learning model can more efficiently use this input. The vectorization approach, TF-IDF (Term Frequency-Inverse Document Frequency), has been used for this investigation. TF-IDF weights words based on how frequently they occur within a certain piece of text compared with the frequency throughout the entire corpus. This will allow the model to discern words used often throughout the corpus versus those used specifically by individual publications. To accomplish this, we use the text data and translate this information into a sparse numerical features matrix with scikit-learn's TfidfVectorizer. We allow algorithms from machine learning to draw inferences from these inputted matrices in search of patterns and relations in the material that is text.

$$TF-IDF(t,d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}} \log\left(\frac{|D|}{1+\left|\{d \in D : t \in d\}\right|}\right) \quad (3)$$

Where:

- $f_{t,d}$ = frequency of term t in document d
- $\sum_{t'} f_{t',d}$ = total terms in document d
- $|D|$ = total number of documents
- $\{d \in D : t \in d\}$ = number of docs containing term t

Figure 6 provides a word cloud that helps in visual analysis of the words frequently occurring in Section 43A of the Indian IT ACT.

Figure 7 provides a word cloud that helps in visual analysis of the words frequently occurring in Section 66E of the Indian IT ACT.

Figure 8 provides a word cloud that helps in visual analysis of the words frequently occurring in Section 72A of the Indian IT ACT.

3.4. Handling Imbalanced Data with SMOTE

To solve the class imbalance in crime datasets, the study employs SMOTETomek, a strategy that combines SMOTE and Tomek Links. While Tomek Links

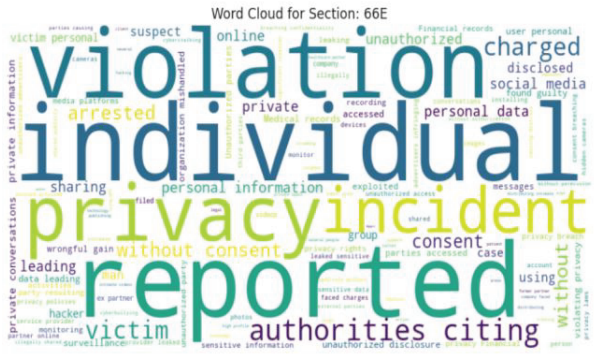

Figure 7. Word cloud for Section 66E

Figure 8. Word cloud for Section 72A

eliminates borderline occurrences, SMOTE creates synthetic examples for the minority class, guaranteeing a balanced class distribution and better predictive performance, particularly for uncommon crime categories.

$$x_{\text{new}} = x_i + \lambda(x_{nn} - x_i), \quad \lambda \in [0,1] \tag{4}$$

Where:
- x_i = a minority class sample.
- x_{nn} = one of the nearest neighbors of x_i
- λ = a random number.
- x_{new} = newly generated synthetic sample.

3.5. Model Selection

The three text categorization algorithms employed in this study are RF, GB, and Ensemble classifier (stacked), which combines RF and GB for increased accuracy and a soft-voting scheme. The models are evaluated using the following metrics as depicted in Figure 9, which shows that the Ensemble classifier performed the best.

This graphical representation shows that the IT Act's automated classification of cybercrime complaints may accelerate case processing, increasing accuracy and response times. Targeted legal measures are made possible by precise classification under certain provisions, which helps the court and law enforcement better combat cybercrime [32].

For RF:

$$\hat{y} = mode\left\{\widehat{y^{(1)}}(x), \widehat{y^{(2)}}(x), \ldots, \widehat{y^{(B)}}(x)\right\} \tag{5}$$

Where:

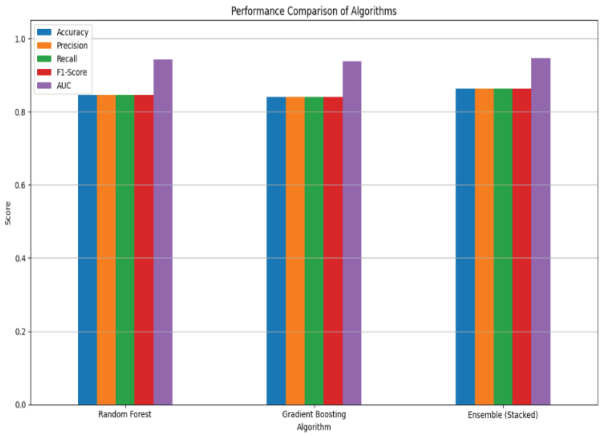

Figure 9. Performance analysis for different models

- $\hat{y}$ = final predicted class.
- $\widehat{y^{(i)}}(x)$ = prediction of the i^{th} base learner for input x.
- B = total number of base learners.
- mode $\{\cdot\}$ = class that appears most frequently among the predictions of all base learners.

For GB:

$$F_m(x) = F_{m-1}(x) + \nu h_m(x) \tag{6}$$

Where:
- $F_m(x)$ = boosted model after m iterations
- $F_{m-1}(x)$ = model from previous iteration (m−1)
- $h_m(x)$ = weak learner at step m
- ν = learning rate
- x = input feature vector

For ECS:

$$\hat{y} = \sigma(w_0 + w_1\widehat{y_{RF}} + w_2\widehat{y_{GB}}) \tag{7}$$

Where:
- $\hat{y}$ = final predicted output
- σ = activation function
- w_0 = bias
- w1i = weights
- y_{RF} = prediction from Random Forest
- $\widehat{y_{GB}}$ = prediction from Gradient Boosting

4. Results and Discussion

4.1. Performance Comparison

The performance of each algorithm is summarized in Table 2, showing the accuracy, precision, recall, F1-score, and AUC based on the training model.

Table 2. Performance comparison of models

Algo	Accuracy	Precision	Recall	F1-Score	AUC
RF	0.84	0.84	0.84	0.84	0.94
GB	0.84	0.83	0.84	0.85	0.93
ECS	0.86	0.86	0.86	0.86	0.94

Table 3. Section-based cybercrime classification result analysis

Section	Algo	Precision	Recall	F1-Score	Support
66E	RF	1.000	1.000	1.000	47
	GB	1.000	1.000	1.000	
	ECS	1.000	1.000	1.000	
72A	RF	0.700	0.903	0.789	31
	GB	0.667	0.903	0.767	
	ECS	0.700	0.903	0.789	
43A	RF	0.927	0.760	0.835	50
	GB	0.923	0.720	0.809	
	ECS	0.927	0.760	0.835	

As per the results shown in Table 2, the Ensemble Classifier performs better than Random Forest and Gradient Boosting in terms of accuracy, precision, recall, F1-score, and AUC.

4.2. Section-Wise Accuracy

A detailed breakdown of accuracy per section (crime category) is provided, showing how each model performs across different crime types. Table 3 illustrates the accuracy of Random Forest, Gradient Boosting, and Ensemble Classifier for each crime section.

In conclusion, the research shows that every algorithm performed exceptionally well, obtaining flawless scores in Section 66E. In Section 72A, RF and Ensemble voting maintained strong recall while marginally outperforming GB in precision and F1-score. Random Forest and Ensemble Voting outperformed Gradient Boosting in Section 43A, achieving better F1 scores. RF is the most dependable and time-efficient algorithm overall since it continuously shows a balance between excellent performance and efficiency throughout all parts.

5. Conclusion

This study demonstrated that ML techniques can be used effectively to classify cybercrime complaints based on textual descriptions. We create a strong feature extraction and preprocessing pipeline by resolving class imbalance with SMOTETomek and applying the TF-IDF vectorization approach. The ensemble classifier outperforms the other models in terms of overall performance across a number of evaluation metrics, demonstrating how well it handles challenging criminal classification tasks.

5.1. Limitations and Model Improvements

The model has been successfully demonstrated for the three sections of the Indian IT Act, including sections 66E, 72A, and 43E. In future models, they can be trained to identify any relevant section of the Indian IT Act based on the reported cybercrime, which requires a huge dataset creation for each section of the Indian IT Act. This model has demonstrated high accuracy, but this can lead to a few small misclassifications in cases where complaints have overlapped phrases between relevant sections. Deeper semantic output may be obtained by improving the models' efficiency by the use of advanced NLP approaches like named entity recognition (NER) and sentiment analysis [28, 32].

5.2. Future Work

In the future, this algorithm can be applied to predict any relevant section of the Indian IT Act. In order to better understand complex patterns of language found in cybercrime complaints, future research must examine advanced approaches such as ensemble algorithms and BERT. These methods could enhance the system's capacity to manage massive volumes of data and adhere to data protection laws when paired with federated learning for privacy-preserving training [26].

AUTHORS

Sukrati Agrawal* – Research Scholar, Department of Computer Science, SAGE University, Indore, Madhya Pradesh, India, +91-8982493161, Kailod Kartal, Indore Rau Bypass Road, Indore, Madhya Pradesh, 452020, e-mail: sukratiagrawalphd@gmail.com.

Hare Ram Sah – Department of Computer Science, SAGE University, Indore, Madhya Pradesh, India, +91-9630689967, Kailod Kartal, Indore Rau Bypass Road, Indore, Madhya Pradesh, 452020, e-mail: ramaayu@gmail.com.

Rajesh Kumar Nagar – Department of Electronics and Communication, IET, SAGE University, Indore, Madhya Pradesh, India, +91-9981850973, Kailod Kartal, Indore Rau Bypass Road, Indore, Madhya Pradesh, 452020, e-mail: errajesh973@gmail.com.

*Corresponding author

References

[1] Indusface Research Team, "The State of Application Security Report Q1 2023," Indusface, 2023. [Online]. Available: https://www.indusface.com

[2] T. Basuroy, "Number of cyber-crimes reported in India 2012–2021," *Statista*, 2022. [Online]. Available: https://www.statista.com

[3] S. Kharat, "Cyber crime – a threat to persons, property, government and societies," *Property, Government and Societies*, Mar. 2017.

[4] S. Vashisth, N. Singh, and S. Dudi, "An in-depth investigation into police operations and their impact on delivering speedy justice in criminal proceedings: Efficiency, challenges, and legal implications," *Library Progress International*, vol. 44, no. 3, 2024, pp. 24434–24446.

[5] A. Kovacs, "Cybersecurity and data protection regulation in India: An uneven patchwork," *CyberBRICS: Cybersecurity Regulations in the BRICS Countries*, 2021, pp. 133–181.

[6] H. Ning, X. Ye, M. A. Bouras, D. Wei, and M. Daneshmand, "General cyberspace: Cyberspace and cyber-enabled spaces," *IEEE Internet of Things Journal*, vol. 5, no. 3, 2018, pp. 1843–1856.

[7] D. M. Katz and J. J. Nay, "Machine learning and law," *Legal Informatics*, 2021, pp. 94–98.

[8] V. Altuglu and R. Schwabe, "Machine learning in litigation," *Handbook of Marketing Analytics*, Edward Elgar Publishing, 2018, pp. 661–670.

[9] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artificial Intelligence Review*, vol. 52, 2019, pp. 273–292.

[10] M. Noguti, E. Vellasques, and L. S. Oliveira, "A small claims court for the NLP: Judging legal text classification strategies with small datasets," *Proc. 2023 IEEE International Conference on Systems, Man, and Cybernetics*(SMC), Oct. 2023, pp. 1840–1845.

[11] P. Payne, R. V. Romould, M. K. Gourisaria, V. Singh, D. K. Behera, and S. Das, "Unveiling bankruptcy risk through advanced data analysis and machine learning techniques," *Proc. 2024 International Conference on Distributed Computing and Optimization Techniques* (ICDCOT), Mar. 2024, pp. 1–6.

[12] S. Agrawal and H. R. Sah, "Study on vulnerability in online social networking: Impact on an individual, community," *Online Social Networks in Business Frameworks*, 2024, pp. 27–46.

[13] G. Bhardwaj and R. K. Bawa, "Performance of machine learning models on crime data," *Proc. International Conference on Computing, Communications, and Cyber-Security*, Springer, Oct. 2022, pp. 811–823.

[14] O. V. Sviatun, O. V. Goncharuk, C. Roman, O. Kuzmenko, and I. V. Kozych, "Combating cybercrime: Economic and legal aspects," *WSEAS Transactions on Business and Economics*, vol. 18, 2021, pp. 751–762.

[15] S. Gangwar and V. Narang, "A survey on emerging cyber crimes and their impact worldwide," *Research Anthology on Combating Cyber-Aggression and Online Negativity*, IGI Global, 2022, pp. 1583–1595.

[16] S. Chen, M. Hao, F. Ding, D. Jiang, J. Dong, and S. Zhang, "Exploring the global geography of cybercrime and its driving forces," *Humanities and Social Sciences Communications*, vol. 10, no. 1, 2023, pp. 1–10.

[17] P. Datta, S. N. Panda, S. Tanwar, and R. K. Kaushal, "A technical review report on cyber crimes in India," *Proc. 2020 International Conference on Emerging Smart Computing and Informatics* (ESCI), IEEE, Mar. 2020, pp. 269–275.

[18] D. Chudasama, D. Patel, A. Shah, and N. Shaikh, "Research on cybercrime and its policing," *American Journal of Computer Science and Engineering Survey*, vol. 8, no. 10, 2020, pp. 14–20.

[19] S. Alami and O. Elbeqqali, "Cybercrime profiling: Text mining techniques to detect and predict criminal activities in microblog posts," *Proc. 2015 10th International Conference on Intelligent Systems: Theories and Applications* (SITA), Rabat, Morocco, IEEE, 2015, pp. 1–5.

[20] A. V. Mbaziira, E. Abozinadah, and J. H. Jones Jr., "Evaluating classifiers in detecting 419 scams in bilingual cybercriminal communities," *arXiv* preprint, 2015, arXiv:1508.04123.

[21] S. Kumari, Z. Saquib, and S. Pawar, "Machine learning approach for text classification in cybercrime," *Proc. 2018 Fourth International Conference on Computing Communication Control and Automation* (ICCUBEA), Pune, India, IEEE, 2018, pp. 1–6.

[22] S. Andleeb, R. Ahmed, Z. Ahmed, and M. Kanwal, "Identification and classification of cybercrimes using text mining technique," *Proc. 2019 International Conference on Frontiers of Information Technology* (FIT), IEEE, Dec. 2019, pp. 227–275.

[23] R. Ch, T. R. Gadekallu, M. H. Abidi, and A. Al-Ahmari, "Computational system to classify cybercrime offenses using machine learning," *Sustainability*, vol. 12, no. 10, 2020, pp. 4087–4098.

[24] K. Veena, K. Meena, R. Kuppusamy, Y. Teekaraman, R. V. Angadi, and A. R. Thelkar, "Cybercrime: Identification and prediction using machine learning techniques," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, 2022, p. 8237421.

[25] H. Pandey, R. Goyal, D. Virmani, and C. Gupta, "Ensem_SLDR: Classification of cybercrime using ensemble learning technique," *International Journal of Computer Network and Information Security*, vol. 14, no. 1, 2022, pp. 81–92.

[26] S. Verma and K. Joshi, "Automated text classification for legal applications," *Indian Law Review*, vol. 11, no. 1, 2023, pp. 75–95.

[27] N. Rao and P. Banerjee, "Data classification techniques in cybercrime analysis," *Journal of Information Security*, vol. 29, no. 4, 2022, pp. 410–430.

[28] S. Gupta, "AI in cyber crime management: A study," *Artificial Intelligence and Law*, vol. 28, no. 1, 2020, pp. 102–123.

[29] X. Luo, "Efficient English text classification using selected machine learning techniques," *Alexandria Engineering Journal*, vol. 60, no. 3, 2021, pp. 3401–3409.

[30] K. Patel and V. Sharma, "Machine learning applications in cyber law," *International Journal of Legal Studies*, vol. 8, no. 2, 2019, pp. 56–72.

[31] S. Batra, M. Gupta, J. Singh, D. Srivastava, and I. Aggarwal, "An empirical study of cybercrime and its preventions," *Proc. 2020 Sixth International Conference on Parallel, Distributed and Grid Computing* (PDGC), IEEE, Nov. 2020, pp. 42–46.

[32] S. Verma, S. Nayak, and D. K. Deshmukh, "A survey of emerging cyber crimes and their probable solutions," *Research Journal of Engineering and Technology*, vol. 11, no. 2, 2020, pp. 82–88.