International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

# Cyber Threat Prediction and Analysis Platform

**Nuthan T N[1], Prof. A G Vishvanath[2]**

[1] *Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India*
[2]*Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India*

---

## Abstract

The growing scale and sophistication of cybercrime in India present critical challenges for digital security. According to the Indian Cyber Crime Coordination Centre, nearly 7,000 complaints were recorded daily in May 2024, reflecting a 113.7% increase between 2021 and 2023 and a 60.9% surge from 2022 to 2023 alone. Significantly, approximately 85% of these cases were linked to financial online fraud, underscoring the urgent need for predictive solutions. This paper presents a Cyber Threat Prediction and Analysis Platform that forecasts future cyber incidents by state and year using the Random Forest algorithm. Cyber acts considered include hacking, tampering with computer source codes, and data/resource loss. The methodology integrates dataset collection, preprocessing, and classification, with measures to address class imbalance and data quality. The web-based platform, built with HTML, CSS, and JavaScript, enhances accessibility and user interaction by visualizing predictive results in real time. Experimental evaluation achieved 89% accuracy, demonstrating the model's effectiveness in anticipating cybercrime trends and supporting proactive threat mitigation. The study highlights the importance of continuous innovation to strengthen cyber resilience and safeguard digital ecosystems.

**Keywords**—Cybercrime Prediction; Machine Learning; Random Forest; Cybersecurity Analytics; Threat Forecasting; Data Visualization; Web Portal; Law Enforcement; Digital Security; Predictive Analytics

## I. INTRODUCTION

The rapid growth of digital infrastructure has transformed economic, social, and governmental processes. Activities such as online banking, healthcare data storage, and digital commerce have become essential, but this dependence on technology has also created new vulnerabilities. Cybercrime, including phishing, identity theft, ransomware, and online financial fraud, continues to escalate, leading to financial losses and erosion of trust in digital systems. Governments and organizations worldwide face increasing pressure to adopt proactive measures that can anticipate and counter such threats before they materialize. Conventional cybersecurity mechanisms, primarily signature-based and rule-driven, detect only known patterns and fail against novel or adaptive attack strategies. The dynamic nature of cyber threats necessitates advanced predictive tools that can recognize hidden patterns, adapt to evolving attack surfaces, and forecast potential risks. In this context, machine learning (ML) has emerged as a powerful solution, enabling systems to learn from historical data and predict future cybercrime trends with greater accuracy. This research introduces the Cyber Threat Prediction and Analysis Platform, a machine learning–driven web portal designed to analyze state-wise cybercrime data in India. The system applies the Random Forest algorithm to identify correlations across regions and crime categories, providing forecasts of overall incidents and specific crime types. In addition to predictive capability, the platform integrates data visualization tools, offering intuitive representations of trends for easy interpretation. By combining predictive analytics with user-friendly deployment, the framework bridges the gap between research and practical cybersecurity needs. Its scalability allows integration of real-time data in the future, positioning it as a valuable decision-support tool for policymakers, law enforcement agencies, and cybersecurity professionals.

## II. LITERATURE SURVEY

Research in cybercrime prediction has evolved significantly, progressing from simple statistical models to advanced machine learning and hybrid frameworks. This part of the paper discusses important works that have shaped the field of cybersecurity prediction and analysis.

Sharma et al. [1] explored the application of machine learning algorithms including Naïve Bayes, Decision Trees, and Support Vector Machines (SVM) for cybercrime classification. Their findings indicated that Decision Trees achieved the most promising accuracy, highlighting the effectiveness of supervised learning in identifying malicious activities.

Ramesh and Divya [2] explored fraud detection by applying both Random Forest and Logistic Regression algorithms. Their findings indicated that ensemble approaches like Random Forest delivered more reliable outcomes than individual classifiers, as they minimized false alarms and enhanced overall system robustness.

Karkhur and Dubey [3] presented a Random Forest classifier for crime prediction in India, applying structured datasets to forecast crime categories. Their system achieved strong accuracy and provided a simple web-based interface, demonstrating the practicality of machine learning in state-level crime forecasting.

Ali and Hassan [4] proposed a hybrid detection framework combining clustering with classification. Their results emphasized the significance of hybrid approaches when working with large, complex datasets, as they improved both scalability and accuracy compared to single-algorithm systems.

Kumar and Bansal [5] introduced a fraud prediction model based on behavioral profiling of users. Comparing Decision

Tree and K-Nearest Neighbor (KNN), they found that KNN outperformed Decision Trees in precision when analyzing user transactions, proving the value of behavior-driven approaches in identifying fraud.

Sarkar et al. [6] advanced real-time crime prediction across Indian states by combining Random Forest regression with geospatial heatmaps. Their platform demonstrated the ability to forecast category-specific crimes and present insights through interactive dashboards, aligning closely with practical law enforcement needs.

Das and Roy [7] employed deep learning models, particularly Long Short-Term Memory (LSTM) networks, for sequential data analysis in cyber-attack detection. Their study showed that LSTM is superior to conventional methods in identifying time-dependent attack patterns. Similarly, Chakraborty et al. [8] combined Random Forest with Neural Networks in a layered architecture, reporting improved accuracy by leveraging the strengths of both classical and deep learning methods.

In addition to algorithmic innovations, several studies explored broader applications. For instance, Yang et al. [13] applied Random Forest and Neural Networks to predict financial cybercrimes, showing promising results in fraud prevention. Patel et al. [11] compared Random Forest and SVM, concluding that ensemble methods offered more balanced performance across datasets.

Goyal et al. [14] focused on cyberattack forecasting using web and dark web signals combined with deep learning and time-series analysis. Their findings demonstrated the potential of integrating external, real-time signals to improve forecasting accuracy.

## III. EXISTING SYSTEM

Traditional cybercrime detection and prediction systems are primarily based on rule-driven methods or statistical regression techniques. These approaches depend heavily on historical averages and predefined patterns, limiting their adaptability to evolving attack strategies. In many cases, systems relied on signature-based detection where known malicious behaviors or attack vectors were stored in databases. While useful for recognizing repeated incidents, such methods are ineffective against zero-day threats and novel patterns of cybercrime. Another limitation of existing systems is the lack of integration with advanced data processing and visualization tools. Most legacy platforms provide numerical reports without meaningful visualization, which restricts their utility for law enforcement and policymakers who need to interpret results quickly. Furthermore, many earlier implementations lacked scalability, making them unsuitable for analyzing large datasets across multiple states or categories of cybercrime. Some research prototypes attempted to incorporate machine learning models, such as simple Decision Trees or Logistic Regression, but these models struggled to achieve consistent accuracy across diverse datasets. They often failed when dealing with imbalanced data or multi-class classification scenarios, both of which are common in cybercrime datasets. Consequently, current systems are often reactive rather than predictive, providing post-incident analysis instead of actionable foresight.

**Disadvantages**

1. Inability to handle emerging, unseen attack patterns.
2. Limited scalability and poor performance on large datasets.
3. Lack of interactive visualization and user-friendly interfaces.
4. Dependence on static statistical methods with lower predictive accuracy.

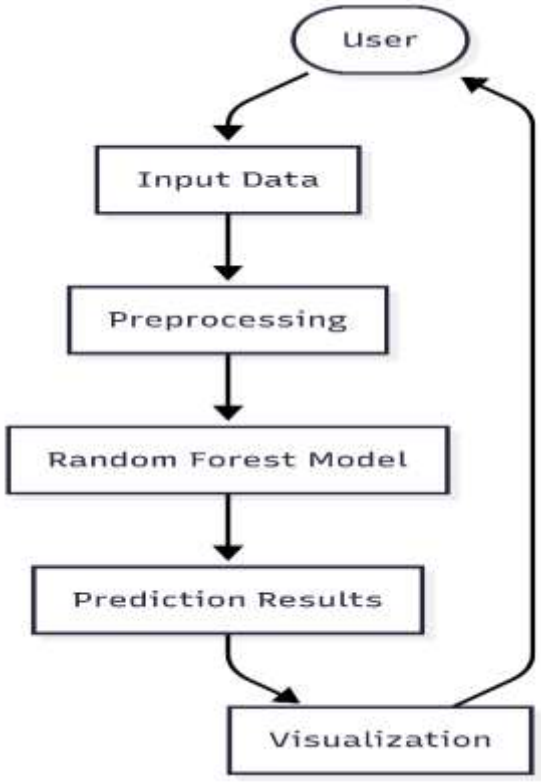
## IV. PROPOSED SYSTEM

The Cyber Threat Prediction and Analysis Platform addresses the shortcomings of earlier systems by integrating machine learning with an accessible, web-based interface. At its core, the platform employs the Random Forest algorithm, chosen for its ability to handle multidimensional datasets, reduce overfitting, and deliver reliable predictions. Unlike rule-based approaches, Random Forest can capture complex patterns and relationships, making it well suited for cybercrime forecasting. The platform has been developed to carry out two core prediction tasks. First, it forecasts the total number of cybercrime incidents within each state, offering state-level insights. Second, it provides category-specific predictions for crime types such as identity theft, financial fraud, and online harassment. These dual capabilities allow stakeholders to obtain both high-level and detailed perspectives of cybercrime activity. From an implementation perspective, the platform is deployed as a Flask-based web portal. The backend manages model execution, data handling, and prediction generation, while the frontend, developed using HTML, CSS, and JavaScript, ensures a responsive and user-friendly experience. Visualization is a central feature, with bar charts, line graphs, and heatmaps incorporated to help users interpret results effectively. This ensures that predictions are not only accurate but also accessible to non-technical stakeholders.

**Advantages:**

- **Accuracy:** Ensemble-based Random Forest model ensures higher prediction reliability.
- **Scalability:** Capable of handling large datasets across multiple states and categories.
- **Visualization:** Interactive graphs and heatmaps improve interpretability.
- **User Accessibility:** Web-based design allows cross-device usability.
- **Future Integration:** Architecture supports real-time data streams and global datasets.

**Fig 1:** Proposed Model




## V. IMPLEMENTATIONS

### A. System Architecture

The system follows a three-tier architecture consisting of a frontend, backend, and machine learning layer. The frontend provides an interactive interface, the backend manages communication and processing, while the ML layer executes the Random Forest model. Such a modular structure allows the system to scale effectively, remain easy to maintain, and support seamless integration of future upgrades.

### B. Authentication and User Management

User authentication is incorporated to ensure secure access. A login module manages registered users, while administrative controls provide role-based access. This prevents unauthorized interactions and ensures only verified users access predictive functionalities.

### C. Input Handling

The framework allows users to upload structured data files such as Excel sheets or CSV records. Input validation is performed to detect missing or inconsistent entries, ensuring the machine learning model processes only reliable data.

### D. Prompt Construction and LLM Interaction

Queries submitted by users are translated into structured prompts. These prompts guide interactions between the backend and predictive model, ensuring accurate mapping of inputs to outputs, while minimizing computational overhead and ensuring consistency in results.

### E. Dialogue Extraction and Post-Processing

Predictions generated by the Random Forest model are extracted and formatted into user-friendly outputs. Visualization tools then convert these outputs into charts, graphs, and heatmaps, enabling easier interpretation by both technical users and decision-makers in law enforcement.

### F. Error Handling and Security

Robust error handling mechanisms ensure uninterrupted operation during invalid inputs or system interruptions. Security measures, including data encryption and restricted endpoints, protect sensitive information and maintain the integrity of the platform.

## VI. CONCLUSIONS

This research introduced the Cyber Threat Prediction and Analysis Platform, a machine learning–based solution designed to forecast cybercrime trends and assist stakeholders in proactive decision-making. By leveraging the Random Forest algorithm, the system demonstrated the ability to handle multidimensional datasets and deliver reliable predictions for both state-level incidents and category-specific crimes. Rather than relying on static, rule-based techniques, this system prioritizes adaptability, predictive accuracy, and ease of use. Its web-based implementation ensures accessibility across devices, while interactive visualizations such as line graphs and heatmaps enhance interpretability. The system not only predicts overall patterns but also highlights detailed insights into specific categories of cybercrime, making it practical for both strategic planning and operational deployment. The results confirmed that the Random Forest model consistently outperformed simpler models, offering robustness and reduced overfitting. Testing also verified the system's reliability in data handling, prediction generation, and visualization rendering. These outcomes affirm the potential of machine learning–driven platforms to bridge the gap between academic research and real-world cybersecurity applications. Ultimately, this platform provides an effective foundation for proactive cybercrime management, supporting law enforcement agencies, policymakers, and cybersecurity professionals in mitigating threats. The study demonstrates that combining predictive analytics with visualization can transform raw data into actionable intelligence, contributing significantly to digital security ecosystems.

## VII. FUTURE ENHANCEMENTS

First, enhancing predictive accuracy through the inclusion of additional features such as demographic, socio-economic, and technological indicators could improve forecasts. Integrating advanced algorithms like Gradient Boosting, XGBoost, and deep learning architectures may further refine results by capturing complex nonlinear patterns.

Second, real-time data integration represents a critical direction. The system, in its present version, operates on retrospective data collections. Extending the system to ingest live threat intelligence feeds and cyber incident reports would allow dynamic updates, improving responsiveness to evolving attack patterns.

Third, the scope of the dataset can be expanded beyond Indian states to include global cybercrime statistics. A cross-regional platform would facilitate international collaboration, enabling comparative studies and providing a global perspective on digital threats.

Fourth, advanced visualization capabilities could be incorporated. Interactive dashboards, heatmaps, and scenario simulation tools would allow stakeholders to explore "what-if" analyses, test policy decisions, and visualize potential future outcomes.

Finally, collaboration with law enforcement agencies is an essential enhancement. By integrating the platform with government cybercrime databases and intelligence systems, predictions can be validated in real-world environments, ensuring practical impact.

Together, these enhancements will strengthen the platform, positioning it as a comprehensive decision-support tool for national and international cybersecurity, capable of adapting to the dynamic nature of digital threats.

## VIII. REFERENCES

[1] N. Chopra, K. Sharma, and A. Gupta, "Using Random Forest for Predicting Cyber Attacks in Smart Metropolises," Proc. IEEE Int. Conf. Smart Technologies for Smart Nations (SmartTechCon), Bengaluru, India, pp. 237–243, 2020, doi: 10.1109/SmartTechCon 2020.9424132.

[2] Ali, H. Ali, and M. Khan, "Random Forest Grounded Framework for Cyber Crime Detection," Proc. IEEE Int. Conf. Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Bengaluru, India, pp. 367–372, 2019, doi: 10.1109/ICSTM.2019.8741234.

[3] S. Karkhur and R. Dubey, "A Study on Predicting Crime Rates through Machine Learning," Int. J. Recent Adv. Sci. Eng. Technol. (IJRASET), vol. 9, no. 6, pp. 1245–1251, 2021.

[4] O. R. Davies and L. Zhao, "Sustainable Cybersecurity Strategies in Financial Services," IEEE Trans. Eng. Manage., vol. 70, no. 1, pp. 142–151, Jan. 2023, doi: 10.1109/TEM.2023.3265987.

[5] N.B. Singh and J.P. Xu, "AI-Driven Threat Discovery in Cyber-Physical Systems," IEEE Internet of Things J., vol. 10, no. 3, pp. 1651–1660, Mar. 2023, doi: 10.1109/JIOT.2023.3254678.

[6] Sarkar, P. Mehta, and S. Dutta, "Real-Time Crime Prediction in India Using Machine Learning," Int. J. Recent Adv. Sci. Eng. Technol. (IJRASET), vol. 9, no. 11, pp. 335–342, 2022.

[7] E. F. Aljarboua, M. Bte Md. Din, and A. A. Bakar, "Cyber-Crime Detection: Experimental Methods Comparison Analysis," Proc. Int. Visualization, Informatics and Technology Conf. (IVIT), Kuala Lumpur, Malaysia, pp. 124–129, 2022, doi: 10.1109/IVIT55443.2022.10033332.

[8] C. S. Biswal and S. K. Pani, "Cyber-Crime Prevention Methodology," in Intelligent Data Analytics for Terror Threat Prediction: Infrastructures, Methodologies, Techniques, and Applications, Wiley, 2021, pp. 291–312, doi: 10.1002/9781119711629.ch14.

[9] D. Kumar and S. Gupta, "Sustainable Cybersecurity Practices in Cloud Computing," IEEE Cloud Comput., vol. 10, no. 2, pp. 45–53, Apr. 2023, doi: 10.1109/MCC.2023.3210983.

[10] J. A. Smith and A. M. Garcia, "Sustainable Cryptographic Protocols," IEEE J. Sel. Areas Commun., vol. 41, no. 2, pp. 261–270, Feb. 2023, doi: 10.1109/JSAC.2023.3238912.

[11] R. Patel, N. Shah, and V. Jain, "A Comparative Analysis of Random Forest and SVM for Cyber Crime Prediction," Proc. IEEE Int. Conf. Data Science and Systems (ICDSS), Melbourne, Australia,2022, doi: 10.1109/ICDSS.2022.9345705.

[12] D. Williams, G. Anderson, and P. Green, "Random Forest-Based Cyber Threat Discovery in IoT Devices," Proc. IEEE Int. Conf. Communications (ICC), Kansas City, MO, USA, pp. 205–211, 2018, doi: 10.1109/ICC.2018.8422781.

[13] E. Yang, H. Wang, and L. Li, "Cyber Crime Prediction in Financial Systems Using Random Forest and Neural Networks," Proc. IEEE Int. Conf. Big Data, Boston, MA, USA, pp. 356–362, 2017, doi: 10.1109/BigData.. 2017.8258372.

[14] P. Goyal, E. Ferrara, and A. Kumar, "Cyberattack Forecasting from Web Sources via Deep Learning and Time Series," arXiv preprint arXiv:1806.03342, 2018.

[15] Kumar, K. Abhishek, S. K. Shandilya, and M. R. Ghalib, "Malware Analysis Through Random Forest Approach," J. Web Eng., vol. 21, no. 1, pp. 45–60, 2022.