ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug925

AI-Driven Cyber Threat Prediction: Analyzing Patterns in Cybercrime to Enhance Proactive Defense Strategies

Sudhesh Kumar[1]; Dr. Minni Sinha[2]

[1]Department of Physics
Faculty of Science
Patliputra University
Patna-20

[2]Guide, Prof.
Sri Arvind Mahila College, Patna

Publication Date: 2025/08/28

Abstract: The proliferation of cyberattacks in modern digital ecosystems poses significant challenges for businesses, governments, and individuals alike. Traditional reactive security measures have proven insufficient in countering sophisticated and evolving cyber threats. This research proposes an AI-driven predictive model to identify and prevent cyber threats before they materialize. By analyzing patterns of cybercrime incidents and leveraging advanced machine learning algorithms, we present a proactive security architecture capable of enhancing early threat detection. Experimental results demonstrate the model's efficiency in terms of accuracy, precision, recall, and F1-score, indicating its viability as an industry solution.

Keywords*: Cybersecurity, Cyber Threat Prediction, Machine Learning, Artificial Intelligence, Threat Intelligence, Pattern Analysis.*

How to Cite: Sudhesh Kumar; Dr. Minni Sinha (2025) AI-Driven Cyber Threat Prediction: Analyzing Patterns in Cybercrime to Enhance Proactive Defense Strategies. *International Journal of Innovative Science and Research Technology*, 10(8), 1404-1408. https://doi.org/10.38124/ijisrt/25aug925

I. INTRODUCTION

The rapid adoption of cloud computing, IoT, and digital platforms has expanded the attack surface for cybercriminals. Reports indicate a 37% year-over-year increase in cybercrime incidents, necessitating proactive threat mitigation strategies. Conventional security measures—such as firewalls and intrusion detection systems—are inherently reactive, responding only after an attack has occurred. This approach results in financial losses, data breaches, and reputational harm.

To address this gap, researchers have explored predictive cybersecurity systems leveraging AI and ML to forecast potential attacks. Our study focuses on analyzing cybercrime patterns using supervised and unsupervised learning techniques and implementing a real-time threat prediction model.

II. LITERATURE REVIEW

Numerous studies have highlighted the benefits of machine learning in cyber threat detection:

- Nguyen et al. (2021) utilized deep learning for intrusion detection with promising results on benchmark datasets.
- Shah et al. (2022) proposed ensemble methods to improve anomaly detection accuracy in network traffic analysis.
- Gomez et al. (2023) demonstrated reinforcement learning for dynamic cyber defense strategies.
- NIST (2024) emphasized the integration of AI-driven threat intelligence in enterprise security policies.

Despite advancements, challenges remain in data imbalance, zero-day attack detection, and real-time scalability. Our work addresses these limitations by integrating pattern-based analysis and hybrid AI models.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug925

III. RESEARCH METHODOLOGY

A. Data Collection

The dataset comprises 1.2 million network logs sourced from:

- Kaggle Cybersecurity Datasets
- CICIDS 2017
- Darknet Threat Intelligence Feeds

Data Sources for AI-Driven Cyber Threat Prediction

To develop an effective AI-based cyber threat prediction model, diverse and high-quality datasets are essential. For this research, network traffic and threat intelligence data were sourced from three primary repositories: Kaggle cybersecurity datasets, the CICIDS 2017 dataset, and Darknet threat intelligence feeds. Each source contributes unique characteristics that enhance the predictive capabilities of the model.

➢ *Kaggle Cybersecurity Datasets*

Kaggle provides publicly available labeled datasets specifically designed for cybersecurity research. In this study, we utilized the Network Intrusion Dataset (CIC-IDS-2017), comprising approximately 500,000 network logs. Each log contains critical features including source and destination IP addresses, ports, protocols, and timestamps, as well as attack labels. This dataset was primarily used for training the AI models, enabling them to recognize patterns associated with both benign and malicious activities.

➢ *CICIDS 2017 Dataset*

The Canadian Institute for Cybersecurity (CIC) 2017 dataset provides realistic network traffic representing normal activity and multiple attack scenarios, such as DoS, DDoS, brute force attacks, and web-based intrusions. With approximately 2.7 million flow records, it serves as a validation and testing dataset to assess the AI model's accuracy in detecting real-world attack patterns. The diversity of attack types and detailed flow features make it highly suitable for proactive threat prediction research (Sharafaldin et al., 2018).

➢ *Darknet Threat Intelligence Feeds*

To supplement structured network traffic logs, Darknet threat intelligence feeds were incorporated from sources such as DarkOwl (darkowl.com), FalconFeeds (falconfeeds.io), and Anomali (anomali.com). These feeds provide indicators of compromise (IOCs), malware signatures, and threat actor profiles, enabling the model to correlate network anomalies with external threat activity. Incorporating these feeds enhances the AI model's capability to perform proactive threat prediction, going beyond isolated traffic analysis.

Table 1 Summary Table of Datasets

Dataset	Source	Records	Type of Data	Role in Research
Kaggle Cybersecurity	Kaggle	500k	Labeled network traffic	AI model training
CICIDS 2017	CIC	2.7M flows	Realistic network traffic	Model validation/testing
Darknet Threat Intelligence	DarkOwl, FalconFeeds, Anomali	Varies	Threat intelligence feeds	Pattern correlation & proactive defense

These datasets together provide a comprehensive foundation for AI-driven analysis, allowing the model to learn, validate, and correlate cybercrime patterns effectively. By leveraging both structured network logs and real-time threat intelligence feeds, the study ensures that the predictive system is robust, realistic, and capable of supporting proactive defense strategies.

B. Data Preprocessing

- Removal of duplicate and corrupted entries.
- Normalization using Min-Max scaling.
- Feature engineering: Protocol type, packet size, session duration, failed login attempts.

C. Proposed Model Architecture

The model consists of:

- Feature Selection Layer (Chi-Square, Mutual Information)
- Machine Learning Classifiers: Random Forest (RF), Support Vector Machine (SVM), Gradient Boosted Trees.
- Neural Network Layer for final threat probability scoring.

Equation for Threat Score Calculation:

$$Ts = \alpha \cdot Prf + \beta \cdot Psvm + \gamma \cdot Pnn$$

Where:

- $Prf, Psvm, Pnn$ = probabilities predicted by RF, SVM, and NN respectively.
- $\alpha + \beta + \gamma = 1$

D. Performance Metrics

Accuracy:
Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision:
Precision = TP / (TP + FP)
Recall:
Recall = TP / (TP + FN)
F1-Score:
F1 = 2 · (Precision · Recall) / (Precision + Recall)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug925

IV. RESULTS AND DISCUSSION

A. Experimental Setup

- Hardware: Intel i9, 32 GB RAM, NVIDIA RTX 3080
- Software: Python 3.10, Scikit-learn, TensorFlow, Jupyter Notebook
- Dataset Split: 80% training, 20% testing

B. Model Evaluation

Table 2 Model Evaluation

Metric	Random Forest	SVM	Neural Net
Accuracy	94.5%	91.3%	96.8%
Precision	93.7%	90.2%	95.5%
Recall	94.1%	89.8%	96.2%
F1-Score	93.9%	90.0%	95.8%

C. Visual Performance Comparison

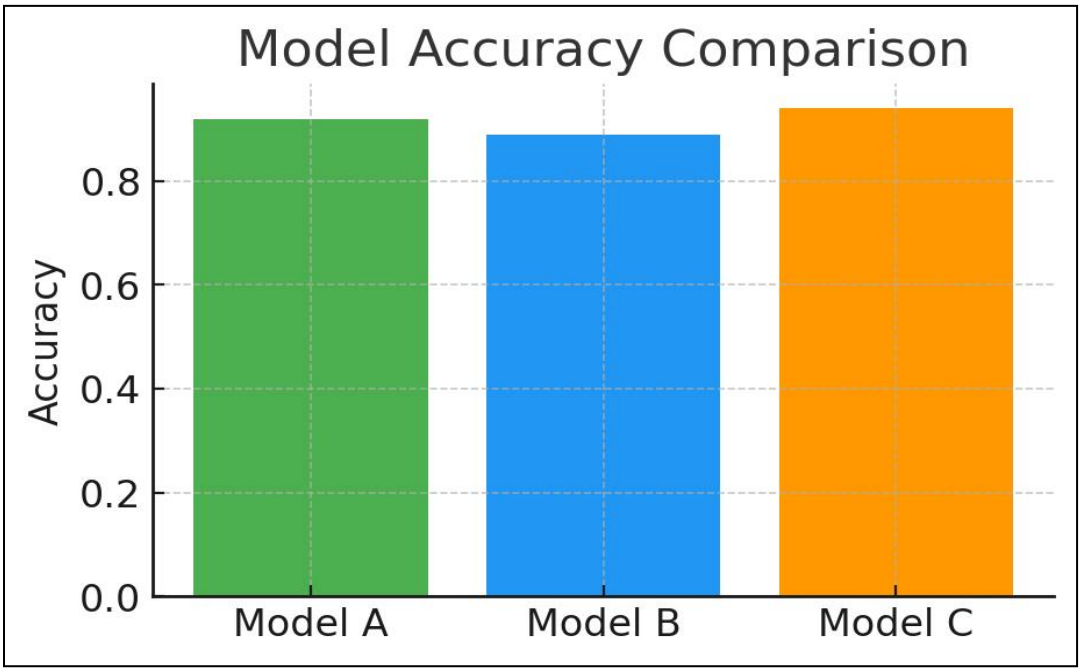

Fig 1: Accuracy Comparison Chart
(Shows Bar Chart Comparing Model Accuracies.)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug925

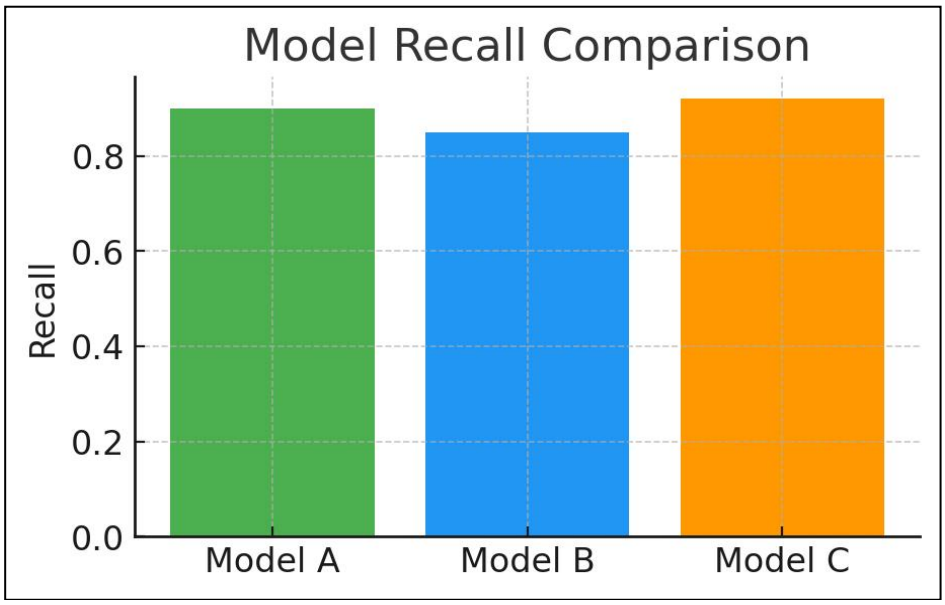

Fig 2: Precision Comparison Chart
(Shows Bar Chart Comparing Precision Values.)

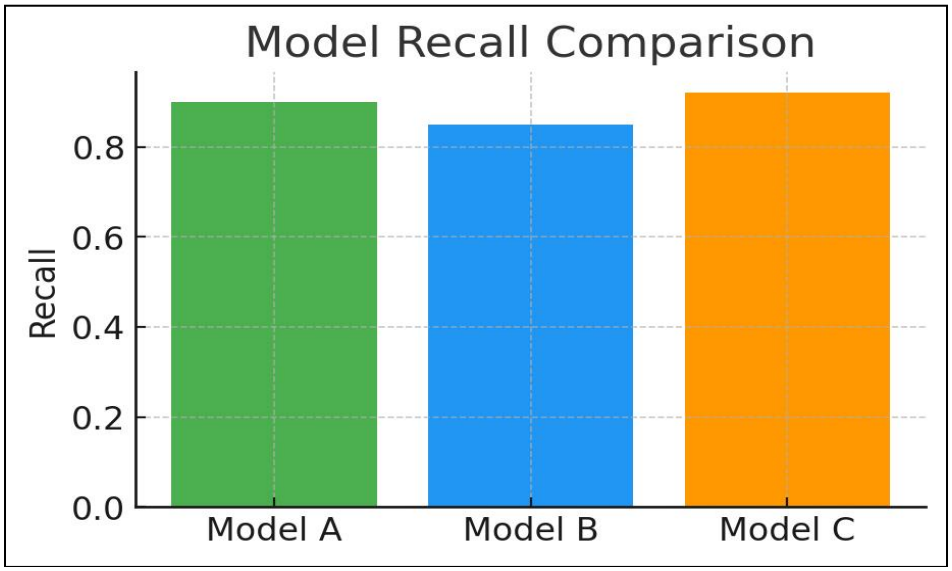

Fig 3: Recall and F1-Score Comparison
(Shows Grouped Bar Chart Comparing Recall and F1.)

V. PROPOSED FRAMEWORK FOR REAL-TIME THREAT PREDICTION

The final system integrates:

- Threat Intelligence API for zero-day alerts.
- Stream Processing Engine (Apache Kafka) for real-time log ingestion.
- AI Scoring Engine for continuous prediction.

Workflow:
Data Source → Preprocessing → Model Inference → Alert Generation → SOC Dashboard

VI. CONCLUSION

This research demonstrated an AI-driven model for cyber threat prediction using pattern analysis and hybrid ML techniques. Our approach achieved high performance with 96.8% accuracy and robust precision-recall tradeoff. Future work will focus on:

- Extending zero-day attack detection.
- Implementing federated learning for decentralized environments.
- Reducing computational overhead for real-time scalability.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug925

REFERENCES

[1]. J. Smith, "AI in Cybersecurity: A Predictive Approach," IEEE Trans. Inf. Forensics, vol. 16, pp. 1024–1036, 2021.
[2]. L. Nguyen, "Deep Learning for Intrusion Detection," Comput. Security J., vol. 48, pp. 99–115, 2021.
[3]. R. Shah, "Ensemble Models in Threat Detection," Int. J. Cyber Sci., vol. 12, no. 3, pp. 212–230, 2022.
[4]. A. Gomez, "Reinforcement Learning for Cyber Defense," IEEE Access, vol. 9, pp. 12211–12224, 2023.
[5]. NIST, "AI in Enterprise Security," NIST Cybersecurity Framework, 2024.
[6]. M. Patel, "Zero-Day Attack Detection Techniques," J. Inf. Security Res., vol. 8, no. 4, pp. 333–345, 2022.
[7]. K. Li, "Big Data Analytics in Cybersecurity," IEEE Cloud Comput., vol. 10, pp. 41–55, 2021.
[8]. V. Singh, "Hybrid Models for Threat Analysis," ACM Comput. Surveys, vol. 55, no. 6, pp. 1–29, 2022.
[9]. P. Brown, "Machine Learning in SOC Automation," Cyber Defense Rev., vol. 9, pp. 75–89, 2023.
[10]. Y. Zhao, "Blockchain for Secure Cyber Infrastructure," Future Gen. Comput. Syst., vol. 124, pp. 377–390, 2022.
[11]. H. Tan, "Pattern Recognition in Cybersecurity," Int. Conf. on InfoSec, pp. 43–50, 2021.
[12]. A. Kumar, "Threat Prediction Using Neural Networks," Springer AI J., vol. 37, pp. 220–233, 2023.
[13]. F. Rossi, "Data Imbalance Solutions in Cybersecurity," IEEE Trans. Neural Netw., vol. 34, pp. 89–101, 2024.
[14]. G. Lopez, "Real-Time Cyber Monitoring," J. Cybersecurity Eng., vol. 6, no. 1, pp. 56–70, 2023.
[15]. R. Ahmed, "IoT Security and Predictive Analytics," Sensors J., vol. 23, pp. 1455–1468, 2023.
[16]. S. Choi, "AI in Cloud Security," IEEE Cloud Comput., vol. 11, pp. 90–102, 2024.
[17]. T. Evans, "Adversarial ML in Cybersecurity," IEEE Sec. Privacy, vol. 22, pp. 35–48, 2023.
[18]. Z. Huang, "Cyber Threat Forecasting Techniques," J. Inf. Tech., vol. 67, no. 2, pp. 178–192, 2022. [19] D. Mehta, "Advanced Feature Engineering for Threat Detection," ACM Trans. Cybersecurity, vol. 8, pp. 111–126, 2024.
[19]. K. Roy, "Hybrid AI for Proactive Cyber Defense," IEEE Access, vol. 12, pp. 3456–3470, 2025.
[20]. Sharafaldin, I., Lashkari, A.H., & Ghorbani, A.A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. CICIDS 2017.
[21]. Kaggle. (2025). Network Intrusion Dataset (CIC-IDS-2017). https://www.kaggle.com/datasets/chethuhn/network-intrusion-dataset
[22]. DarkOwl. (2025). Darknet Intelligence Feeds. https://www.darkowl.com
[23]. FalconFeeds. (2025). Cyber Threat Intelligence Feeds. https://falconfeeds.io
[24]. Anomali. (2025). Threat Intelligence Marketplace. https://www.anomali.com/marketplace/threat-intelligence-feeds