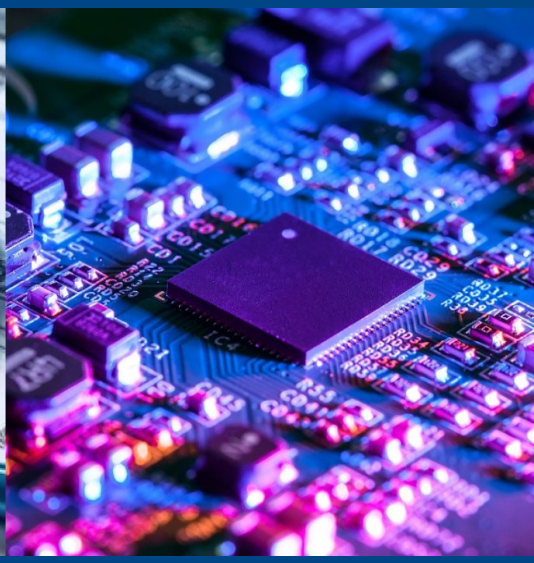
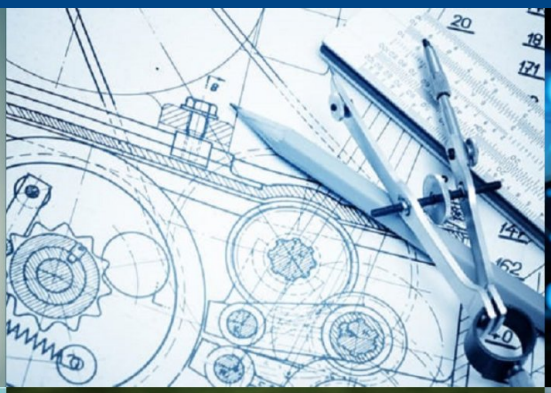
ISSN: 2582-7219

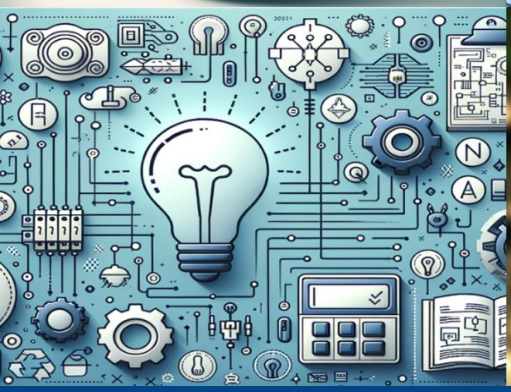




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904387

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI Driven Network Intrusion Detection System

Dr. Amsavalli.S, Payvati Mujeeb AAFAQ

Department of Computer Applications B.S.Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai, Tamil Nadu, India

ABSTRACT: The rapid growth of internet usage and digital communication has significantly increased the risk of cyber threats and network-based attacks. Traditional Network Intrusion Detection Systems (NIDS) rely primarily on signature-based techniques, which are ineffective against zero-day and evolving attacks. To address these limitations, this paper proposes an AI-Driven Network Intrusion Detection System that leverages Machine Learning algorithms to detect malicious activities in network traffic. The system analyzes network packets, performs data preprocessing and feature extraction, and classifies traffic as either normal or intrusive using supervised learning models such as Random Forest, Support Vector Machine, and Logistic Regression. Publicly available benchmark datasets such as NSL-KDD and CICIDS2017 are utilized for training and evaluation. Experimental results demonstrate improved detection accuracy, reduced false positive rates, and enhanced capability to identify previously unseen attack patterns compared to traditional rule-based approaches. The proposed system provides a scalable and intelligent solution for modern cybersecurity challenges and can be further extended for real-time monitoring and automated alert generation. This research contributes to strengthening network security infrastructures through the integration of artificial intelligence techniques.

KEYWORDS: Artificial Intelligence, Network Intrusion Detection System, Machine Learning, Cybersecurity, Anomaly Detection, Random Forest**.**

I. INTRODUCTION

The rapid expansion of digital communication and internet-based services has significantly increased the vulnerability of computer networks to cyber threats. Organizations across sectors, including finance, healthcare, education, and government, rely heavily on networked systems for critical operations. As a result, ensuring the security and integrity of network infrastructures has become a major challenge. Cyberattacks such as Distributed Denial of Service (DDoS), phishing, malware injection, brute-force attacks, and unauthorized access attempts continue to evolve in complexity and frequency.Traditional Network Intrusion Detection Systems (NIDS) are primarily based on signature-based and rule-based detection techniques. While these methods are effective in identifying known attack patterns, they fail to detect novel or zero-day attacks that do not match predefined signatures. Furthermore, maintaining and updating signature databases requires continuous manual effort, making conventional systems less adaptable to emerging threats. This limitation highlights the need for intelligent and adaptive security mechanisms.Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have opened new opportunities for enhancing cybersecurity systems. AI-driven intrusion detection systems are capable of learning patterns from historical network traffic data and distinguishing between normal and malicious behavior. By leveraging supervised learning algorithms such as Random Forest, Support Vector Machine (SVM), Logistic Regression, and deep learning models, these systems can achieve higher detection accuracy and improved generalization to unseen attack types.

II. LITERATURE REVIEW

Recent research between 2023 and 2025 has significantly Recent research has significantly advanced the development of AI-driven Network Intrusion Detection Systems (NIDS). In [1], Md. Alamin Talukder et al. (2024) proposed a machine learning–based intrusion detection framework addressing big and imbalanced datasets using oversampling, stacking feature embedding, and Principal Component Analysis (PCA). Their model was evaluated on UNSW-NB15 and CIC-IDS-2017/2018 datasets and achieved over 99% accuracy using Random Forest and Extra Trees classifiers., K. Mythily Sai Chandana et al. (2024) presented a deep learning–based NIDS that leverages neural network architectures to enhance detection performance in modern network environments. The study emphasized improved feature learning and pattern recognition capabilities for detecting complex cyberattacks.Farhan et al. (2025) in [3] developed a deep neural network model for network-based intrusion detection, demonstrating high detection

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

performance on real-world attack scenarios. Their approach highlighted the effectiveness of deep learning in identifying sophisticated and emerging threats.Similarly, Mustafa Al Lail et al. (2023) in [4] conducted a comparative study of multiple machine learning algorithms for intrusion detection using modern attack datasets. Their findings indicated that the Random Forest classifier achieved approximately 97% detection accuracy, outperforming several other traditional models.

III. PROBLEM DEFINITION

The rapid growth of computer networks, cloud computing, and Internet-based services has significantly increased the exposure of organizational systems to cyber threats. Modern networks generate massive volumes of traffic data, making manual monitoring and traditional security mechanisms insufficient for effective threat detection. Conventional Network Intrusion Detection Systems (NIDS), primarily based on signature or rule-based approaches, are limited in their ability to detect unknown or zero-day attacks. Additionally, these systems often suffer from high false positive rates and lack adaptability to evolving attack patterns.With the increasing complexity and sophistication of cyberattacks such as Distributed Denial of Service (DDoS), brute force attacks, malware propagation, and advanced persistent threats, there is a critical need for intelligent detection mechanisms capable of analyzing large-scale network traffic in real time. Traditional systems also struggle with imbalanced datasets and high-dimensional feature spaces, which reduce detection accuracy and overall system efficiency.Therefore, the core problem addressed in this work is the design and development of an AI-driven Network Intrusion Detection System capable of accurately identifying and classifying malicious network activities while minimizing false alarms. The proposed system aims to leverage machine learning techniques to enhance detection performance, handle large and imbalanced datasets, and provide a scalable and adaptive solution for modern network security environments.

IV. PROPOSED SYSTEM

The proposed system presents an Artificial Intelligence–driven Network Intrusion Detection System (AI-NIDS) designed to enhance network security by accurately detecting and classifying malicious activities in real time. The system utilizes benchmark datasets such as CIC-IDS-2017 and UNSW-NB15 for training and evaluation. Initially, network traffic data undergoes preprocessing steps including data cleaning, normalization, encoding of categorical features, and handling of class imbalance to ensure data quality and model reliability. Feature selection techniques such as Principal Component Analysis (PCA) are applied to reduce dimensionality and improve computational efficiency. Supervised machine learning algorithms, including Random Forest and Support Vector Machine (SVM), are implemented to classify traffic as normal or malicious. The model performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score to ensure robustness and reduced false alarm rates. Once validated, the trained model can be deployed in a real-time network monitoring environment to continuously analyze traffic and generate alerts upon detection of suspicious activities. The proposed AI-NIDS aims to provide a scalable, adaptive, and intelligent solution capable of detecting both known and unknown cyber threats in modern network infrastructures.

V. SYSTEM ARCHITECTURE

The architecture follows a layered design:

Physical Layer: This layer includes network infrastructure components such as routers, switches, servers, firewall systems, and GPU-enabled hardware used for model training and real-time traffic monitoring.

- **Data Layer:** The Data Layer is responsible for network traffic collection, storage, and preparation before feeding into the machine learning model. It captures raw packet data, converts it into flow-based features, removes noise or missing values, and performs normalization and encoding.
- **Application Layer:** This layer performs data preprocessing, feature extraction, feature selection, and classification using machine learning algorithms such as Random Forest and Support Vector Machine (SVM). It identifies whether the incoming traffic is normal or malicious.
- **Analytics Layer:** The Analytics Layer evaluates model performance and provides quantitative analysis using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. It also generates alerts and reports for detected intrusions.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904387

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. METHODOLOGY

The proposed AI-Driven Network Intrusion Detection System follows a structured processing pipeline to accurately detect and classify malicious network activities.

Data Collection
Network traffic data is collected from benchmark intrusion detection datasets such as CICIDS 2017 and UNSW-NB15. These datasets contain both normal and attack traffic, including DoS, DDoS, brute-force, and probing attacks. The collected data is divided into training, validation, and testing sets to ensure unbiased model evaluation.

Data Preprocessing
The raw network traffic data is cleaned to remove missing, duplicate, and inconsistent records. Feature normalization and scaling techniques are applied to standardize numerical values. Categorical features such as protocol type and service are encoded into numerical form. Data balancing techniques are used to address class imbalance and improve model generalization.

Feature Extraction and Model Design
Relevant flow-based features such as packet count, flow duration, byte rate, and protocol information are extracted from network traffic. A supervised machine learning model, such as Random Forest or Support Vector Machine, is designed to learn patterns from these features. The model is trained to distinguish between normal traffic and various intrusion types based on learned behavioral characteristics.

Model Evaluation
The trained model is evaluated using performance metrics such as Accuracy, Precision, Recall, and F1-Score. The predicted intrusion labels are compared with ground truth labels to assess detection effectiveness. Confusion matrix analysis is used to measure classification performance and false alarm rates.

Decision Engine
The Decision Engine analyzes model predictions to identify potential intrusions in real time. Detected attacks are categorized based on their type and severity. This component enables timely alerts and supports automated or manual security responses to mitigate network threats.

Reporting and Visualization
The system generates detailed reports and visual dashboards showing detected intrusions, attack distribution, and network behavior trends. This methodology ensures accurate intrusion detection with reduced false positives, improved detection efficiency, and enhanced reliability for real-time network security monitoring.

ALGORITHM
The AI-Driven Network Intrusion Detection Algorithm detects malicious activities in network traffic using machine learning techniques. Initially, raw network traffic data is collected and preprocessed to ensure consistency and reliability. Relevant traffic features are then extracted and analyzed by a trained AI model to identify intrusion patterns. Finally, the system classifies traffic as normal or malicious and generates appropriate alerts.

AI-DrivenNetwork Intrusion Detection Algorithm
The proposed AI-NIDS Algorithm is an intelligent, data-driven framework designed to accurately detect and classify network intrusions from large-scale network traffic data. The algorithm combines effective data preprocessing, feature extraction, and supervised machine learning techniques to achieve high detection accuracy while minimizing false positives.

The algorithm begins with data preprocessing, where raw network traffic records are cleaned to remove missing, duplicate, and noisy entries. Numerical features are normalized to standardize value ranges, while categorical attributes such as protocol type and service are encoded into numerical representations. To improve model robustness and address skewed attack distributions, data balancing techniques are applied.

In the feature extraction stage, flow-based and statistical features such as packet count, flow duration, byte rate, source

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

and destination ports, and protocol behavior are extracted from the network traffic. These features capture both short-term and long-term traffic characteristics essential for identifying intrusion patterns.

Next, the processed feature vectors are passed to a supervised machine learning classifier, such as Random Forest or Support Vector Machine. The model learns complex relationships between traffic features and attack types during the training phase. Hyperparameter tuning is performed to optimize classification performance and generalization capability.

During the classification stage, the trained model performs traffic classification to distinguish between:

- Normal Traffic
- Denial-of-Service (DoS / DDoS) Attacks
- Brute-Force Attacks
- Probe and Scan Attacks
- Other Malicious ActivitiesThe algorithm outputs probability-based predictions, enabling accurate intrusion identification.In the decision and alert generation stage, detected intrusions are analyzed based on severity and attack type. Real-time alerts are generated for network administrators to enable rapid response and threat mitigation.

Finally, in the evaluation and reporting stage, the system performance is assessed using metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. The algorithm ensures efficient intrusion detection, reduced false alarms, and reliable performance for real-time network security applications

VII. IMPLEMENTATION

The proposed AI-driven Network Intrusion Detection System (NIDS) was implemented using Python with machine learning libraries such as Scikit-learn, Pandas, and NumPy. Benchmark datasets including NSL-KDD and CICIDS2017 were used, and the data was divided into training and testing subsets for unbiased evaluation.

Data preprocessing involved encoding categorical features, handling missing values, and normalizing input data. The class labels were converted into binary form to distinguish between normal and malicious traffic. Relevant features were selected to reduce dimensionality and improve computational efficiency.

Supervised machine learning models, namely Random Forest, Support Vector Machine (SVM), and Logistic Regression, were trained to classify network traffic. Random Forest, as an ensemble method, improved robustness, while SVM handled high-dimensional data effectively, and Logistic Regression served as a baseline model.

Model performance was evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score, along with confusion matrix analysis. Experimental results indicated that the Random Forest model achieved the highest detection accuracy with reduced false positives.

The system was further tested for real-time prediction by classifying incoming network traffic instances. The implementation demonstrates that machine learning-based NIDS can effectively detect both known and unknown cyber threats, providing a scalable and efficient solution for modern network security.

VIII. EXPERIMENTAL RESULTS

The performance of the proposed AI-driven Network Intrusion Detection System (NIDS) was evaluated using the NSL-KDD and CICIDS2017 datasets. The dataset was split into training and testing subsets to ensure unbiased evaluation. Three machine learning models—Random Forest, Support Vector Machine (SVM), and Logistic Regression—were trained and compared.

The evaluation was conducted using standard performance metrics, including Accuracy, Precision, Recall, and F1-Score. Among the models, Random Forest achieved the highest performance with an accuracy of approximately 98%, followed by SVM with around 95%, and Logistic Regression with about 92%. The Random Forest model also demonstrated a lower false positive rate, making it more reliable for intrusion detection.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The confusion matrix analysis indicated that the proposed system effectively distinguished between normal and malicious traffic, with high true positive and true negative rates. Additionally, the models showed the ability to generalize well on unseen data, indicating robustness against unknown attack patterns.

Overall, the experimental results confirm that the proposed AI-based NIDS significantly outperforms traditional signature-based approaches, providing improved detection accuracy and enhanced capability to identify evolving cyber threats.

IX. DISCUSSION

The experimental results demonstrate that the proposed AI-driven Network Intrusion Detection System (NIDS) effectively improves the detection of malicious network activities compared to traditional signature-based methods. The superior performance of the Random Forest model can be attributed to its ensemble learning capability, which enhances classification accuracy and reduces overfitting. While Support Vector Machine and Logistic Regression also produced satisfactory results, their performance was comparatively lower, particularly in handling complex and high-dimensional network data.

The system showed strong generalization ability by accurately identifying both known and previously unseen attack patterns, highlighting the advantage of machine learning in anomaly detection. The low false positive rate further indicates the reliability of the model in real-world scenarios, where minimizing false alarms is critical.

However, the system has certain limitations. Its performance depends heavily on the quality and diversity of the training dataset, and it may require periodic retraining to adapt to emerging cyber threats. Additionally, real-time deployment may introduce computational overhead, especially with large-scale network traffic.

Overall, the proposed approach provides a scalable and efficient solution for modern cybersecurity challenges, and future work can focus on integrating deep learning techniques and real-time monitoring systems to further enhance detection capabilities.

ADVANTAGES

- High detection accuracy due to machine learning-based classification
- Capable of detecting zero-day and unknown attacks
- Reduced false positive rate compared to traditional methods
- Scalable and efficient for large network traffic data Robust performance using ensemble models like Random Forest
- Easily extendable for real-time monitoring and advanced AI integration

X. LIMITATIONS

- Performance depends on the quality and diversity of the training dataset
- Requires periodic retraining to adapt to new and evolving cyber threats
- Computational overhead for real-time detection in high-speed networks
- May struggle with highly imbalanced datasets without proper handling
- Limited interpretability of some machine learning models
- Initial setup and tuning can be time-consuming

XI. FUTURE ENHANCEMENT

- Integration of deep learning models such as LSTM and CNN for improved detection
- Implementation of real-time intrusion detection using live network traffic capture
- Development of a web-based dashboard for monitoring and visualization
- Incorporation of automated alert and response systems (email/SMS notifications)
- Use of advanced feature selection and dimensionality reduction techniques
- Deployment on cloud platforms for scalability and remote access
- Integration with threat intelligence systems for proactive attack prevention .

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

XII. CONCLUSION

The proposed AI-driven Network Intrusion Detection System (NIDS) effectively demonstrates the potential of machine learning techniques in strengthening modern network security. By leveraging supervised learning models such as Random Forest, Support Vector Machine, and Logistic Regression, the system is capable of accurately classifying network traffic into normal and malicious categories. The implementation highlights the importance of proper data preprocessing, feature selection, and model evaluation in achieving reliable performance.

The experimental results show that the proposed system achieves high detection accuracy with reduced false positive rates, particularly with the Random Forest model, which proved to be the most effective among the implemented approaches. Unlike traditional signature-based systems, the proposed model is capable of detecting zero-day and previously unseen attacks, making it more adaptable to evolving cyber threats.

Furthermore, the system demonstrates good generalization ability when tested on unseen data, indicating its suitability for real-world applications. Although certain limitations exist, such as dependency on dataset quality and computational requirements, the overall performance validates the effectiveness of integrating artificial intelligence into intrusion detection systems.

In conclusion, the proposed NIDS provides a scalable, efficient, and intelligent solution for detecting network intrusions. It serves as a strong foundation for future enhancements, including real-time deployment, integration with deep learning models, and incorporation of automated response mechanisms for improved cybersecurity infrastructure.

REFERENCES

[1] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009.
[2] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization (CICIDS2017),” in Proc. ICISSP, 2018.
[3] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[4] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, “Applied Logistic Regression,” 3rd ed., Wiley, 2013.
[5] S. Axelsson, “Intrusion detection systems: A survey and taxonomy,” Technical Report, Chalmers University, 2000.
[6] A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
[7] N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems,” in Proc. Military Communications and Information Systems Conference, 2015.
[8] H. Hindy et al., “A taxonomy of network threats and the effect of current datasets on intrusion detection systems,” IEEE Access, vol. 8, pp. 104650–104675, 2020.
[9] M. Ring, D. Schlör, D. Landes, and A. Hotho, “Flow-based network traffic generation using GANs for intrusion detection,” Computers & Security, vol. 82, pp. 156–172, 2019.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com

Crime Pattern Analysis and Hotspot Prediction Using Machine Learning System

Sabaria S, Musfira Zaka, Lavanya .S

Assistant Professor, Department of Computer Applications B.S.Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai Tamil Nadu, India

B.sc Computer Science III year, Department of Computer Applications, B.S Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai Tamil Nadu, India

B.sc computer science III Year, Department of Computer Applications, B.S Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai Tamil Nadu, India

ABSTRACT: Crime analysis and hotspot identification play a critical role in enhancing public safety and assisting law- enforcement agencies in proactive crime prevention.Tradition crime monitoring methods rely primarily on manual reports and static statistical summaries, which Patterns and hotspot localization's are frequently not visible in the raw data of large datasets. This paper proposes a machine learning-driven framework for crime hotspot detection and spatial crime analysis using crime data collected from various districts of Tamil Nadu. The proposed system utilizes crime attributes including crime type, geographic location to analyze crime distribution patterns.

Multiple data analysis Spatial analysis and machine learning techniques are employed to pre-process crime data, perform cluster analysis, produce visualization maps to help locate hot spots and spatial models that describe the environmental causes of crime and detect high-risk regions. Spatial visualization methods such as geographic heat maps are employed to represent crime density across different locations, enabling intuitive identification of crime hotspots.

This work relies heavily on a web-based crime data mining system that uses historical crime incident data and geographical information system data to build a crime forecasting model.visualization dashboard is developed to present crime density maps and analytical insights for decision support. The proposed approach enables efficient crime monitoring, improved hotspot identification, and supports data-driven policing strategies for better law enforcement resource allocation.

KEYWORDS: Crime prediction, Machine Learning, Random Forest, Heat Map, Hotspot Analysis, Data Mining

I. INTRODUCTION

Crime, necessitating advanced analytical methods, is among the most significant issues globally. Law enforcement addresses crime due to its focus on the way prevention appears in complex environments. The issue involves manual observation and basic statistics, and the hidden patterns of data reflect the analyst's idea of the connection between urbanization and population.

The complex and growing crime data contains various elements of analysis, such as time and error, as well as the principles, for instance, effectiveness and conventionality..

undertaken to predict the crime using machine learning techniques. Previous studies employed K Nearest Neighbors (KNN) and Decision Trees, yielding fundamental accuracy with interpretability. Recent research is directed towards more complicated models, for example Random Forest model, Support Vector Machines (SVM) and Deep Learning approaches such as Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM). Another approach

This project focuses on crime data from Tamil Nadu and aims to develop an intelligent system for crime analysis and prediction. The system performs comprehensive data visualization using dashboards and charts to analyze past crime trends across different cities.

I. LITERATURE REVIEW

There are several research works that have been

that gained strong popularity is heat map visualization using Geographic Information Systems (GIS) tools to identify crime hotspots and spatial patterns.

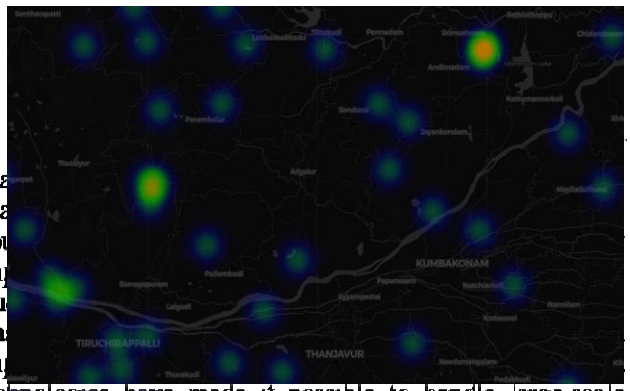


Fig 1 Heat Map

technologies have made it possible to handle large scale crime datasets effectively.

However, most existing systems have limitations in real time implementation and user friendly visualization. Many models are costly to run and need large datasets for training, which makes them less practical for smaller organizations. Plus, there is a lack of systems that integrate prediction, visualization, and decision support into one platform. As a result, there is a need for a system that is scalable, efficient, and easy to use. Such a system should not only predict crime accurately but also offer clear visual insights for law enforcement agencies

IV. METHODOLOGY

Using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

II. DATASET DESCRIPTION

The dataset used in this study comes from crime records gathered from several open data sources and public reports about crime statistics in Tamil Nadu. It includes different types of criminal activities recorded in various cities and regions. This information allows for both spatial and statistical analysis. The collected data serves as the basis for understanding crime patterns, identifying hotspots, and creating predictive models.

The dataset includes several important attributes that affect crime analysis and prediction. These features were chosen to reflect both the time and place of criminal activities.

The main attributes considered in this study are:

- Crime ID
- Crime type(e.g.Theft,Assault,Cyber Crime,Robbery)

- Location (City/Region)
- Date and Time of Occurrence
- Crime Category (Domestic / Non-Domestic)
- Latitude and Longitude (for geographic mapping)
- Number of Crimes per Location
- Crime Density Information

Figure 1 illustrates the geographic visualization of crime distribution, highlighting different regions where crime intensity varies. The map provides a clear understanding of high-density and low-density areas using heatmap techniques.

A. Data Collection

The data is collected from various government, non government websites, available datasets, and other websites. Also some of the datasets are collected from other resources such as online news using web scraping, etc.

B. Data Preprocessing

The data is cleaned and pre-processed to remove redundancy and fill the gaps in the data for achieving a smooth and complete data set. This dataset results in a smooth and accurate prediction. The data is arranged as required.

C. Data Analysis

The data is analyzed for required information which will become an input to the predicting algorithm later. Data analyses helps to know the data and take required measures for the machine learning model to perform accurately.

D. Data Prediction

The data is then feed to the prophet tool which predicts the crime rate of certain crimes in a specific area. This tool works on the date time column i.e., the time series, to produce its output.

E. Data Visualization

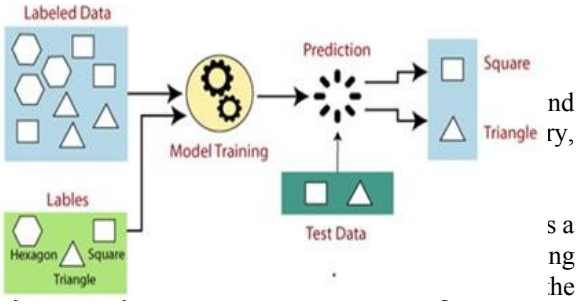
This website provides various forms in which the data can be visualized such as heat map, pie chart, bar graph, etc. It helps in understanding large datasets.

V. FRAMEWORK

A. Flask

Flask is a web framework that allows developers to build lightweight web applications quickly and easily with Flask Libraries. It was developed by Armin Ronacher, leader of the International Group of Python Enthusiasts(POCCO). It is based on the WSGI toolkit and Jinja2 templating engine

VI. LIBRARIES



A. Pandas

Pandas is a Python library designed for data manipulation and nd functions tailored for working with numerical tables and time s ry, NumPy

B. Matplotlib

Matplotlib, a plotting library compatible with both Python and its s a potent tool for individuals engaged in Python and NumPy based ng visualizations. And for making statistical interference, it becomes ve he tool that can be very helpful for this purpose.

C. Seaborn

Seaborn, built upon matplotlib, serves as a Python data visualization library offering a sophisticated interface for crafting visually appealing and informative statistical graphics. Seaborn is a library for making statistical graphics in Python. Expanding upon matplotlib and tightly integrating with pandas data structures, Seaborn's plotting functions are tailored to operate seamlessly on dataframes and arrays encompassing entire datasets. Internally, they execute essential semantic mapping and statistical aggregation, culminating in the creation of insightful plots. Its dataset oriented, declarative API lets us focus on what the different elements of our plots mean, rather than on the details of how to draw them

D. Numpy

NumPy enhances Python with support for large, multi- dimensional arrays and matrices, complemented by a vast array of highlevel mathematical functions tailored for manipulating these arrays. NumPy is open-source software and has many contributors

VII. ALGORITHM

A. Supervised machine learning

- This system uses supervised learning algorithm for prediction. Supervised learning falls within the realm of machine learning, where labeled datasets are employed to train algorithms, enabling them to predict outcomes and identify patterns. Labelled data refers to input data that has been pre-assigned with corresponding correct output values. In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly)

.

Fig2. Supervise algorithm

B. LINEAR REGRESSION

The piecewise linear regression model in Prophet captures the overall trend in the data while allowing for flexibility and adaptability to changes over time. This approach differs from traditional linear regression models, which assume a single linear relationship between the predictor variables and the target variable.

By incorporating piecewise linear regression, Prophet can capture complex trends and patterns in the time-series data, making it particularly suitable for forecasting tasks

where the trend may exhibit non-linear behavior or undergo changes over time.

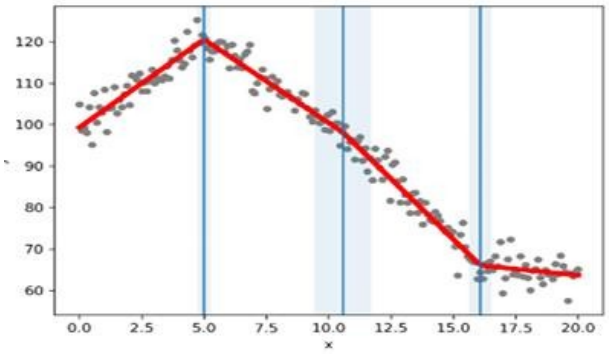


Fig3. Linear Algorithm

VIII. PROPOSED SYSTEM

The proposed system uses Machine Learning and Data Visualization to overcome existing limitations.

A. Features:

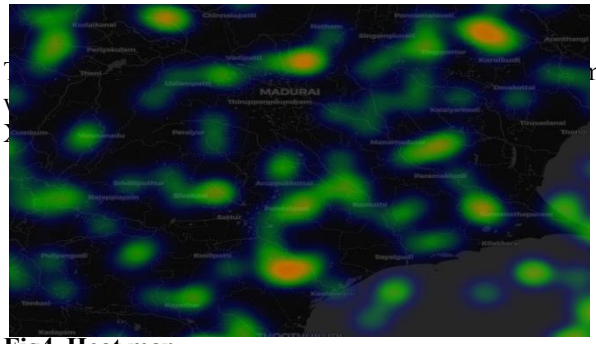
- Data Pre-processing and Cleaning
- Crime prediction using Random Forest Machine Learning algorithm
- Heat map generation using latitude & longitude
- User-friendly visualization

B. Advantages**:**

- Faster prediction with accuracy
- Identify hotspots where crimes are happening
- Facilitates decision-making for authorities

IX. RESULT

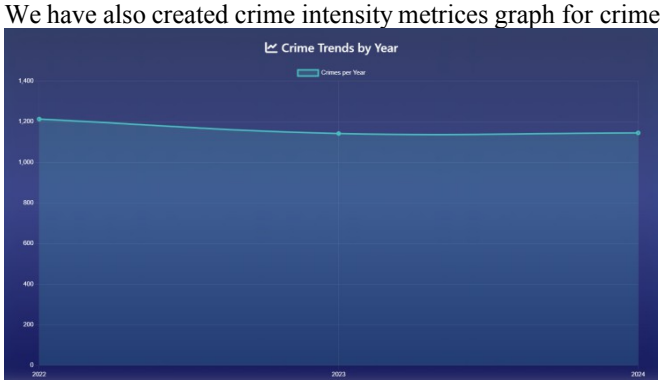
The required data is collected and preprocessed as required. Then we have created a heat map for depictions of areas with high crime rate.



ne rate and visualizes crime intensity of different areas in various

Fig4. Heat map

We have also created crime intensity metrices graph for crime trends over year



.

Fig5. Crime Trends

Then we have also created crime distribution for better understanding.

Fig6. Domestic and Non-Domestic

Then we have analysis part of the project to show the core crime details

The proposed crime analysis and prediction system has a wide range of practical applications across multiple domains. By combining data visualization, geospatial analysis, and machine learning, the system supports informed decision-making and enhances public safety.

- Police Departments:

Law enforcement agencies can use the system to identify crime hotspots, monitor trends, and allocate resources more effectively. Predictive insights help in proactive patrolling, faster response planning, and reducing crime rates in high-risk areas.

- Government Agencies:

Government bodies can utilize the system for policy formulation, urban planning, and safety initiatives. It enables data-driven decisions for improving infrastructure, surveillance systems, and law enforcement strategies.

- Crime Analysts:

Analysts can leverage the dashboards and heat-maps to study crime patterns, correlations, and trends over time. The system simplifies complex datasets into meaningful visual insights, aiding in accurate reporting and forecasting.

- Research and Academic Use:

The system serves as a valuable tool for researchers and students to study crime behavior, test machine learning models, and develop advanced analytical solutions.

XI. CONCLUSION

Crime is an unlawful act which disturbs the peace and harmony of the society. This projects aims to successfully predict

crime and their locations based on the historical crime data. The project uses machine learning which is an advanced and latest technology for accurate prediction. The web application will display crime rate in various areas. It is extremely useful for both the higher investigating authorities and officers designated to handle low level crime for tracking and stopping the crime. The predictions will help to ensure increased security and thus could help in lowering the crime rate. Overall, the project demonstrates the potential of data analysis and mapping technologies to improve public safety and inform decision-making. Proactive measures can be
taken to prevent crime and improve public safety by using data to identify crime hotspots and trends. Although there's more work needed to enhance the precision and breadth of the project, it marks a significant stride towards employing data-driven strategies to tackle intricate social challenges.

XII. REFERENCES

[1] Varun Mandalapu , Lavanya Elluri, Piyush Vyas, and Nirmala Roy, "Crime Prediction Using Machine Learning and Deep Learning: A
Systematic Review and Future Directions" , IEEE, Volume 11.

[2] Raza, D. M. & Victor, D. B. ''Data mining and region prediction based on crime using random forest'', 980–987 (IEEE, 2021). .

[3] Prakash Maurya, Tahir Shaikh, Imran Ahmed, Amaan Firdosi, Prof. Kiran Deshmukh, "Crime Analysis and Prediction Using Machine Learning", IJRASET, Volume 11, Issue 4.

[4] Tzu-Wei Hung1 ,Chun-Ping Yen, "Predictive policing and algorithmic fairness", Synthese (2023)

[5] Ruaa Mohammed Saeed, Husam Ali Abdulmohsin, "A study on predicting crime rates through machine learning and data mining using text", Journal of Intelligent Systems, Volume 32, Issue 1

.