Cybercrime Rate Prediction of Each City Using the KNN Algorithm

**1. Miss. Neha Gandham, 2. Prof. Pragati Malusare, 3. Miss. Aishwarya Bhoir,**

[neha.gandham@indiraicem.ac.in](mailto:neha.gandham@indiraicem.ac.in), [pragati.malusare@indiraicem.ac.in](mailto:pragati.malusare@indiraicem.ac.in) , [aishwaya.bhoir@indiraicem.ac.in](mailto:aishwaya.bhoir@indiraicem.ac.in),

**4. Miss. Vanshika Bhan, 5. Miss. Gloriya Francis**

[vanshika.bhan@indiraicem.ac.in](mailto:vanshika.bhan@indiraicem.ac.in), [gloriya.francis@indiraicem.ac.in](mailto:gloriya.francis@indiraicem.ac.in)

\*1, 3,4,5Student, \*2Faculty

Department of Computer Engineering, Indira College of Engineering and Management, Pune

Savitribai Phule Pune University, Pune, Maharashtra, India

-----------------------------------------------------\*\*\*\*---------------------------------------------------

***Abstract-*** This paper introduces an intelligent, interactive system for predicting the rates of cybercrime in Indian cities based on the K-Nearest Neighbors (KNN) algorithm. With the model being trained on past cybercrime data by state, city, crime category, and year, the system can produce future crime incidents with a great deal of accuracy. The site includes a dynamic interface whereby the user may input specified crime and location details, display results in interactive bar and line charts, and identify a ranked list of the top five high- and low-risk cities as well as the same for solved and unsolved cases. A map-based integrated interface facilitates enhanced spatial comprehension of the distribution of cybercrime. Drawing on the findings of recent machine learning reports, which have shown the efficacy of the KNN as well as other algorithms as predictive tools for crime, the study emphasizes their application for proactive planning and resource deployment. It seeks to assist law enforcers and policymakers in the identification of trends, the prioritizing of intervention, and the enhancement of the public's human security through data-informed decision-making.

**Keywords*:* KNN algorithm, Prediction, Cybercrime rate, Solved and unsolved cases, Indian Cities**

I. INTRODUCTION

With the proliferation of digital connectivity and internet usage in India, cybercrime increasingly threatens cities and states. Traditional methods of tracking and prevention proved ineffective with the rise in frequency and sophistication of such crimes. Therefore, the use of data-driven technologies started witnessing increased interest in the prospect they have in predicting and preventing cyber-attacks.

Machine Learning (ML) methods have been found useful in pattern identification and predictive analysis of past crime data. Among them, the K-Nearest Neighbors (KNN) algorithm is a straightforward yet effective classification and regression tool. KNN is based on the proximity concept prediction based on the class with the highest majority among the closest neighbors in the feature space. Its simplicity in implementation, as well as being non-parametric, makes it suitable for analyzing crime data, as the data tends to be labeled and structured by dimensions including the city, state, year, and the type of crime.

Although KNN's simplicity and low computational requirement make it valuable, it suffers when dealing with imbalanced data as well as data with a higher number of dimensions. In comparison with other algorithms including Random Forests, Support Vector Machines, and Neural Networks, the predictive accuracy of KNN can be lower on bigger and noisy data. Nevertheless, the virtue lies in the simplicity and efficacy when the dataset is clean, well-balanced, and structured as government crime data often is. This project envisions a smart, interactive system based on the KNN algorithm for forecasting cybercrime rates in Indian cities. With the model being taught with historical data, the system will be capable of predicting crime incidence in the future, classifying the case as resolved or unresolved, and presenting the results graphically through interactive dashboards. Users will be able to engage with the system to analyze the forecast by location and by type of crime, view the risk rankings, and analyze the results through graph-based and map-based visualizations. Overall, the vision is for law enforcement and citizens alike to be empowered by accessible, predictive knowledge of cybercrime patterns.

II. LITERATURE SURVEY

In this paper, the author [1] highlights the use of machine learning algorithms, including KNN, for cyberbullying detection, which can be applied to cybercrime rate prediction. It emphasizes feature selection, data challenges, and Big Data analysis, which are crucial for improving crime prediction accuracy. The study’s insights on classification, prediction, and handling imbalanced data can help refine KNN-based cybercrime forecasting in your system.

In this paper, the author [2] “Cyberspace and Geopolitics: Assessing Global Cybersecurity Norm Processes at a Crossroads" provides insights into global cybersecurity trends, geopolitical factors, and cybercrime patterns, which can help structure the cybercrime rate prediction model. It highlights regional cybersecurity policies, threat landscapes, and international cyber norms, which can be useful for defining crime rate prediction parameters. The study's discussion on cyber threats and risk assessment can aid in feature selection for the KNN algorithm, improving accuracy in identifying high-risk areas. Additionally, it emphasizes the role of data-driven approaches in cybersecurity, aligning with the use of KNN for crime trend forecasting in your system.

In this paper, the author [3] emphasizes predicting cybercrime using machine learning and data-driven approaches, which align with implementing KNN for cybercrime rate prediction. They highlight the importance of feature selection, data preprocessing, and risk assessment to improve accuracy. Additionally, both studies discuss the impact of globalization and cybersecurity trends on crime patterns, reinforcing the need for real-time crime analysis. Their focus on IT-driven crime forecasting supports the use of KNN to classify and predict cybercrime rates across different regions.

In this paper, the author [4] highlights the importance of Big Data analysis in crime prediction and prevention, which supports implementing KNN for cybercrime rate forecasting. It emphasizes data preprocessing, feature extraction, and real-time analysis to enhance prediction accuracy. The study also discusses identifying crime patterns and high-risk areas, which aligns with using KNN to classify and predict cybercrime trends. Additionally, it underscores how intelligent data analytics improve crime prevention strategies, reinforcing the need for data-driven decision-making in cybersecurity.

In this paper, the author [5] provides insights into cybercrime trends in India, which helps in implementing KNN for cybercrime rate prediction. It discusses various types of cybercrimes, their impact, and statistical analysis, aiding in feature selection and crime categorization. The study highlights the growing need for predictive models to analyze cybercrime rates, aligning with KNN’s ability to classify and forecast crime trends. Additionally, it emphasizes real-time monitoring and data-driven approaches, supporting the development of an efficient cybercrime prediction system.

In this paper, the author [6] focuses on malware detection using feature engineering and behaviour analysis, which supports feature selection for implementing KNN in cybercrime rate prediction. It highlights data preprocessing, pattern recognition, and anomaly detection, which are crucial for improving KNN-based crime forecasting. The study’s approach to identifying malicious activities and trends can be adapted to classify and predict cybercrime rates. Additionally, it reinforces the importance of machine learning in cybersecurity, aligning with using KNN to analyze cybercrime patterns effectively.



Table (1). Comparison of Proposed System with Existing System

III. METHODOLOGY

The cybercrime rate prediction system using KNN collects and preprocesses historical crime data, including reported cases, solved/unsolved cases, and population density. The KNN algorithm is applied for classification (categorizing cities into low, moderate, or high-risk zones) and regression (predicting future crime rates using K-nearest neighbors). An interactive map system allows users to click on locations to view real-time crime statistics, while charts and dashboards visualize trends.

Fig 1. Shows the cybercrime dashboard that allows users to log in using their Aadhaar card for authentication. Once authenticated, users can select a state, city, and cybercrime type to view relevant crime statistics. The system retrieves data from a MySQL database and uses a KNN algorithm to predict and classify cybercrime rates based on population and solved and unsolved cases. The results are displayed dynamically, showing crime rates, the number of solved and unsolved cases, and the top five cities with the lowest and highest cybercrime rates. Additionally, the system integrates an admin panel for database management and analytics.

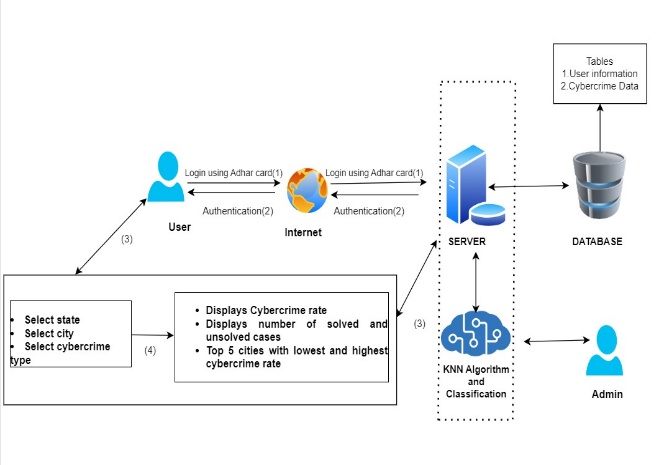
******

Fig (1). Proposed System

A screenshot of a diagram

Description automatically generated

Fig 2).DataFlow Diagram

Fig 2. Shows The Cyber Crime Prediction System follows a structured workflow, beginning with user authentication via Aadhar verification. Upon successful login, users select a state and city, enabling location-based cybercrime analysis. The system then displays crime rates, top 5 high-risk cities, and solved/unsolved cases. Using machine learning algorithms (KNN and classification models), the system predicts future cybercrime trends. Finally, prediction results are displayed, aiding law enforcement in proactive decision-making.

IV. RESULT AND DISSCUSION

The performance of the proposed K-Nearest Neighbors (KNN) model was evaluated using a labeled dataset of historical cybercrime records from multiple Indian cities.

The interactive system successfully:

* Predicted future cybercrime rates by city and crime type.
* Displayed bar and line charts for trend analysis.
* Ranked cities by risk levels (high-risk and low-risk).
* Presented a map-based interface for geographic crime visualization.

V. CONCLUSION AND FUTURE WORK

A. Conclusion

The proposed system aims to leverage machine learning techniques, specifically the K-Nearest Neighbors (KNN) algorithm and classification methods, to predict and analyze cybercrime trends across various cities in India. By utilizing historical crime data, the system will predict the cybercrime rate for each city, classify cases as solved or unsolved, and highlight the top 5 cities with the highest and lowest crime rates.

B. Future Work

1. Enhanced Predictive Accuracy

• Data Enrichment: By integrating diverse datasets (e.g., social media activity, economic indicators, and demographic data), KNN can improve its predictive accuracy.

• Feature Engineering: Identifying and incorporating relevant features, such as local cybersecurity incidents or known vulnerabilities, can refine predictions.

2. Real-Time Analytics

• Dynamic Monitoring: Implementing KNN in real-time systems could help continuously update crime predictions based on incoming data, allowing law enforcement to respond more swiftly.

• Anomaly Detection: KNN can assist in identifying unusual patterns or spikes in cyber activities, aiding in the early detection of

VI. REFERENCES

1. Mohammed Ali-Garadi1, Mohmmad Rashid Hussain2, Nawsher Khan2, Ghulam Murtaza1,3, Henry Friday Nweke 1, Ihsan Al. “Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges.” Received April 21, 2019, accepted May 14, 2019, date of publication May 22, 2019, date of current version June 11, 2019.
2. Christian Ruhl, Duncan Hollis, Wyatt Hoffman, and Tim Maurer. “Cyberspace and Geopolitics: Assessing Global Cybersecurity Norm Pro cesses at a Crossroads.” FEBRUARY 2020
3. Segundo Moisés Toapanta Toapanta. “Analysis to Predict Cybercrime Using Information Technology in a Globalized Environment”. 2020 3rd International Conference on Information and Computer Technologies (ICICT).
4. Galal A. AL-Rummana1, Abdulrazzaq H. A. Al-Ahdal1 and G.N. Shinde. “The Role of Big Data Analysis in Increasing the Crime Prediction and Prevention Rates”. Subhendu Kumar Pani, Sanjay Kumar Singh, Lalit Garg, Ram Bilas Pachori and Xiaobo Zhang (eds.) Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications, (209–220) © 2021 Scrivener Publishing LLC
5. Priyanka Datta, Surya Narayan Panda, Sarvesh Tanwar. “A technical Review Report on Cyber Crimes in India.” 2020 International Conference on Emerging Smart Computing and Informatics (ESCI) AISSMS Institute of Information Technology, Pune, India. Mar 12-14, 2020.
6. Manuel Torres, Rafael Alvarez, And Miguel Cazorla. “A Malware Detection Approach Based on Feature Engineering and Behaviour Analysis”. 2020 International Conference on Emerging Smart Computing and Informatics (ESCI)
7. Pratibha, Akanksha Gahalot, Uprant, Suraina Dhiman, Lokesh Chouhan. “Crime Prediction and Analysis.” Conference Paper ·February 2020 DOI: 10.1109/IDEA49133.2020.9170731
8. Akash Kumar, Aniket Verma, Gandhali Shinde “Crime Prediction Using K-Nearest Neighboring Algorithm.”2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)