ISSN: 2633-4828 Vol. 5 No.2, June, 2023

International Journal of Applied Engineering & Technology

A MACHINE LEARNING APPROACH TO ANALYZE AND FORECAST CYBERCRIME PATTERNS ACROSS INDIAN STATES

Manisha Bharambe[1*], Poonam Ponde[2], Harshita Vachhani[3] and Abhishek Vijhani[4]

[1]Associate Professor, Department of Computer Science, MES Abasaheb Garware College, Pune, India

[2]Associate Professor, Department of Computer Science, Nowrosjee Wadia College, Pune, India

[3]Assistant Professor, Department of Computer Science, Pratibha College of Commerce and Computer Studies, Pune, India

[4]SDE II, Expedia Inc., Seattle, Washington, USA

*mgb.agc@mespune.in

ABSTRACT

India's socioeconomic landscape has undergone significant changes as a result of the country's rapid adoption of digital technologies and increased internet access. In the Modern era, digital infrastructure is essential to all industries, including governance, healthcare, education, and finance. India is poised for digital dominance with over 750 million internet users and a thriving digital economy. There are also serious cyber security issues as a result of this change. The past decade has witnessed a significant rise in cybercrimes, ranging from ransomware attacks and online harassment to phishing and identity theft, underscoring the urgent need for robust cybersecurity awareness and policies. This study shows the rise in cybercrimes across states by using line visualization and scatter plot analysis. This research uses a Machine learning model to predict cybercrimes in the future and identifies the socioeconomic factors that contribute to the rise of cybercrimes. This study aims to understand the growth of cybercrime in India and predict future trends by applying regression analysis, a widely used machine learning technique.

Keywords: *Cybercrime, Machine learning, Regression analysis, visualization, phishing, ransomware*

1. INTRODUCTION

Cybercrimes are illegal activities conducted through computers or other digital devices. These offenses include financial fraud, malware attacks (where malicious software is installed), data theft, hacking, phishing (which involves obtaining sensitive information like credit card details or passwords), online fraud, and cyberstalking. Essentially, cybercrime is any criminal activity that either targets or uses a computer, a computer network, or a networked device. The increasing digitization of society has directly contributed to a rise in cybercrime, making security measures crucial to combat it.

To understand and address this growing problem, a study utilized historical data from 2013 to 2022, obtained from credible sources such as the National Crime Records Bureau (NCRB) and the Press Information Bureau (PIB). This research investigates patterns in cybercrime, with scatter plot analysis and trend line visualization clearly illustrating a consistent increase in incidents across various states, highlighting the critical need for immediate action. The study further employs machine learning and predictive analytics, specifically using linear regression, to develop a robust predictive model. This model not only validates the ongoing growth in cybercrime but also projects its accelerating rate up to the year 2026. By pinpointing significant socio-economic, technological, and demographic factors that influence cybercrime rates, the research provides valuable, actionable insights. These insights can empower policymakers, law enforcement agencies, and citizens to implement preventative measures. Ultimately, the study's findings are intended to foster a proactive cybersecurity strategy, where prevention and education play key roles in reducing vulnerabilities and building greater digital resilience. Machine learning and predictive analytics offer valuable tools for understanding complex trends and making informed decisions. In this research, linear regression is used to establish a predictive model that not only confirms the growth in cybercrime but also projects the rate of increase up to 2026. By identifying key socio-economic, technological, and demographic factors influencing cybercrime rates, this study provides actionable insights that can help policymakers, law enforcement agencies, and citizens take preemptive measures. Our

findings aim to support a proactive cybersecurity strategy, where preventive action and education reduce vulnerability and enhance digital resilience.

2. LITERATURE REVIEW

Apoorva Bhangla et al. [1] focus on the legal aspects of cybercrime in India that affect women. The study examine the rise of cyberstalking, cyber pornography, and impersonation. They discussed the role of the Information Technology (IT) act of 2000 in addressing these crimes. However, they argue that the current legal framework does not fully protect women, as many forms of online violence, including psychological abuse remain inadequately covered.

Thamidela Mythri Devi [2] explores the intersection of technology and criminality in India, highlighting the role of digital systems in accelerating cybercrimes. Devi emphasizes the need for legal frameworks and technological advancements to counteract increasing cyber threats as the use of computer systems in daily life. Prof. Saquib Ahmad Khan[3] outlines the rising incidence of cybercrime due to rapid technology adoption, including phishing and cyberbullying. Khan examines how the 2000 Information Technology Act serves as India's legal foundation to control cyber offenses and protect citizens from digital threats, highlighting the need for updated cybersecurity measures. Animesh Sarmahand et al. [4], discuss the necessity of "Cyber Law" to address the proliferation of cyber-related offenses, describing cyber law as integral to safeguarding online privacy and security. The paper emphasizes that without punishment, cybercriminals are unlikely to desist from illegal activities, indicating the need for stricter law enforcement. Krishna Kumar et al. [5] discussed that Cybercrime is causing increasing harm to individuals, finances, and even national security. To address this, the Government of India introduced the IT Act, 2000, to regulate cyberspace. However, legal measures alone are not enough—public awareness is also essential. This paper aims to spread knowledge about cybercrime and related laws to help build a safer digital environment. Nikita Goyal [6], This study examines common cybercrimes, their impact on society, and suggests preventive measures while addressing future challenges. Sharma, D., Jain, et al. [12], focuses on using deep learning technique Residual Neural Networks (ResNets) to analyze cybercrime datasets in India. The study aims to understand cybercrime trends, patterns, and characteristics by leveraging a comprehensive dataset from official sources.

3. DATA COLLECTION

Data on cybercrime incidents were collected from credible government sources, including the Press Information Bureau (PIB), the National Crime Records Bureau (NCRB), and sansad.gov.in, as well as from academic publications. This dataset, covering various states, offers an overview of cybercrime trends from 2013 to 2022. Specifically, the National Crime Records Bureau (NCRB) provided detailed statistics on cybercrime incidents by state and year, enabling a state-wise trend analysis. The Press Information Bureau (PIB) contributed official reports on cybercrime, including information on government interventions and policy updates. Additionally, Parliamentary Records from sansad.gov.in documented legislative discussions on cybercrime, highlighting policy considerations and areas requiring improvement

sr.no	states	2014	2015	2016	2017	2018	2019	2020	2021	2022
1	Andhra Pradesh	282	536	616	931	1207	1886	1899	1875	2341
2	Arunachal Pradesh	18	6	4	1	7	8	30	47	14
3	Assam	379	483	696	1120	2022	2231	3530	4846	1733
4	Bihar	114	242	309	433	374	1050	1512	1413	1621
5	Chhattisgarh	123	103	90	171	139	175	297	352	439
6	Goa	62	17	31	13	29	15	40	36	90
7	Gujarat	227	242	362	458	702	784	1283	1536	1417
8	Haryana	151	224	401	504	418	564	656	622	681
9	Himachal Pradesh	38	50	31	56	69	76	98	70	77
10	Jharkhand	93	180	259	720	930	1095	1204	953	967
11	Karnataka	1020	1447	1101	3174	5839	12020	10741	8136	12556
12	Kerala	450	290	283	320	340	307	426	626	773
13	Madhya Pradesh	289	231	258	490	740	602	699	589	826
14	Maharashtra	1879	2195	2380	3604	3511	4967	5496	5562	8249
15	Manipur	13	6	11	74	29	4	79	67	18
16	Meghalaya	60	56	39	39	74	89	142	107	75
17	Mizoram	22	8	1	10	6	8	13	30	1
18	Nagaland	0	0	2	0	2	2	8	8	4
19	Odisha	124	386	317	824	843	1485	1931	2037	1983
20	Punjab	226	149	102	176	239	243	378	551	697
21	Rajasthan	697	949	941	1304	1104	1762	1354	1504	1833
22	Sikkim	4	1	1	1	1	2	0	0	26
23	Tamil Nadu	172	142	144	228	295	385	782	1076	2082
24	Telangana	703	687	593	1209	1205	2691	5024	10303	15297
25	Tripura	5	13	8	7	20	20	34	24	30
26	Uttar Pradesh	1737	2208	2639	4971	6280	11416	11097	8829	10117
27	Uttarakhand	42	48	62	124	171	100	243	718	559
28	West Bengal	355	398	478	568	335	524	712	513	401
	total states	9322	13312	13697	22979	26931	44511	49708	52430	64907

Table 1: Dataset used for predictive model

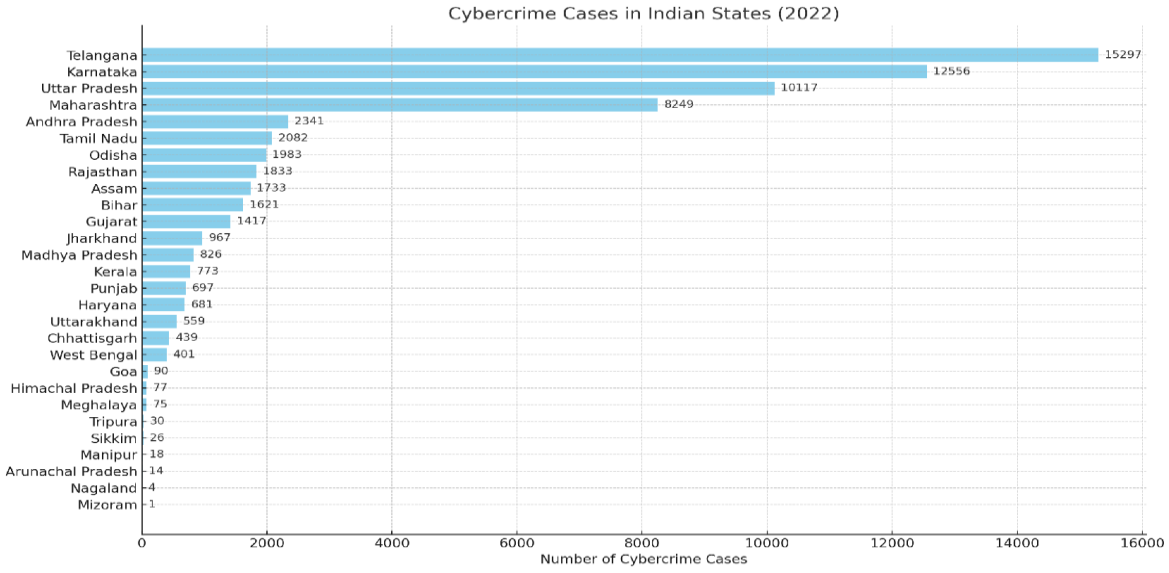

Fig 1: State-wise cybercrime cases in India

Data Preprocessing: Data cleaning and standardization were conducted to ensure uniformity and to handle missing or inconsistent entries. Preprocessing steps included normalization, formatting, and validation.

DATA ANALYSIS METHODS :

Exploratory Data Analysis (EDA): Visualizations like scatter plots helped us explore data trends, detect anomalies, and confirm a consistent increase in cybercrime rates over time.

Regression Analysis and Modeling: A linear regression model was applied to the preprocessed data to determine a trend line, which was validated and used to project future cybercrime counts.

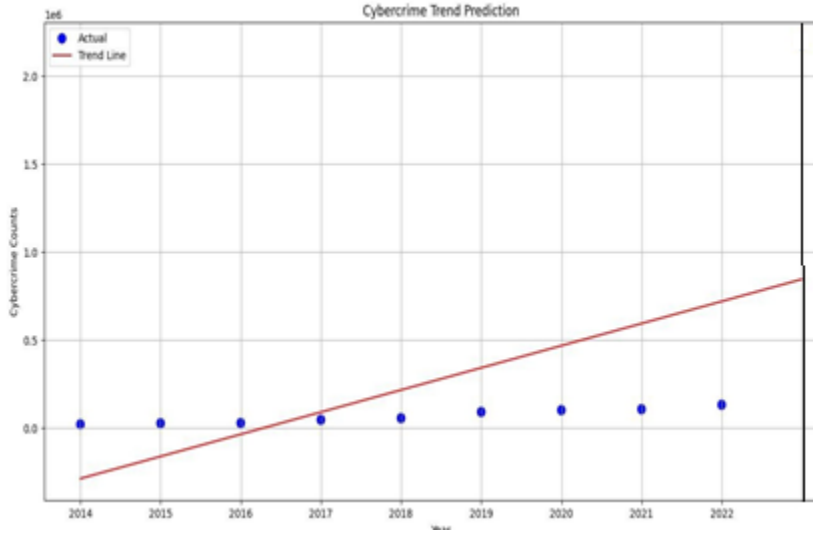

Fig 2: Predictive model using ML

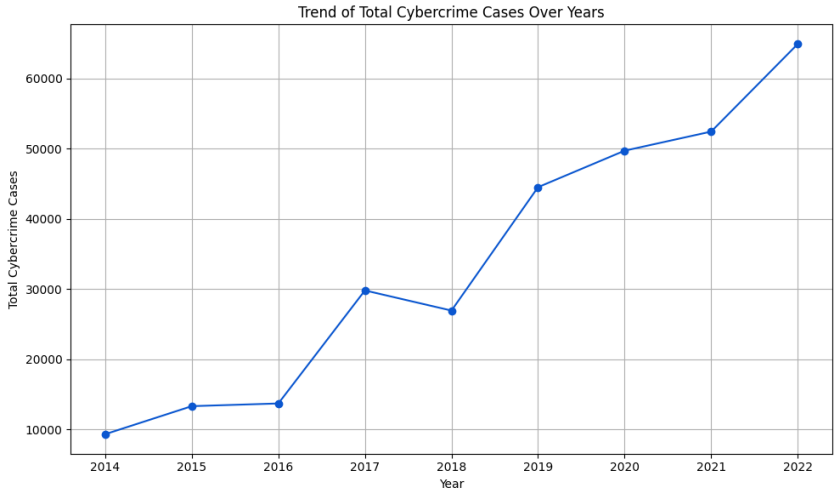

Fig 3: Data viusalization plot for trend of total cyber crime cases

Visualization and Interpretation:
The scatter plot and regression line were used to visualize the upward trend in cybercrime and predict future growth (Fig 2). The results are displayed as an annotated plot, showing historical data and future projections.

Trend Visualization: The scatter plot of yearly cybercrime counts shows a steady with increase, clear upward momentum over the last decade.

TREND PREDICTION USING ML MODEL:
The graph in Fig 2 represents a regression analysis conducted using Machine Learning (ML) in Python to study and predict the trend of cybercrime cases across Indian states from 2014 to 2022. The blue dots indicate the actual number of cybercrime incidents recorded each year, while the red trend line shows the predicted linear growth pattern derived from a regression model. The python program uses scikit-learn's LinearRegression model to predict the trend in cybercrimes and visualize it using Matplotlib. sklearn.linear_model.LinearRegression is the

machine learning model used for fitting a linear trend line to the data. This model can be used to forecast the number of cybercrimes from 2023 onwards, aiding policymakers and cybersecurity teams in planning preventive measures. The model likely uses linear regression to fit a line through the historical data and estimate future values based on this trend. Using this ML module, future cybercrime Count will be predicted.

4. RESULT AND DISCUSSIONS

The regression analysis and scatter plot visualizations reveal a substantial increase in cybercrime incidents from 2013 to 2022. This upward trend is closely associated with the rapid growth of internet usage, the expansion of digital services, and the widespread adoption of social media. Projections from the regression model for the years 2023 to 2026 suggest a continued rise in cybercrime, underscoring the urgent need for effective cybersecurity measures. The model's high accuracy in fitting historical data highlights the reliability of machine learning techniques in forecasting cybercrime rates. These predictive insights provide valuable guidance for policymakers and law enforcement agencies, enabling proactive planning and response. By identifying future risks and high-vulnerability regions, this analysis supports the formulation of targeted strategies to combat cybercrime more effectively. The analysis shows that cybercrime rates differ significantly across Indian states due to socio-economic and regional factors. States like Maharashtra, Karnataka, and Telangana with high digital adoption and major IT hubs report the highest incidents, as their large digital footprint attracts cybercriminals. Conversely, states with low digital literacy and limited cybersecurity resources also face rising cybercrime, driven by different challenges. This indicates that both digitally advanced and economically weaker states are vulnerable, though the nature and motives behind cybercrimes may vary.

Measures to Reduce Cybercrime

i. **Enhancing Digital Literacy**: Educating citizens about safe online practices is vital to reduce cybercrime. Initiatives in schools, colleges, and community centers play a key role in building awareness and resilience.

ii. **Strengthening Cyber Laws**: Cyber laws must be regularly updated to address new threats such as AI driven fraud and ransomware. Strict penalties should be enforced to act as a deterrent against cybercriminal activities.

iii. **Improving Cybersecurity Infrastructure**: Establishing advanced cybersecurity frameworks and Security Operations Centers and strong data protection systems is essential for early detection and prevention of cyber threats.

iv. **Conducting Public Awareness Campaigns**: Government bodies should lead campaigns to inform the public about online threats and safe internet usage, especially targeting vulnerable groups such as children, seniors, and rural users.

v. **Promoting Corporate Responsibility**: Organizations should adopt robust cybersecurity measures and conduct regular audits to protect user data. Businesses must also ensure continuous training and awareness among employees.

5. CONCLUSION

This study presents a comprehensive analysis of cybercrime trends in India. Using regression analysis and machine learning techniques, this study has projected that the upward trend in cybercrime is likely to continue, underscoring the need for immediate, preventive measures. The findings of this study underscore the urgent need for a holistic approach to combat cybercrime in India. Legislative reforms, public awareness initiatives, enhanced law enforcement capabilities, and community outreach must converge to form a proactive defense against cyber threats. India must invest in ongoing cybersecurity research, adopt global best practices, and foster a culture of cyber vigilance to stay ahead of evolving cyber risks. Through sustained, collaborative efforts across sectors, India can aspire to create a safer and more resilient digital environment, protecting its citizens and bolstering its digital economy.

REFERENCES

1) Apoorva Bhangla and Jahanvi Tuli, "A Study on Cyber Crime and its Legal Framework in India", IJLMH, ISSN 2581-53-69, Volume 4, Issue 2, Page 493 – 504

2) Thamidela Mythri Devi , *Cyber Crime in India: A Critical Study in Modern* Perspective, JETIR, ISSN 2349-5162,February 2022, Volume 9, Issue 2

3) Prof. Saquib Ahmad Khan - *Cyber Crime in India: An Empirical Study,* International Journal of Scientific and Engineering Research · May 2020, Volume 11(Issue 5):690-694

4) Animesh Sarmahand and Amlan Jyoti Baruah , "*A Brief Study on Cyber Crime and Cyber Laws in India",* International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, Volume: 04 Issue: 06 | June -2017

5) Krishna Kumar, Praveen Gupta, "A Study About Cyber Crimes And Cyber Law In India", JETIR May 2019, Volume 6, Issue 5, ISSN-2349-5162.

6) Shivani ReddyAnalytical Study on Cyber Crimes in India, CMR School of Legal Studies, Date Written: June 15, 2021

7) Nikita Goyal, Deepam Goyal, "Cybercrime in the society, Security Issues, Prevention and Challenges, RJET,2321-581X (Online), Vol 8, Issue 2, 2017.

8) McAfee & CSIS (2020). The Hidden Costs of Cybercrime, Center for Strategic and International Studies. Retrieved from https://csis.org

9) Gupta, A., Singh, B., & Sharma, C. (2022). "Cybercrime Trends in India: A Comprehensive Analysis." International Journal of Cybersecurity and Digital Forensics, 10(3), 45-58.

10) Patel, R., Desai, S., & Shah, P. (2020). "Deep Learning Techniques for Cybercrime Detection: A Survey." In Proceedings of the IEEE International Conference on Cybersecurity and Privacy (ICCSP), New Delhi, India, pp. 112-125.

11) Sharma, D., Jain, M., & Gupta, S. (2021). "A Novel Approach to Cybercrime Dataset Analysis Using Deep Learning." Journal of Computer Security and Privacy, 28(4), 589-602.

12) Websites: https://rbi.org.inNational Crime Records Bureau (NCRB), Crime in India Report 2013-2022. Ministry of Home Affairs, Government of India. Retrieved from (https://ncrb.gov.in)

13) Website: https://www.statista.com/search (2022)

14) Website: Press Information Bureau (PIB) https://pib.gov.in