Manisha M. More, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Analyzing Decadal Cyber Crime Trends in India: A Data Driven Approach Using Clustering and Predictive Modeling

Manisha M. More[1], Mayuri More[2], Varsha Gidde[3], Sheetal Sandeep Patil[4]
[1,2,3]School of Computer Studies, Sri Balaji University, Pune, Maharashtra, India
[4]MITCOM, MIT ADT University, Pune, Maharashtra, Inida

Abstract- The proposed study explores the NCRB i.e. National Crime Record Bureau data for the period of 2012 To 2022 and assessed the trends cybercrime across the states, the metropolitan cities, and the sectors such as IT and IPC. The analysis is conducted using Power BI for data visualization and statistical methods in Python looking for any trends or correlations. In addition, clustering techniques were utilized in grouping states and sector data observation into three groups, i.e., low risk, medium risk and high risk. The analysis showed that most states were classified as low risk and only a few instead showed an extremely high intensity of Cyber Crime suggesting a degree of geographical concentration for Cyber Crime. The authors also used two types of machine learning model (Linear Regression and Random Forest) to identify any trends based on time. However, both machine learning models displayed limited success in their predictability meaning that Cyber Crime occurrences cannot simply be explained by reference to time alone. The overall findings of this study then align with the conclusion that while some Cyber Crime occurrences are similar and concentrated to a particular location or point in time, they are also dissimilar based upon when they occur and where they occur. As a whole the study illustrates a methodical analytical technique of using both data visualization through statistical methods and explorative AI methods to provide better insight into Cyber Crime trends.

Keywords: Cyber Crimes Trends, Cyber Crime Analysis, Clustering analysis cybercrime.

I. INTRODUCTION

Cybercrime has continued to grow in India over the period of 2012 to 2022, with a notable variation of incidents occurring between the states (by metropolitan area) and by legal classification under either the IT Act or the IPC. Upon analysis of the NCRB data, it is apparent that incidents of cybercrime do not naturally occur on a constant basis; there are many regions where higher than average numbers of cybercrime occur, therefore requiring a structured analysis to identify regional and sector based patterns of cybercrime. In this analysis, the data was analyzed using Power BI for visualizing and processing the data statistically and analytically through Python.

The data was analyzed on the basis of state-wise distribution, state-level on a city-wise basis, and an overall sector-wise comparison between the IT and IPC crimes. Clustering techniques were applied to remove states and sectoral observations from risk-based classification, creating low-, medium-, and high-risk categories. The results from the clustering reveal that most states fall into the low-risk category; only a few states are accountable for a majority of the cybercrime incidents contributing to the total number of cybercrime incidents in India. Additionally, based on IT and IPC cybercrime clustering of sectors, it can also be seen that the area of where cybercrime occurrences varies from low- to high-based averages based on the geographic location and time basis on when the incidents occurred.

The majority of observations fall into low-risk based averages with only a few high-risk cases contributing a disproportionately larger number of cybercrime incidents. The research confirms that there is a concentration of cybercrime in certain geographical locations and times; however, that the modelling results show very little in predictive performance, in that there is no strong relationship between time and cybercrime trends when using these two different types of machine learning models (Linear Regression & Random Forest). It is likely that there

are multiple other factors (such as digital adoption, the existence of reporting channels or social economic conditions) that also influence the variations in cybercrime behaviours over time or by region.

Overall this research provides a data-driven approach to size, analyse and interpret cybercrime data through multiple means (i.e., graphics, statistical analysis, data clustering, and predictive modelling) to gain an understanding of how cybercrime behaves over time and between different locations in order to aid in the identification of at-risk locations and to assist in developing strategies for improving cyber-security.

II. RELATED LITERATURE WORK

The following table (Table 1) shows the extensive literature review showcase the existing research.

Table 1: Related Literature Review

Authors	Key Findings	Research Gap
Dr. R. D. Padmavathy et. al.[1]	Identified key cyber threats specific to the North East region of India, emphasizing the need for localized strategies.	Focuses only on the North East region, lacks quantitative data
Tiwari R. et.al.[2]	Highlighted the socio-economic impacts of cybercrime on scheduled tribes, suggesting tailored intervention policies.	Limited to scheduled tribes, lacks broader applicability
Sharma B. et. al.[3]	Discussed various cyber threats faced by children, calling for enhanced protective measures and awareness programs.	Focuses solely on children, lacks generalizability
Singh R. et. al.[4]	Analyzed cyber vulnerabilities in remote working environments during COVID-19, proposing targeted cybersecurity practices.	Context-specific to COVID-19 pandemic, may not apply to other scenarios
Parmar S. D. et. al.[5]	Provided a comprehensive framework for national cybersecurity, recommending multi-layered defense mechanisms.	Broad scope, lacks detailed focus on specific cyber threats
Kumar V. A. et. al.[6]	Explored cyber vulnerabilities in the power sector, advocating for robust cybersecurity infrastructure.	Sector-specific, lacks general cybercrime insights
Mangat H. S. et. al.[7]	Identified urban cybercrime trends, suggesting urban-specific cybersecurity strategies.	Focuses on urban areas, lacks insights into rural crime patterns
Goel S. et. al.[8]	Examined cyber threats in the banking sector, recommending improved security protocols and customer awareness.	Banking-specific, lacks broader applicability
Gupta R. et. al[9]	Compared cyber threats across emerging economies, highlighting common vulnerabilities and suggesting regional cooperation.	Broad comparative study, lacks in- depth analysis of individual economies
Ghate S.et. al.[10]	Provided an overview of global cybersecurity trends, calling for international collaboration to address challenges.	Broad overview, lacks specific focus on actionable strategies
Khalid M. et. al.[11]	Analyzed domestic factors affecting national security, emphasizing the importance of integrating cybersecurity in national policies.	General analysis, lacks specific focus on cybercrime
Kaur M. et. al.[12]	Evaluated government initiatives against cyberbullying in higher education, suggesting improvements in policy implementation.	Focuses on higher education institutions, lacks broader applicability
Alam N. et. al.[13]	Assessed cybercrime awareness levels among students and teachers, recommending enhanced cybersecurity education.	Educational context-specific, lacks generalizability
Agrawal S. et. al.[14]	Investigated cybercrime in the banking sector, proposing measures to strengthen banking cybersecurity.	Banking-specific, lacks broader applicability

Morya K. K. et. al.[15]	Analyzed cybersecurity threats to IT and OT systems, advocating for integrated strategies to protect critical infrastructure.	Sector-specific, lacks general cybercrime insights
Jiyauddin M. et. al.[16]	Explored cybercrime under Indian criminal law, highlighting gaps in enforcement and suggesting legal reforms.	Legal framework-focused, lacks practical enforcement insights
Malik J. K. et. al.[17]	Reviewed the cybercrime policy landscape, recommending practical strategies for policy implementation.	Policy-focused, lacks practical implementation strategies
Singh M. et. al.[18]	Analyzed Indian cyber law and cybercrime, suggesting improvements in legal enforcement mechanisms.	Legal framework-focused, lacks practical enforcement insights
Sharma A. et. al.[19]	Reviewed cyber-attack techniques and patterns, recommending targeted countermeasures.	Broad review, lacks detailed focus on specific countermeasures
Iqbal J. et. al.[20]	Analyzed evolving cybercrime trends in India, calling for adaptive and proactive cybersecurity measures.	Broad analysis, lacks specific focus on actionable strategies
More M. M. et.al.[21]	Provided an overview of contemporary cybercrime trends, emphasizing the need for updated cybersecurity policies.	General overview, lacks specific focus on actionable strategies
More M. M. et.al.[22]	Investigated online banking vulnerabilities, recommending enhanced security measures for online transactions.	Banking-specific, lacks broader applicability

Current research tends to primarily look at descriptors of cyber crime trends, individual cyber crime categories, or cases within specific geographic areas. Most of the studies do not take a consolidated / combined analytical approach that uses visualization, statistical methods, grouping (clustering), and predictive models. Also, there is little / no mention of risk classification of States and industry-specific differences using a data-based approach. The current study delineates to improve upon the existing literature by using a combination of Power BI Visualization, statistics, clustering, and introductory machine learning models to demonstrate an overall structure for analyzing cyber crime patterns across state and time.

III. RESEARCH GAP AND SIGNIFICANCE OF THE STUDY

The comprehensive literature survey reveals that, lacks a thorough analysis of cybercrime trends at statewide, metropolitan city-wise, and the cybercrime sector levels. In addition, there is a gap in the correlation between the cybercrime trends across various sectors, and cybercrime heads are not mentioned in the existing studies. There is also a deficiency in the ML i.e. machine learning applications and the techniques for prediction of cybercrime trends and cybercrime heads. Researchers have addressed these crucial gaps in existing studies for developing effective strategies and policies for the policy makers. Our proposed study helps in aiding in resource allocation for law and enforcement and policymakers.

The proposed study helps to understand the patterns and relationship between the different types of cybercrimes and thorough correlation analysis will help to target the intervention. A greater degree of accuracy in predicting and forecasting future cybercrime trends will allow for better, more proactive prevention actions improving the cyber landscape of India The research will provide a thorough analysis of nationally collected data on cybercrime to identify pattern's in multiple regions and sectors throughout India incorporating advanced machine learning technologies to make well supported predictions and provide actionable insights for varying regions and industries.

IV. RESEARCH QUESTIONS

The goal of our study is to obtain a comprehensive understanding of and address the rapidly changing

landscape of cybercrime across India through an assessment of patterns, regional differences and sector vulnerabilities. The following research questions will be the basis for guiding our study: What are some examples of the evolution of state wise cybercrime trends within India for the period of 2012 to 2022 and what types of regional differences can be determined?. What is the trend observed in cybercrime incidents across major metropolitan cities in India from 2012 to 2022, and which cities exhibit strong positive correlations?. How do cyber threats vary across different sectors (IT and IPC) in India, and what sector-specific vulnerabilities are most prevalent?. What is the trend observed in IT and IPC cybercrime cases in India from 2013 to 2022, and which states show a positive correlation with this trend?. How do cybercrime patterns vary across states and sectors, and what insights can be delivered to identify the high risk areas.

V. OBJECTIVES OF THE RESEARCH

To ascertain the research questions following are the objectives, To investigate the patterns of cybercrime across different states in India from 2012 to 2022. To evaluate the increase in cybercrime incidents within major metropolitan areas of India. To analyze the growth trends in IT-related and IPC (Indian Penal Code) cybercrime sectors. To examine the rise in specific categories of cybercrime on a case-by-case basis. To predict future trends in cybercrime using machine learning techniques.

VI. HYPOTHESIS OF RESEARCH

H_0: There is no significant increase in the Cyber Crimes incidence in India from 2012 to 2022 . H_0: There is no significant difference between cyber crimes registered under IT Act and under IPC in India.

VII. METHODOLOGY USED

A. Data Collection and Analysis

The dataset for the current research study is based on the NCRB (National Crime Record Bureau) of India that includes data collected between the years 2012 to 2022. The dataset consists of all variables that are pertinent to the current research, such as when (i.e. years) and where (i.e. states or union territories) the cyber-crimes occurred, which metropolitan (i.e. urban) areas in India suffered from cyber-crime, what types of cyber-crime occurred, and the number of reported cyber-crimes (both under the IT Act and IPC). The study's data is entirely from secondary (i.e. published or otherwise generated) sources. The methodology for conducting this research has been divided into two phases. The first phase of the research was carried out with the aide of Power BI, a business analytics service and visualization tool, to clean, prepare, and plot (i.e. visualize) the data in order to identify trends and patterns in the cyber-crime data. Many different types of graphs were used to demonstrate the trends of the data across time/years, across locations/states, or across sectors; such as bar charts, line charts, and pie charts.

B. Testing of Hypothesis

The second phase of the research (hypothesis testing) was completed using advanced statistical techniques combined with the Python programming language. The Chi-Square test was applied to check the incidence of cybercrimes over one decades across all the states and union territories. The ANOVA test was applied on the IPC and IT Act incidents over the selected period to find the significant relationship between these two sectors.

C. ML Models

KMeans Clustering were employed to make the clusters of States based on the levels of the Risk of the Cyber Crimes on total Cyber Crimes State wise and Sectors wise(IT/IPC). The Regression and Random Forest were employed on the Cyber Crimes Data Machine Learning Methods (including Linear Regression) used to Analyze Trends in Cyber-Crime Along With Statistical Background. Linear Regression Method Used as Base Model to Determine Potential Relationship Between Year and Growth of Cyber Crime. Random Forest Method Applied to Determine the Potential For Non-Linear Relationships and Provide More Detailed Analysis of Cyber-Crime Data.Through the use of Training Set and Testing Set (70% & 30%) Used to Validate Models. Performance of the Models were Measured via MAE, RMSE, & R^2 and Output Visualized in Power BI and/or Python.

The data used in this research article comes from the NCRB dataset and consists solely of absolute measures of cybercrime that have been reported. While normalization using other factors such as population or internet penetration may provide greater insights into cybercrime trends, these data were not utilized in this study due to limitations. Therefore, there is an inherent limitation of not accounting for digital adoption and the evolving nature of reporting practices in the time since the start of this research project, which may also impact how the results should be interpreted.

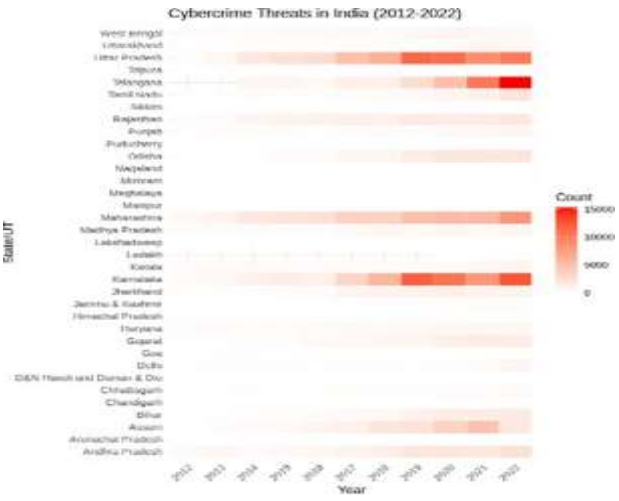
VIII. DATA ANALYSIS, RESULT AND DISCUSSION

This section contains an analysis of the cybercrime data using data visualization, statistical techniques, clustering techniques and predictive modeling approaches. It discusses the results by breaking them down by state/state models, by sector/sector models and by data model-based approaches to illustrate all of the key trends and variations. The results provide a structured view of the trends of cybercrime among regions and change in time.

A. State wise and Union Territories Wise Cybercrimes in India

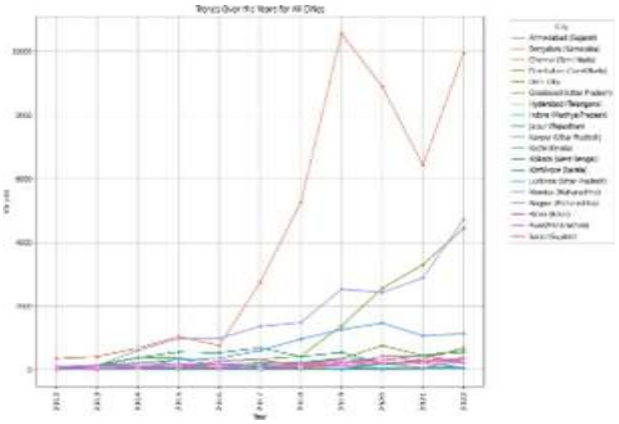
The correlation matrix i.e. heat map shows the cybercrimes incidents in India from 2012 to 2022. It highlights a significant rise in the several states and Union territories. It is notably highlighting Uttar Pradesh, Maharashtra and Karnataka with the highest peak in Uttar Pradesh in 2022. The growth of Maharashtra since 2019 has been consistent; however, the growth of incidents in Karnataka has been increasing since 2017. The northeastern states of Nagaland, Mizoram, Arunachal Pradesh have had very little variation. Daman and Diu have also stayed below average in terms of the number of cybercrime incidents that have occurred over the last five years.

The graphs show states with darker shades marked high on the scale of the number of incidents of cybercrime reported, whereas Uttar Pradesh, Maharashtra, and Karnataka will be marked with a dark red shade; the severity of the rise in the number of incidents of cybercrime is alarming when you look at the states with the lightest shade; they will be those located in the northeastern quadrant of India. (see Graph 1).



Graph 1. Cybercrimes in India States/UTs

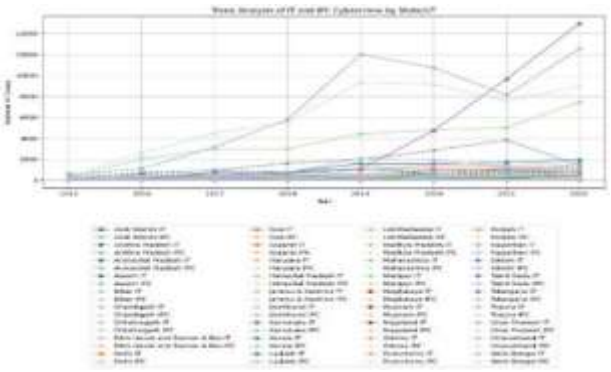
B. Cybercrimes Status Metropolitan City Wise



Graph 2. Cybercrimes Metropolitan City Wise 2012-2022)

The line graph shows a dramatic rise in the number of reported cyber crimes across most of the major cities in India from 2012 through 2022. The cities with the largest reported increases were Bengaluru, Hyderabad, Mumbai and Lucknow; all were significantly higher than other cities from 2017 onward (Graph 2). Conversely, there appeared to be little growth or consistent trends in the cities of Kochi and Kozhikode, indicating that these regions have implemented more effective measures to combat cyber crime or that their crime rates have remained consistent over time.

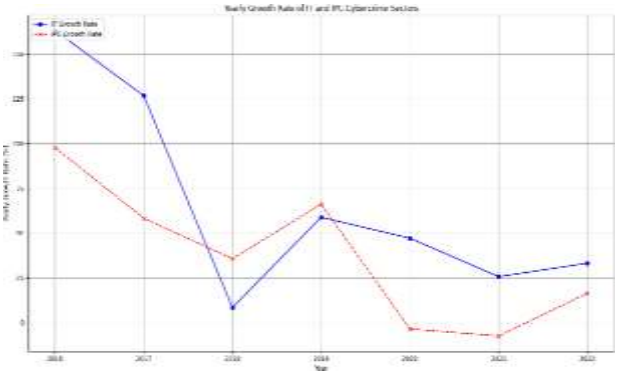
C. Sector Wise Cybercrime in India



Graph 3. Sector Wise Trend Analysis 2013-2022

In total, the graph displays the continued rise in the number of Information Technology-based crimes (IT) and crimes under the Indian Penal Code (IPC) in India from 2013 to 2016 to an increasing annual average between the two periods (through 2022). Some of the states with some of the highest correlations to increasing cyber crime rates are Bihar, Maharashtra, and West Bengal; while other states such as A&N Islands and Himachal Pradesh exhibit little or no correlation at all. This suggests that there is an immediate need for improving cyber security and knowledge across the country. (see Graph 3).

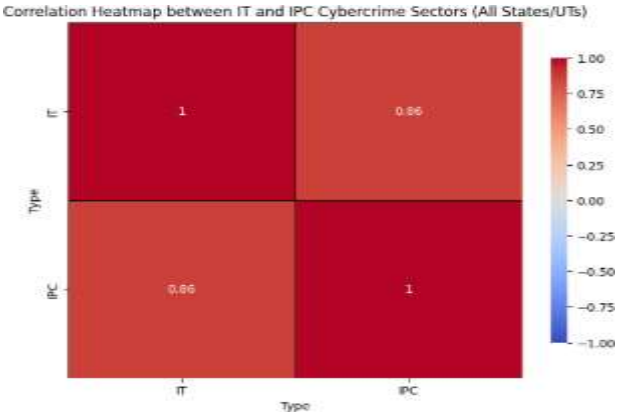
D. Yearly Growth of Cybercrime in India



Graph 4. Sector Wise Yearly Growth of Cyber Crimes in India

Graph 4 shows annual growth rates for the IT cybercrime and IPC cybercrime sectors in India for the years 2016 through 2022. The IT cybercrime sector had an extremely high growth rate of over 160% in 2016 but quickly dropped to a growth rate of less than 0% in 2018 before rebounding to approximately 60% in 2019 and then remaining stable with a growth rate of approximately 25% in 2021. (See graph 4)

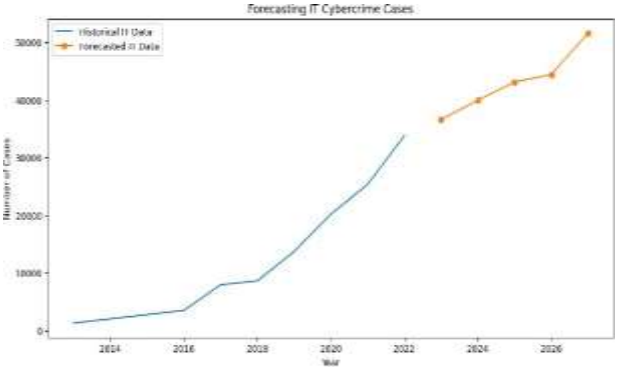
E. Correlation Analysis



Graph 5. Correlation Analysis of Sector Wise Cybercrimes in India 2013-2022

There are many strong positive correlations found in a correlation analysis of all the different types of cybercrime. These correlations suggest that many types of cybercrime may have similar increases over time (Graph 5). However, this correlation could also be impacted by other factors like increases in the number of people using the internet, the level of digital adoption, and increasing numbers of incidents being reported. The overall increases in cybercrime activities across all different types are more likely to be reflected in the correlations than any single type of crime correlated to one another. (See Graph 5)

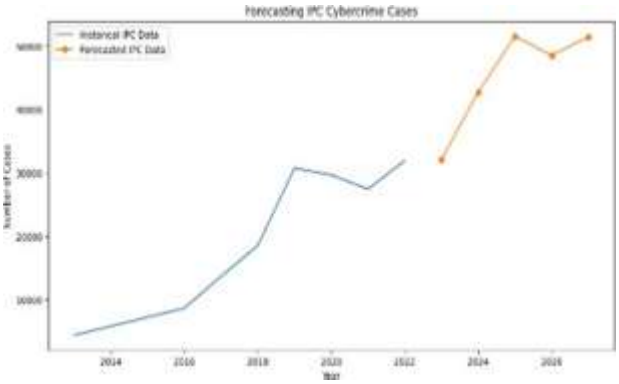
F. ML Model for Cybercrimes Prediction as Per Sectors



Graph 6. Sector Wise Cybercrimes in India 2013-2022 (IT ACT)

The graph 6 illustrates the rise in IT cybercrime cases in India from 2013 to 2027, with historical data shown by the blue line and forecasted data by the orange line. The graph indicates a significant upward trend, with an anticipated rapid escalation in the number of cases between 2025 and 2027, therefore necessitating increased vigilance and security measures to combat rising levels of cybercrime. (See Graph 6)

F.I. IT Sector

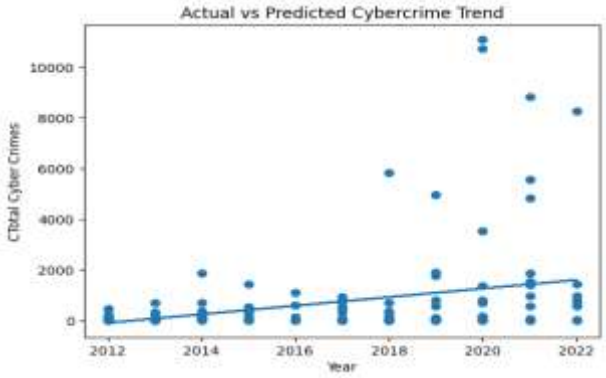


Graph 7. IPC Sector Wise Cybercrimes in India 2013-2022

Graph 7 details historical and projected trends in IPC (Indian Penal Code) cybercrime occurrences from 2013 to 2012, where historical data reveal a consistent increase and future data forecasted based upon these findings will show an even higher rate of increase. This upward movement represents serious difficulties that will be presented to the criminal justice system; these will likely be exacerbated by increasing reliance upon technology, along with continuously evolving tactics employed by cyber criminals. Therefore, it is imperative that there are continued improvements to existing cybersecurity policies and procedures.

F.II. Machine Learning Based Trend Analysis

Machine-learning techniques were applied on the dataset to further analyse the predictive capabilities of cybercrime trends. The first model used was linear regression as a baseline for determining the linear relationship of cybercrime growth and year. A second model, Random forests, was also used to determine if non-linear trends could be captured by this data.



Graph 8. Actual Vs Predicted Cyber Crime Trends

The models were created and tested with a 70/30 split of data(Graph 8). Following the completion of the models, predictions were made using the linear regression model and subsequently were compared with actual amounts of cybercrime to help determine its level of fitness. Besides using Linear Regression we wanted to see if there were any non-linear patterns in the data for cyber crime. To do this we used a Random Forest model which is considered an ensemble learning method, allowing for more complex relationships within the data to be found. As such this will allow us to see how Random Forest compares to a simpler linear model. (See graph 8)

F.III. Model Evaluation Metrix

Using standard evaluation metrics (Mean Absolute Error, Root Mean Squared Error, and R^2 Score), as shown Table 2 we evaluated the performance of both models.

Table 2. Evaluation Metrix Comparision

Model	MAE	RMSE	R^2 Score
Linear Regression	964.88	1890.65	0.1
Random Forest	967.73	1943.51	0.05

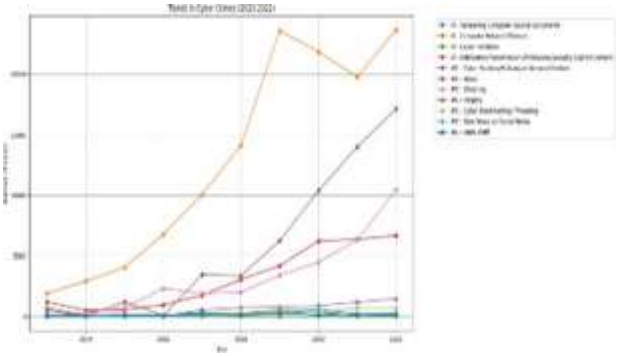
F.IV. Result Interpretation

Based on the results, both of the models that were evaluated have limited ability to predict future events. The Linear Regression Model is limited by a low R^2 score of 0.10, indicating that only a small portion of the variation in cybercrime trends is modelled by using time (i.e., year) as an only input feature. Despite being typically a strong model for capturing complex patterns, the Random Forest model also fails to demonstrate strong forecasting

ability in this case, yielding an even lower R^2 score of 0.05. The data shows that there is not enough input features in the dataset to accurately estimate the complexity of cybercrime behavior.

The graphical display of predicted values and actual values shows this clearly; the difference between the predicted value and the actual value shows that both models cannot accurately predict what will happen in the real world when it comes to predicting how likely an individual is to commit a crime against another individual via the internet. Based on these findings, it is clear that there are many different types of influences that will have an impact on the trends of future cybercrime; for example, these influences include an increasing number of users of Internet technology, a growing population, a shift in how crimes are reported, etc. There is some value to using machine-learning methods to examine possible trends in cyber-criminal activity, but based on the small number of input features, indicators, variables, etc., there are limited opportunities to utilize machine-learning methods for forecasting and/or predicting trends in cybercrime.

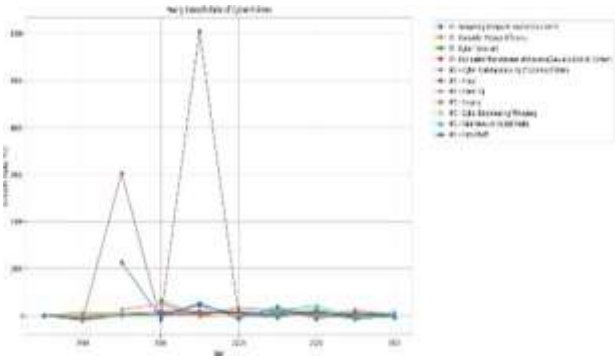
G. Head Wise Cybercrimes in India



Graph 9. Head Wise Trend of Cybercrimes in India 2013-2022

Cybercrime data shows a consistent increase within specific categories from 2013 through 2022 in India. For instance, there are substantial gains in the number of computer-related offenses, fraud, and cyber blackmailing occurrences. These trends illustrate the increased need for advanced security practices to prevent future attacks as trends can change quickly over time(Graph 9).

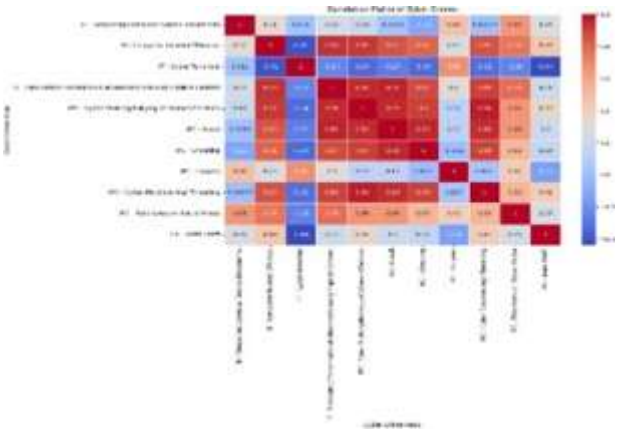
H. Head Wise Yearly Growth of Cybercrimes in India



Graph 10. Head Wise Yearly Growth Rate 2023-2022

As per the Graph 10 data for cybercrime shows that the total number of crimes increased from 2014 - 2016; the largest increases were seen with computer-related crimes, fraud, and data theft. Despite a decrease of most of the crime's rates from 2017 onward, cyber terrorism has seen consistent growth along with some examples of long-standing issues such as blackmailing and creating false news do not have a declining trend; therefore, ongoing prevention and response are necessary.

I. Head Wise Correlation Analysis



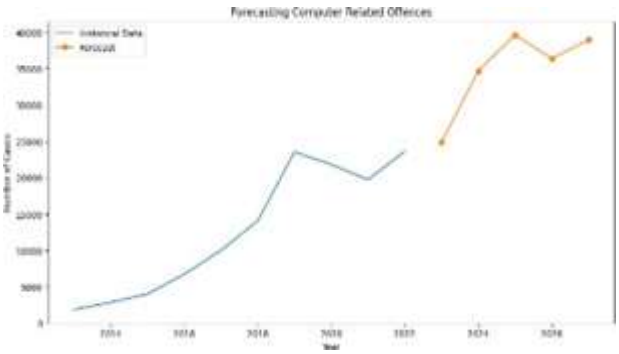
Graph 11. Head Wise Correlation Analysis of Cybercrimes in India 2023-2022

Graph 11 displays the correlation matrix showing an overwhelmingly strong positive correlation for many of the different types of cyber-crimes. In other words, numerous cyber-crimes frequently happen together. For instance, there is a strong positive correlation

between the dissemination and/or transmission of explicit content and fraud and between fraud and the occurrence of fraud computer based crime, which indicated that an underlying factor or method may be shared between these two different criminal patterns. There is also a strong positive correlation between the occurrence of fake news on social media and the occurrence of computer-related crimes and the dissemination and/or transmission of explicit content.

This suggests that there is a potential relationship between these two criminal activities, possibly supporting the argument that there is a similarity between and possibility of misinformation on social media and other criminal activity committed via the Internet. Meanwhile, there are some crimes with a negative correlation, such as the occurrence of data theft with cyber terrorism and forgery with data theft. This indicates that there are many different types of mechanisms through which these crimes occur. The information in the analysis identifies that there are many interrelated cyber-crimes and indicates the necessity of a shared approach to cybersecurity. The data collected from 2013 through to 2022 demonstrates a consistent pattern of these types of correlations therefore establishing a strong basis for the need to continuously evolve and adapt cybersecurity measures to effectively combat changing threats.

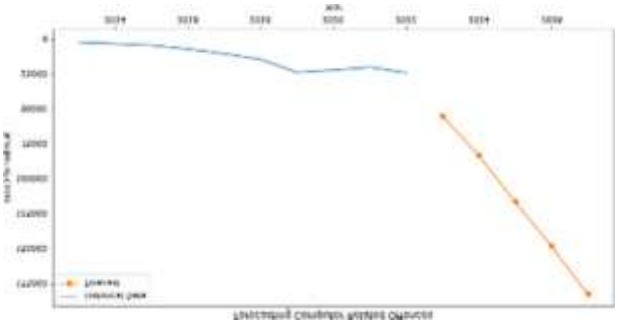
J. Forecasting Related offences



Graph 12. Forecasting of Cyber Crimes in India 2023-2022

Graph 12 shows the rise of computer-related crimes over the last ten years in India. There is a large synergy between the year and the number of cases; higher years typically mean more cases, and lower years typically mean fewer cases. Based on these trends, graph 12 makes predictions about the forecast for 2023-2027. It forecasts a very volatile (unpredictable) future for cybercrime; it does not propose a clear upward or downward trajectory for cybercrime over the forecast period. Although the industry can expect continued volatility in the patterns of computer-related crimes as technology, business, and culture continue to evolve, the industry must remain vigilant in terms of taking proactive measures to address the changing threat environment.

K. Model Valuation



Graph 13. Cybercrimes in India Head wise 2023-2022

Graph 13 illustrates a marked increase in computer-related offences in India since 2013. It clearly shows that there has been a dramatic increase in the number of cybercriminal acts throughout the country. From 2013-2016 there was a steady increase in the number of cases reported until a minor drop occurred in 2015. The drop in 2015 could be attributable to the fact that there had been an increase in digital infrastructure, and therefore it is likely that it contributed to the upward trend in computer-related criminal offences. The rise in the number of computer-based crimes from 2016 to 2018 could be mostly due to hackers taking advantage of the new technologies and the associated vulnerabilities with those technologies; therefore, there is a very strong relationship between technology adoption and occupational crime rates over the last ten years.

This will likely continue to happen through the forecast period, and the forecasting model indicates

a sustained increase in the number of cybercriminals over the next five years. Therefore, it is important to recognise that the number of reported cases may only be a fraction of the total number of crimes committed in India. The actual number of crimes committed could be significantly higher.

L. Hypothesis Testing

L.I. Testing First Hypothesis

There is no significant increase in the incidence of cybercrimes in India from 2012 to 2022 compared to other regions, the Chi-Square test was applied. This test was chosen to assess whether there were significant differences in cybercrime rates over the years in India relative to other regions. Table 3 presents the results of this test.

Table 3. Hypothesis First

Metric	Value
Chi-Square Statistic	48,062.10
P-value	0
Degrees of Freedom	350
Alpha	0.05
Decision	Reject H_0. There is a significant increase in cybercrimes in India from 2012 to 2022.

.

L.I.I. Observations

Chi-Square Statistic: 48,062.10. P-value: 0.0. Degrees of Freedom: 350 Alpha Level: 0.05

L.I.II. Interpretation

The chi-square test results reveal a chi-square statistic of 48,062.10, which is substantially high, and a p- value of 0.0. Given that the p-value is well below the significance threshold of 0.05, the results are statistically significant. This significance indicates that there is a substantial and statistically significant increase in the incidence of cybercrimes in India from 2012 to 2022. Hence we reject the null hypothesis, which was there is no significant increase in cybercrimes over the selected period. The hypothesis test findings underscore a notable trend in the rise in cybercrimes, and it suggests that the rise in the reported cases is not only due to the random variation but also reflects a genuine increase. So the test results highlight the need for enhanced cyber security measures and strict policies to address the growing cyber threats of cybercrimes effectively in India.

L.II. Testing Second Hypothesis

The proportion of cybercrimes prosecuted under the IT Act is significantly higher than those charged under the IPC in India from 2012 to 2022. To perform covariance analysis, we employed ANOVA test, to determine if there is any significant differences in the proportions cybercrime prosecuted under these two legal frameworks. The following table 4 displays the results of the ANOVA.

Table 4 Hypothesis Second

Metric	Value
F-statistic	1.856
P-value	0.174
Decision	Fail to reject the null hypothesis (H_0). There is no significant difference in the proportion cybercrime prosecuted under the IT Act and IPC in India from 2012 to 2022.

L.II.I. Observation:

The second hypothesis was tested by employing the ANOVA, compared the proportions of cybercrimes presented under IT Act and IPC in India from the selected period i.e. from 2012 to 2022. The ANOVA test result shows the F-statistic of 1.856 and the calculated p-value is 0.174

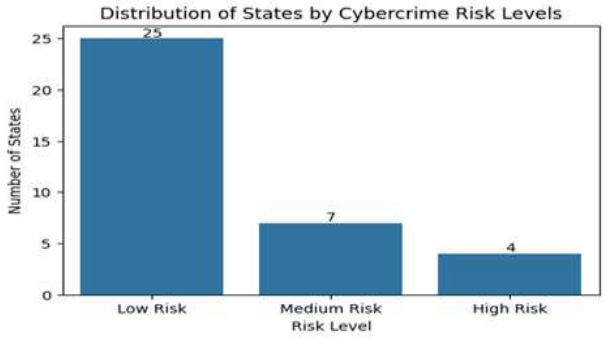
L.II.II. Interpretation:

We can see the p-value i.e. 0.174 is greater than the alpha value (significance value) i.e. 0.05, hence we fail to reject the null hypothesis. The test result indicates that there is no statistically significance difference in the proportions of cybercrimes prosecuted under the IT Act and IPC sections in a given period. The test evidence does not support to claim that the proportion of cybercrimes prosecuted under IT Act is significantly higher than those the IPC sections in India in during the 2012 – 2022.

M. Clustering Analysis cybercrime Risk Levels

A K-Means clustering technique was used for more detailed research into the distribution of cybercrime within each state (i.e., The United States). K-Means

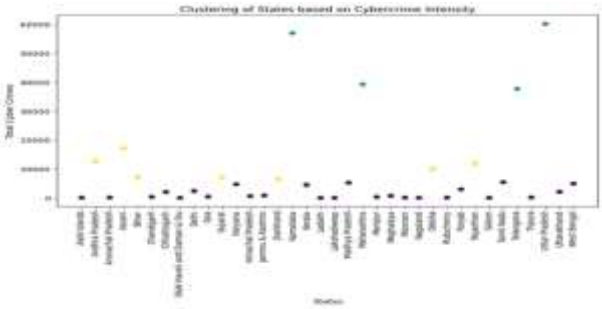
classified the states by their level of risk (i.e., Low Risk, Medium Risk, High Risk) based on the total number of cybercrimes reported in each state.



Graph 14. Distribution of States by Cyber Crimes Risk Level

According to the clustering, most states (25 states) are classified as low risk(Graph 14). Therefore, many states have lower rates of reported cybercrime than others. Seven states fall into the medium-risk category, so they have a moderate number of reported cybercrimes occurring within each state. Only four states fall into the high-risk category and have a concentration of cybercrimes in a limited part of the country.There is an uneven distribution of cybercrime in India, and it is concentrated in a small number of states that can all contribute to increased levels of reported cybercrime due to digital adoption, urbanization, and reporting methods used by law enforcement in those high-risk states. The states are segmented based on risk using K-Means clustering, which is beneficial for law enforcement and policymakers when determining the most efficient use of available resources and what cybersecurity measures should be applied to which states.

Using K-Means clustering contributes to the analysis of the research and provides an AI-based classification of the amount of cybercrime across regions while going beyond just using descriptive statistics to explain the reason for the cybercrime that occurs in the individual states. States are visually clustered based on total cybercrime incident counts on this scatter plot: Each state is represented as an individual data point, with the risk category indicated by its color. The upper right quadrant contains the individual states which are spatially distinct from each other, as they each have high total cybercrime incident counts and create the high-risk cluster. Conversely, the states grouped in the lower left quadrant, indicating low total cybercrime incident counts and representing the low-risk cluster, contain most of the states plotted. A smaller subset of states cluster within the medium-risk range. Therefore, this analysis indicates that cybercrime is disproportionately concentrated in a small number of states, while the majority of the states experience comparatively low total cybercrime incidents. The strength of regional variation illustrated in the distribution of the states further supports this concept.



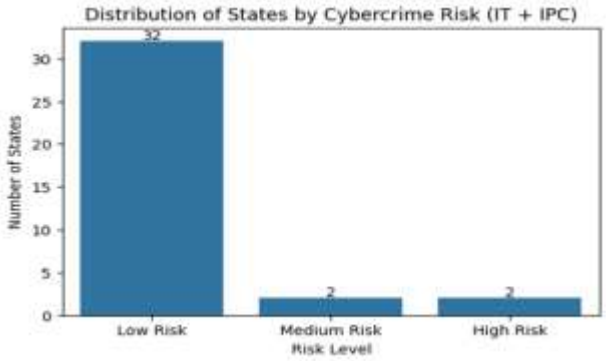
Graph 15. Clustering States Based on Cyber Crime Intensity

The combination of clustering quantitatively and visual representations of the distribution of the data in this scatter plot(Graph 15) enhances the interpretability of the results for decision-makers.

N. Sector Wise Clustering Analysis cybercrime(IPC/IT)

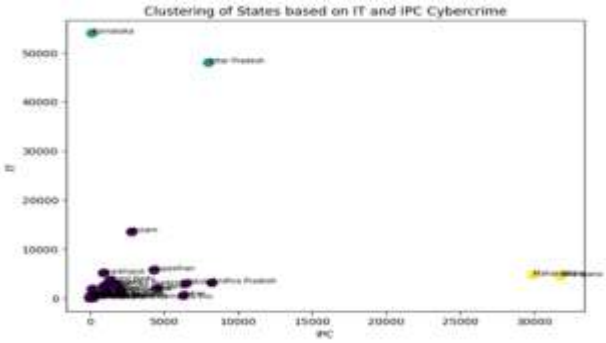
Clustering of the data for cybercrime (by combining IT + IPC crime) indicates clearly that the majority of states (32 states with low-risk) have not experienced significant cybercrime activity(Graph 16). A few states (2) fall into medium and high-risk levels of cybercrime: therefore, cybercrime is concentrated in geographic regions.

The difference between high-risk and low/med-risk states is easily seen using the scatter plot shown in Graph 17 (high-risk states have many more incidents of IT and IPC cybercrime than do other states).



Graph 16. Sector Wise Clustering Analysis

High-risk states contain distinct and easily recognizable values from other states and indicate the existence of intense cybercrime levels.



Graph. 17. Distribution of Cluster Wise Cyber Crime Analysis based on Sector Type(IPC/IT)

The results of this analysis show that the distribution of cybercrime is not evenly distributed (a few states are responsible for a large proportion of total cybercrime), and that many states remain within low-risk levels. By combining IT and IPC crime data together, clustering of the data will identify states that require assistance and a larger focus for cybersecurity efforts; therefore, clustering provides a more comprehensive picture of state-level vulnerabilities to cybercriminal activity.

This study identifies cybercrime trends that vary by state, city, and sector. The clustering analysis helps identify high/medium/low risk areas, while the Sector Analysis shows the differences in the types of crimes that are committed against information technology (IT) versus intellectual property (IPC). The predictive modeling results provide a complete view of historical cybercrime trends through a time-based perspective and highlight the limitations of using historical data to forecast future crimes.

IX. RECOMMENDATIONS AND FUTURE ENHANCEMENT

It is very crucial to address the rising cybercrime trends in India from 2012 to 2022 and enhance the basic cybersecurity infrastructure across all high-incidence states like Uttar Pradesh, Maharashtra, and Karnataka. The need to include the advanced technology and establishing the dedicated cybercrime units. These tailored strategies should be implemented with high training and public awareness campaigns in high-incident regions. The preventive measures in areas where low cyber incidents rate. Need to strengthen the public education on cybersecurity that improve interagency collaboration and update the legal frameworks, which will address evolving threats. In addition, fostering innovation through partnerships and maintaining a system for monitoring cybercrime trends that will support proactive measures and enhance overall security.

X. CONCLUSIONS

The current study examines and explores the trends associated with cybercrimes in India by using data obtained from the National Crime Records Bureau for a period of ten years (2012 to 2022). Our findings showcase how there is an uneven distribution of cybercrime, with a few states quickly becoming major contributors to all types cybercrime. Most states are considered low risk regarding their contributions; thus, showing evidence of inequities within the nation. When we analyze how many incidences are created per sector, be it IT or IPC related, we can determine that there are a wide range of incidents when comparing across time periods.

When we applied clustering methods to our data set(s) we found that we could clearly classifying the states into simple risk levels as well as identify sector specific observations. Lastly, the machine learning models used in our analysis provide only minimal

predictive capabilities, yet further emphasize that cybercrime trends cannot be explained solely based on time alone. Limitations exist for this research analysis. Absolute values were used for each state and no additional factors were considered to influence potential cybercrime; such as population size and digital adoption. Future research could help improve the accuracy of our predictions through better data and more sophisticated models. By employing a combination of statistical and cluster analysis with visualization techniques and introductory prediction modelling, we will be able to develop a far greater understanding of the trends associated with cybercriminal activity, thus allowing decision-makers to develop a detailed cyber security strategy.

REFERENCES

1. R. D. Padmavathy, "Cybercrime in India: A Trend Analysis Specific to North East," Aayushi International Interdisciplinary Research Journal (AIIRJ), vol. 5, no. 4, pp. 88–96, Apr. 2018.
2. R. Tiwari and S. N. Dixit, "Trends and Patterns of Crime Against Scheduled Tribes in India," Transactions of the Institute of Indian Geographers, vol. 45, no. 1, pp. 41–52, 2023.
3. B. Sharma and G. Kataria, "Internet Ubiquity and Cybercrime Targeting Children in India," Journal of Positive School Psychology, vol. 6, no. 9, pp. 5103–5112, 2022.
4. R. Singh and M. K. Verma, "Work From Home and Associated Cyber Vulnerabilities in Metropolitan Cities," vol. 10, no. 5(I), Jun. 2023.
5. S. D. Parmar, "Cybersecurity in India: An Evolving Concern for National Security," The Journal of Intelligence and Cyber Security, vol. 1, no. 1, 2018.
6. V. A. Kumar, K. K. Pandey, and D. K. Punia, "Cyber Security Threats in the Power Sector: Need for a Domain-Specific Regulatory Framework in India," Energy Policy, vol. 65, pp. 126–133, 2014.
7. H. S. Mangat and L. S. Gill, "Magnitude and Spatial Dimensions of Criminality in Urban India: Insights from Million Plus Cities," in Massive Urbanisation: Town Planning, Pedagogy and Research, Institute for Spatial Planning and Environment Research, Jul. 2021, pp. 217–232. [Online]. Available: https://www.researchgate.net/publication/353119842
8. S. Goel, "Cyber-Crime: A Growing Threat to Indian Banking Sector," International Journal of Science Technology and Management, vol. 5, no. 12, pp. 552–559, Dec. 2016.
9. R. Gupta and S. P. Agarwal, "A Comparative Study of Cyber Threats in Emerging Economies," Globus: An International Journal of Management & IT, vol. 8, no. 2, pp. 24–28, Jun. 2017. [Online]. Available: https://globusjournal.com/wp-content/uploads/2018/07/826Ruchika.pdf
10. S. Ghate and P. K. Agrawal, "A Literature Review on Cyber Security in Indian Context," Journal of Computer & Information Technology, vol. 8, no. 5, pp. 30–36, Oct. 2017.
11. M. Khalid, "Emerging Challenges to India's National Security: A Domestic Dimension," Scholars Journal of Arts, Humanities and Social Sciences, vol. 9, no. 11, pp. 600–603, Nov. 2021. [Online]. Available: https://saspublishers.com/media/articles/SJAHSS_911_600-603.pdf.
12. M. Kaur and M. Saini, "Indian Government Initiatives on Cyberbullying: A Case Study on Cyberbullying in Indian Higher Education Institutions," Education and Information Technologies, vol. 28, no. 1, pp. 581–615, Jan. 2023, doi: 10.1007/s10639-022-11168-4.
13. N. Alam, O. Qamar, and A. Husain, "Cybercrime Awareness, Vulnerability and Measures Among Higher Education Students and Teachers: An Exploratory Study of India," American International Journal of Research in Humanities, Arts and Social Sciences, vol. 24, no. 1, pp. 108–112, Sep.–Nov. 2018.
14. S. Agrawal, "Cyber Crime in Banking Sector," vol. 3, May 2016.
15. K. K. Morya and M. Singh, "Study of Latest Cybersecurity Threats to IT/OT and their Impact on e-Governance in India," vol. 11, no. 2, pp. 939–947, 2020. [Online]. Available: researchtrend.net.
16. M. Jiyauddin and S. Banerjee, "An Examination of the Concept of Cybercrimes Under Indian Criminal Law with Particular Reference to the Information Technology Act," International

Journal of Advances in Engineering & Management (IJAEM), vol. 6, no. 4, pp. 566–571, Apr. 2024.
17. J. K. Malik and S. Choudhary, "Policy Considerations in India Against Cyber Crime," vol. 9, no. 12(A), pp. 29811–29814, Dec. 2018.
18. M. Singh, J. A. Husain, and N. K. Vishwas, "A Comprehensive Study of Cyber Law and Cyber Crimes," vol. 3, no. 2, Feb. 2014.
19. A. Sharma, A. Tyagi, and M. Bhardwaj, "Analysis of Techniques and Attacking Pattern in Cyber Security Approach: A Survey," International Journal of Health Sciences, vol. 6, no. S2, pp. 13779–13798, 2022.
20. J. Iqbal and B. M. Beigh, "Cybercrime in India: Trends and Challenges," International Journal of Innovations & Advancement in Computer Science, vol. 6, no. 12, pp. 187–195, Dec. 2017.
21. M. M. More and K. M. Nalawade, "Cyber Crimes and Attacks: The Current Scenario," in Proceedings of the 1st National Conference Organized by NESGOI, Pune, India, 2014.
22. M. M. More, M. P. Jadhav, and K. M. Nalawade, "Online Banking and Cyber Attacks: The Current Scenario," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, no. 12, pp. 743–749, Dec. 2015.