Regional Crime Data Analysis and Insights Using Fb Prophet

Sonali Pakhmode[1], Vanshita Gaikwad[2], Bhargavee Talekar[3], Neha Golatkar[4], Vipin Jaiswal[5]

[1,2,3,4,5]Department of Information Technology, VPPCOE & VA, Sion, Mumbai, India.

Email Id: spakhmode@pvppcoe.ac.in[1], vu4f2021022@pvppcoe.ac.in[2], vu4f2021086@pvppcoe.ac.in[3], vu4f2021127@pvppcoe.ac.in[4], vu4f2021129@pvppcoe.ac.in[5]

Abstract

Crime is one of the most dominate and formidable aspect of the society. Crimes are committed daily, which has made the lives of the common people restive. Therefore, preventing the crime from occurring can be achieved through a systematic approach by recognizing and evaluating the criminal trends specific to a certain location. Data analysis can assist us to explore former criminal incidents and help us analyze the patterns and hidden correlations which can be used for prevention of crime. Our system aims to analyze crime-prone regions by using the location and time of the crime. Data visualization is a powerful tool for investigating crime rates across the country. Heat maps can be useful for determining the hotspots of crime. The proposed system analyzes the data using the FBprophet model to forecast crime rates in different regions. The Facebook Prophet library, maintained by Facebook, provides forecasting of time series data. This application uses past data and web-scraped data from various sites in order to get detailed studies of various crimes in multiple places. It also enables you to get data about crimes in many places across India by using the pin codes. Government organizations and police departments can effectively utilize it to monitor crime events within a locality.

Keywords: *Analyze; Correlation; Crime; FB prophet; Patterns.*

1. Introduction

With an increasing population and fast urbanization, crime trends in India are always shifting. There is a necessity for tools that can help us analyze and visualize the crime trend to ensure proper surveillance. The web application provides this functionality. This data visualization can also be used for research. In today's world, understanding and analyzing crime is necessary. There are now more options available to law enforcement, government organizations and other entities for monitoring criminal activity.[1] In addition to that government, the spotlight is now on emerging modern technologies to reduce crime. Crime data is a significant tool for us to examine in which year, how many kinds of crimes were committed along with regions.[2] Experts conducted numerous research to help them analyze crime patterns and connections in a specific location. Some of the examined hotspots have grown easier to classify criminal trends. This leads to assisting officials in handling situations faster.[3] In this study underscores the importance of graphical representation in compressing and analyzing data across diverse fields, facilitating insights like hotspot detection and tracking crime occurrences. Various techniques such as bar graphs, heatmaps, and line charts are utilized to visualize crime rates, with heatmaps particularly effective for hotspot detection. Analyzing time series data of crime rates in India, the study utilizes FB Prophet, a forecasting library that employs curve fitting within a Bayesian framework. This data-driven approach aids law enforcement and policymakers in strategic planning, resource allocation, and proactive crime prevention measures.

2. Literature Survey

The proposed study by N. Kanimozhi et al. present preprocessing methods such as data segregation and integer mapping for crime type labeling. Naive Bayes classification, applying extending supported features and employing both Multinomial and Gaussian NB to predict crime types, addressing

challenges associated with continuous target variables.[4] The authors Mary Shermila, et al. illustrate a workflow starting with homicide data extraction from Kaggle, preprocessing it into a crime database by categorizing qualitative data into binary form, and adding dummy columns. The model focuses on crime pattern detection and perpetrator description prediction, with two key phases as Crime Analysis, identifying unsolved crimes, and Prediction, estimating perpetrator descriptions.[5] S. Sathyadevan, et al. elaborate on collecting unstructured data from various websites, storing it in MongoDB, and employing Naive Bayes for classification, Apriori algorithm for pattern mining, and decision trees for prediction. They assess accuracy using different datasets and predict crime-prone regions in India, suggesting improved accuracy by considering specific states/regions.[6] R. Garg, et al. emphasize preprocessing using tools like MS Excel, R, and Tableau for exploratory analysis to identify data issues and create visualizations. Focus lies in demonstrating the utility of data science, particularly through R programming and Tableau, in uncovering insights from historical crime data to aid in crime prediction when integrated with other pertinent data sources.[7] Catlett, et al. propose an algorithm combining data analysis, machine learning, and geographic information systems to improve public safety in urban communities. The approach involves detecting crime-dense regions, extracting crime predictors, and implementing spatial and temporal aggregation, tailoring autoregressive models for each detected region to predict crime.[8] G. Borowik, et al. utilize forecasting tools to identify variance patterns and estimate prediction model parameters, focusing on ordinary police activities and interventions in social events like hooliganism. Emphasize the importance of exploratory analysis and data mining in crime detection and forecasting, highlighting how analytical algorithms enhance resource allocation for law enforcement agencies and improve crime prediction accuracy.[9] Biswajit Panja, et al. focuses on algorithms like KNN and ANN for crime mapping. Research emphasizes

evidence-based analysis funded by the Office of Community Oriented Policing Services (COPS) and proposes a systematic approach involving data collection, grouping, clustering, and forecasting to reduce crime rates and enhance public safety.[10] Chung-Hsien Yu et al. present a crime forecasting system that was created in partnership with a US municipal law enforcement agency highlights the need for more theoretical knowledge in order to provide working solutions for crime prevention that are suited to particular locations and periods. Using a variety of data mining classification approaches, the paper describes the architecture for datasets taken from actual crime reports in order to predict crime trends and hotspots.[11] Jianfa Hu et al. systematically examines five analysis of data methods for use in criminal investigations, talking about the problems and uses of large-scale criminal investigation in security and analysis of crimes. In order to address the analysis of large amounts of data in the context of criminal investigations, the paper's conclusion makes several study suggestions.[12] Sunil Yadav, et al. utilize regression modeling to forecast crime rates in different Indian states based on historical data from 2001 to 2014, and uses various machine learning approaches on criminal records to improve predicted accuracy and assist in crime reduction. The study highlights the importance of quick crime resolution for society and provides insightful information for local law enforcement agencies.[13] Romika Yadav et al. deploy a combination of 'Big Data' techniques and Linear Regression to examine crime data in various time frames. They have used ARIMA models to forecast crime sites with minimal variance. The research focuses on enhancing predictive accuracy and certainty in crime prediction, leveraging Auto Regression Techniques and "R" tool experimentation for improved inference.[14] Kai Seidensticker et al. has elaborated the crime forecasting methods in the state of North Rhine-Westphalia, Germany, with an emphasis on analysis of time series, risk area modeling, and policing prediction. It provides insights into maximizing efforts to avoid crime and

emphasizes the possible use of spatiotemporally-based analysis in improving operative and strategic decisions for security.[15] Hongyu Lv et al. develop a method that uses deep learning models to extract criminal components and forecast excavation-related heritage crimes, focusing on the importance of holidays and days of celebration. Decision-making assistance to significant security departments and governmental authorities, with better predicted accuracy through the integration of spatiotemporal analysis and deep learning techniques.[16] The authors elaborate about the IoT technologies pivotal for data analysis through mobile applications, emphasizing communication protocols, middleware, fog, and cloud computing. Integrating these enables real-time data collection and analysis for enhanced analysis projects utilizing mobile IoT applications.[17]

3. Methodology

The following diagram outlines a three-step process for crime data analysis, encompassing data sourcing, extraction, and visual analysis. Crime data is gathered from sources such as Kaggle and web-scraped websites, with a focus on parameters like Crime, time and location of incidents.

The visualization methods encompass bar charts, line graphs, radar charts, pie charts, and nested pie charts. Meanwhile, data analysis involves the creation of heat maps and mapping crime data onto India's geographical map. Google Maps Embed API has been used for locating the area on map. The map and the table will be updated according to the pin code entered by the user. The FB Prophet Library is then used for time series analysis in Figure 1.
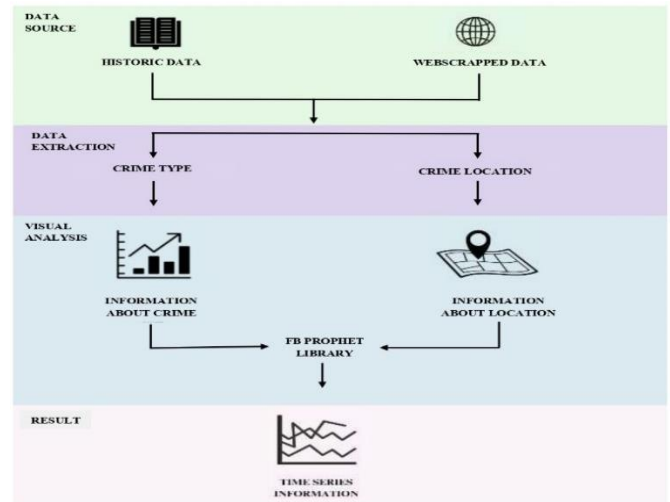

Figure 1 System for crime visualization

3.1. Data

The study utilizes historical data sourced from Kaggle [18], consisting information on various crimes across multiple regions in India since 2001. Additionally, data is acquired through web scraping from diverse online sources in Table 1.

3.2. Features

The parameters being analyzed include both the location and time dimensions of crime. By scrutinizing crime location, we can identify hotspots— areas where criminal activity is concentrated. Simultaneously, analyzing the timing of crimes provides valuable insights into patterns and trends for understanding criminal behavior during specific periods. Both these parameters, helps in the development of effective crime prevention strategies.

Table 1 Comparison of Crime with Locations

Types of Crime & Places	Murder	kidnapping	Robbery
	Value (%)	Value (%)	Value (%)
Maharashtra	1.37%	0.81%	0.36%
Uttar Pradesh	2.53%	4.35%	1.61%
Bihar	2.35%	3.14%	1.02%

3.3. Visual Analysis

Figure 3.1 demonstrates Prophet, a robust time series forecasting model capable of handling seasonality and holiday effects on monthly, weekly, and daily scales. It's the preferred choice for forecasting time-dependent data with significant seasonality. Facebook relies on Prophet for accurate projections, leveraging its resilience to missing data and outliers as described in Prophet's GitHub documentation. The Prophet API is easily available in both Python and R for predictive tasks is Figure 2.
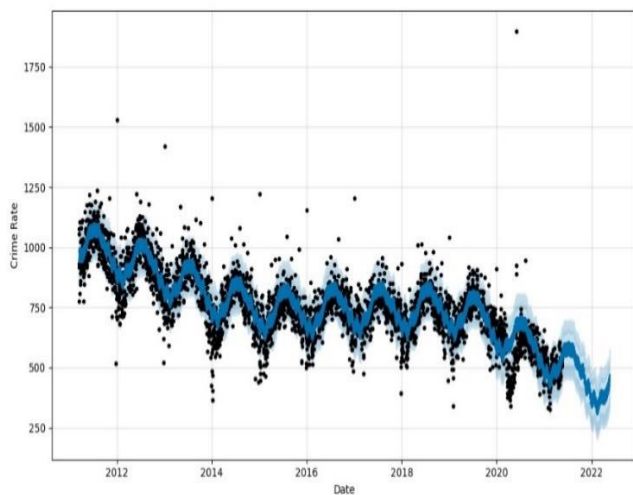

Figure 2 Predictive Analysis Using FB Prophet Library

The trend plot graph depicts a decreasing trend observed over the years, with the x-plane spanning from 2012 to 2022 and the y-plane labeled as "trend," ranging from 400 to 1000. This trend showcases a consistent decline in values over the specified time period. Moving to the weekly plot graph, it represents weekly data, with the x-plane indicating the days of the week and the y-plane labeled as "weekly," ranging from -20 to 40. This graph illustrates fluctuations in values occurring throughout the week, providing insights into weekly patterns or trends. Lastly, the yearly plot graph illustrates yearly data, with the x-plane displaying specific months and dates. The y-plane is labeled as "yearly," ranging from -100 to 50. This graph portrays a wave-like trend, indicating variations in values over the course of the year, potentially reflecting seasonal patterns or fluctuations.

The homepage displays in Figure 3&4 a table featuring crime data from various Indian states for the year 2013, encompassing total crimes, murders, kidnappings, and robberies. Each state is represented with its corresponding crime rates. Notably, the map highlights Gujarat's crime data, showcasing detailed statistics separately. In Gujarat, for instance, there were 1113 murders, 1133 kidnappings, 1136 robberies, and a total of 105469 crimes recorded.
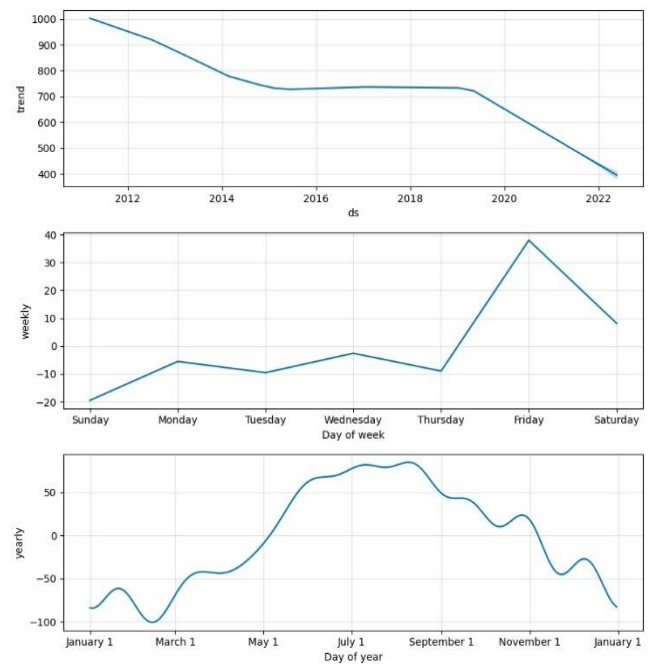

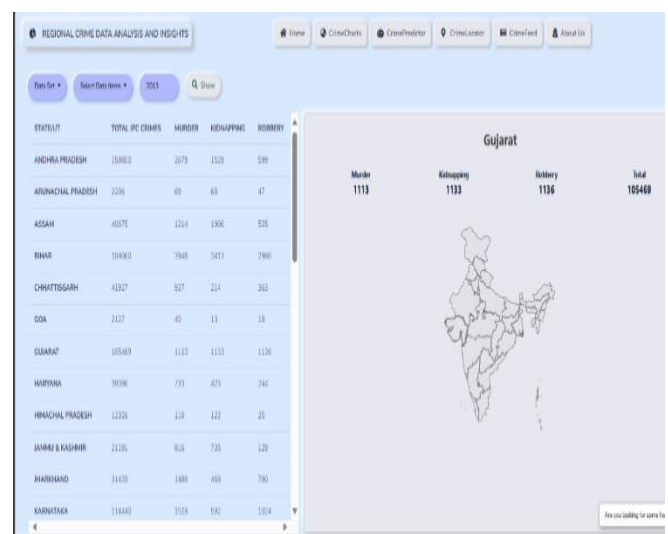**Figure 3 Patterns in Crime Data**

Figure 4 Crime Rates for Different States

The Heat Map displays in Figure 5 shows, regional crime data analysis with varying colors indicating different intensity levels of criminal activity across specific locations from India. The visual representation of complex data through the heat map makes it easier to interpret and analyze crime statistics.
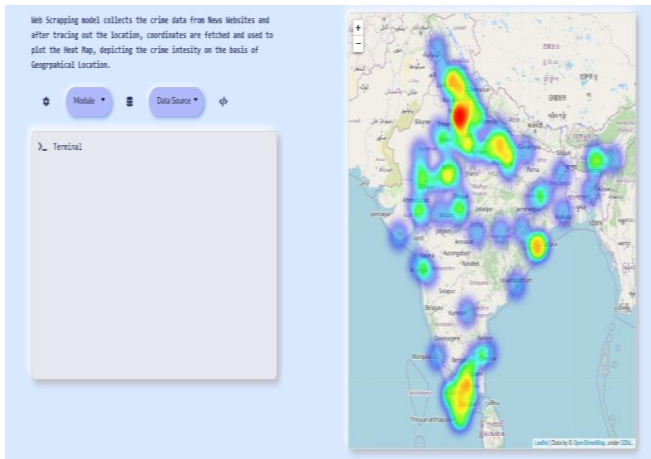

Figure 5 Heat Map for Hotspot Regions

The bar graph in Figure 6 representing crime data for the year 2007, 2009, and 2011 across Madhya Pradesh, Maharashtra, Assam, Gujarat, and Kerala, with each state represented by a distinct color. Among these states, Madhya Pradesh exhibits the highest crime rates, while Assam records the lowest, showcasing variations among states over the specified years. The graph offers a clear visual comparison of crime trends, highlighting differences in occurrence across the selected states.

Figure 6 Crime Result for Different States

The nested pie chart displays in Figure 7 shows crime data for the years 2007, 2009, and 2011 across Madhya Pradesh, Maharashtra, Assam, Gujarat, and Kerala. Each outer slice corresponds to a different year, while individual states are represented by distinct colors within each year. This visualization offers a concise comparison of crime rates over time and across states, facilitating analysis of regional and temporal trends.

Figure 7 Crime Results for Different States and Years

Pie chart in Figure 8 showcases crime data segmented by states and years, with each portion representing a specific region's crime rate for a given year. The chart visually compares crime proportions across Madhya Pradesh, Maharashtra, Assam, Gujarat, and Kerala, utilizing different colors for clarity. It offers a clear and concise insight of crime distribution for regional comparisons within the crime dataset.

Figure 8 State Wise Representation of Crime Data

This page provides the functionality to locate and analyse the crime rate for the pin code of the location provided by the user. This is the crime locator page on the where application where the data and map changes with the pin code is Figure 9.
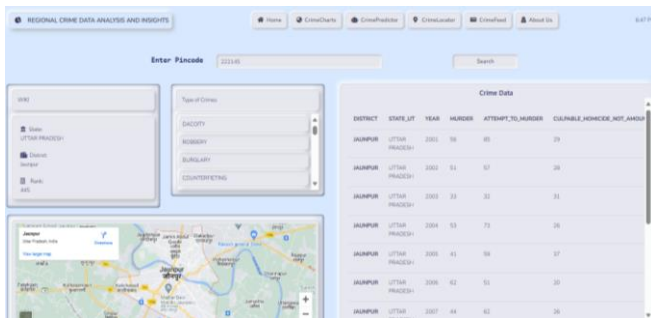

Figure 9 State Wise Representation of Crime Data

The Google Feedback Form facilitates users in sharing experiences, suggesting improvements, and concerns, enhancing tool usability and aligning analysis with user needs is Figure 10.
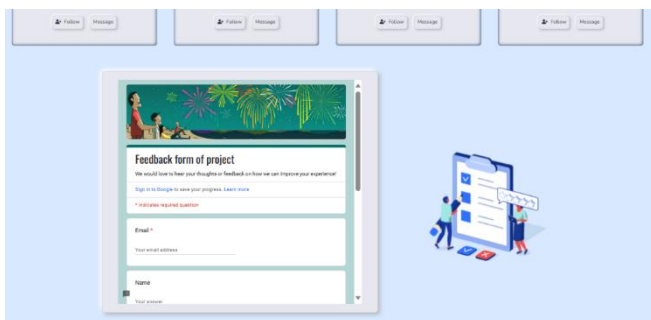

Figure 10 Feedback Form of Project

4. Future Scope

The training model for the purpose of future improvement will be used to identify India's crime-prone locations. More specific data visualization tools can be used. The project will be Open source for developer community to contribute. We can accept feedback from the site visitors and implement the modules as per the necessity.

Conclusion

The analysis of crime rate for Indian states is very useful in studying the crime trend specifically in a certain area. If there has been a rise in crime, the authorities can react suitably. Any nation's dataset may be used with data visualization tools and the prediction model. The community may learn about crimes in various regions and locations of a country by identifying places that are prone to crime.

References

[1]. Nag, R. Ranjan and C. N. S. Vinoth Kumar, "An Approach on Cyber Crime Prediction Using Prophet Time Series," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-5, doi: 10.1109/I2CT54291.2022.9825386.

[2]. D. M. Raza and D. B. Victor, "Data mining and Region Prediction Based on Crime Using Random Forest," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 980-987, doi: 10.1109/ICAIS50930.2021.9395989.

[3]. N. Kanimozhi, N. V. Keerthana, G. S. Pavithra, G. Ranjitha and S. Yuvarani, "CRIME Type and Occurrence Prediction Using Machine Learning Algorithm," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 266-273, doi: 10.1109/ICAIS50930.2021.9395953.

[4]. Pratibha, A. Gahalot, Uprant, S. Dhiman and L. Chouhan, "Crime Prediction and Analysis," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170731.

[5]. Mary Shermila, A. B. Bellarmine and N. Santiago, "Crime Data Analysis and Prediction of Perpetrator Identity Using Machine Learning Approach," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2018, pp. 107-114, doi: 10.1109/ICOEI.2018.8553904.

[6]. S. Sathyadevan, M. S. Devan and S. S. Gangadharan, "Crime analysis and prediction using data mining," 2014 First International

Conference on Networks & Soft Computing (ICNSC2014), Guntur, India, 2014, pp. 406-412, doi: 10.1109/CNSC.2014.6906719.

[7]. R. Garg, A. Malik and G. Raj, "A Comprehensive Analysis for Crime Prediction in Smart City Using R Programming," 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2018, pp. 14-15, doi: 10.1109/CONFLUENCE.2018.8442491.

[8]. Catlett, E. Cesario, D. Talia and A. Vinci, "A Data-Driven Approach for Spatio-Temporal Crime Predictions in Smart Cities," 2018 IEEE International Conference on Smart Computing (SMARTCOMP), Taormina, Italy, 2018, pp. 17-24, doi: 10.1109/SMARTCOMP.2018.00069.

[9]. G. Borowik, Z. M. Wawrzyniak and P. Cichosz, "Time series analysis for crime forecasting," 2018 26th International Conference on Systems Engineering (ICSEng), Sydney, NSW, Australia, 2018, pp. 1-10, doi: 10.1109/ICSENG.2018.8638179

[10]. B.Panja,P.Meharia and K. Mannem, "Crime Analysis Mapping, Intrusion Detection - Using Data Mining," 2020 IEEE Technology & Engineering Management Conference (TEMSCON), Novi, MI, USA, 2020, pp. 1-5, doi: 10.1109/TEMSCON47658.2020.9140074.

[11]. C. -H. Yu, M. W. Ward, M. Morabito and W. Ding, "Crime Forecasting Using Data Mining Techniques," 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 2011, pp. 779-786, doi: 10.1109/ICDMW.2011.56.

[12]. J. Hu, "Big Data Analysis of Criminal Investigations," 2018 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 2018, pp. 649-654, doi: 10.1109/ICSAI.2018.8599305.

[13]. S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma and N. Yadav, "Crime pattern detection, analysis & prediction," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2017, pp. 225-230, doi: 10.1109/ICECA.2017.8203676.

[14]. R. Yadav and S. Kumari Sheoran, "Crime Prediction Using Auto Regression Techniques for Time Series Data," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2018, pp. 1-5, doi: 10.1109/ICRAIE.2018.8710407.

[15]. Using Forecasting Methods on Crime Data: The SKALA Approach of the State Office for Criminal Investigation of North Rhine-Westphalia by Kai Seidensticker and Katharina Schwarz Eng. Proc. 2022, 18(1), 39; https://doi.org/10.3390/engproc2022018039

[16]. The System for Extracting Crime Elements and Predicting Excavation-Type Heritage Crimes Based on Deep Learning Models by Hongyu Lv, Ning Ding, Yiming Zhai,Yingjie Du and Feng Xie Systems 2023, 11(6), 289; https://doi.org/10.3390/systems11060289

[17]. Sonali Lohbare (Pakhmode) et al. "End-to-End Supporting System for IoT Applications: Survey", 4th Edition of International Conference on Communications and Cyber-Physical Engineering, CDSA 2021.Springer Nature Singapore. https://www.springerprofessional.de/en/end-to-end-supporting-system-for-iot-applications-survey/20404690

[18]. https://www.kaggle.com/datasets/rajanand/crime-in-india