2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)

CRIME DATA ANALYSIS & SAFETY RECOMMENDATION SYSTEM USING MACHINE LEARNING

R.M.Akil
Department of Information Technology
PSG College of Technology
Coimbatore, Tamil Nadu, India
akilraj18122001@gmail.com

Dr. S. Sarathambekai,
Associate Professor
Department of Information Technology
PSG College of Technology
Coimbatore, Tamil Nadu, India
ssi.it@psgtech.ac.in

Dr. T. Vairam,
Assistant Professor (Sl. Gr.)
Department of Information Technology
PSG College of Technology
Coimbatore, Tamil Nadu, India
tvm.it@psgtech.ac.in

R.Sanjay Krishnan
Department of Information Technology
PSG College of Technology
Coimbatore, Tamil Nadu, India
sanjaykrishnan13710@gmail.com

G.S. Dharaneesh
Department of Information Technology
PSG College of Technology
Coimbatore, Tamil Nadu, India
dharan221001@gmail.com

D. Janarthanan
Department of Information Technology
PSG College of Technology
Coimbatore, Tamil Nadu, India
djanarthanan614@gmail.com

***Abstract*—Crime is the intentional commission of an act that is often seen as risky or socially damaging, specifically prohibited by law, and punishable. [1] Judicial rulings used to be the main method of defining crimes in the common-law system. [2] Today, most common-law offenses are governed by statutes. Many people hold the opinion that a crime cannot exist without a law. Behavior and an associated state of mind are the usual elements of a crime. Criminal offences include things like arson, assault, bribery, burglary, child exploitation, counterfeiting, embezzlement, extortion, forgery, fraud, hijacking, murder, kidnapping, perjury, piracy, rape, sedition, smuggling, treason, theft, and usury. [5] The major goal of this initiative is to assist people who often visit new locations by teaching them about the region's criminal history so that they may exercise caution and avoid being a victim of any kind of crime. [7] The user may clearly comprehend the most frequent crimes in the area and how often each crime occurs in different parts of the state given the extensive analysis of the crime data.**

Keywords— Crime Analysis, KMeans Clustering, Visualization, NCRB, Machine Learning.

I. Introduction

In India, the number of crimes rises daily, placing a growing load on the court system as it tries to handle the backlog of old criminal cases.There have been about 2 lakh fewer new cases of robbery, burglary, theft, and crimes against women, children, and the elderly.[22] Most of the serious offences against women were classified as “cruelty by a spouse or his family” (30.2%), “attack on women to insult her dignity” (19.7%), “kidnapping of women” (19%), and “sexual assault” (7.2%).[22] Homicide rose by 1 per cent, but “violent crimes” fell by 0.5%. When it comes to women’s safety, Delhi is one of the country’s cities with many incidents. Over 10,093 instances of violence against women were documented in the city in 2020.[22] On the other side, Uttar Pradesh was the only state to improve in this section as a result of the state's increasing number of claims of property destruction, the majority of which happened during protests against the Citizenship Amendment Act of 2019 rallies.

Law enforcement organisations place a high priority on crime prevention because it gives the public a sense of security and protection. The expenditures associated with a crime and the suffering that a victim endures are reduced when a crime is prevented. Police should take a more active role in activities that could prevent crime because they are the most important factor in crime prevention and have the ability to stop it locally. Everyone has a responsibility to prevent crime; it is not just the police's job. India requires efficient crime prevention policing since the criminal cycle has become unstoppable. In India, reactive policing is most frequently observed, meaning that our legal system only responds to crimes after they have already occurred. India is experiencing a rise in crime as a result of this inadequate police. In order to prevent crimes from happening and to lower the rising crime rate in a nation like India, proactive policing should be used. This will lessen the pressure on the court system. It also entails researching crime patterns and educating local populations on crime prevention.[22] If crime prevention is avoided locally, it may be avoided generally. Since they can regularly watch their surroundings and foresee the possibility

of a crime, the police play a critical role in local crime prevention. India, a developing nation, uses cutting-edge techniques and technology to improve its prospects for economic success.

India is a developing nation, yet despite this, it is frequently observed that it is in the forefront of technology breakthroughs, particularly with regard to the investigative work of the police. Technology developments are unavoidably necessary for a successful reduction in crime since localized crime prevention is only possible with sufficient police resources. Every day, crime rates rise, whether in India or somewhere else, and while the legal system works to punish the guilty, new crimes continue to be committed.

Machine learning (ML) being one of the pillars of data analysis, enables a system to learn and improve automatically from previous experiences without being explicitly coded. It is not always possible to pinpoint a certain pattern or piece of information after evaluating the data. In such circumstances, ML is used to decipher the precise pattern and data. ML advances the notion that a computer may learn and resolve both intricate mathematical issues and particular problems if given access to the appropriate data. ML is often divided into two categories: supervised ML and unsupervised ML. When using supervised learning, a computer is educated using a predetermined set of training examples, enhancing its ability to draw exact inferences from incoming data. Whereas in unsupervised, the machine tries to identify the pattern in the datasets containing data points that are neither classified nor labeled.

The booming machine learning strategies for prediction and data analysis has paved the way to so many new, efficient and simple applications that benefit the greater good. These algorithms, if rightly used with some innovation, can do wonders using just a few thousand numbers. Machine learning helps in avoiding the boilerplate code and lets the machine learn at its own pace and understand the data so that it can give us the right insights.

Section I presents the basics of crime prediction and the need of the prediction model. Section II provides an overview of the literature survey. Section III gives a brief explanation on the existing crime analysis and prediction models. Section IV provides a summary of the proposed methodology that gives a clear understanding on the flow of the project. This section also depicts the system design architecture and the workflow details of the project. All of the hardware and software requirements for this project are also explained. Section V gives a comprehensive in depth understanding on the implementation of the proposed system. It also shows the project results, conclusions and future scope of the project. The last chapter includes the references in form of a bibliography where it has dataset links and API references.

II. LITERATURE SURVEY

Vishan Kumar Gupta [1] proposed a SVM model that aims at addressing a two-class learning problem to find the optimal classification function to discriminate between members of the two classes.To offer the best advice to the public in choosing the ideal residential area and to the police department in using the dataset to combat crime.

Jadhav Payal et al [2] a system that would employ crime data sets to train some future predictions was suggested. The result could then be shown for the user to easily understand it. With the aid of data mining tools, this study recommends a crime mapping analysis based on the "KNN" (K-Nearest Neighbor) algorithm to streamline the process and pinpoint the areas where crimes occur the most frequently.

Neil Shah et al [3] proposed a system that uses the crime mapping analysis based on Deep Neural Network algorithm to simplify this process and identify the most frequently crime occurring zone with the help of data mining techniques.

Ashokkumar Palanivinayagam et al [4] provided a study focusing on approaches for feature generation, such as time zone categorization, estimation of the likelihood of a crime occurring, examination of crime hotspots, and vulnerability analysis. In order to improve the subject machine learning algorithm's accuracy, this study aims to extract the key attributes, such as time zones, crime probability, and crime hotspots.

Boppuru Rudra Prathap et al [5] suggests the K-means method's findings regarding crime hotspots. The KDE strategy is used to handle crime density, and this methodology has overcome the problems of the KDE algorithm that it replaced. The study's findings showed that the ARIMA model and the proposed crimes predicting model were equivalent.

P. B. Fajemiseye et al [6] proposed a Self-Organizing Map (SOM) and K-means clustering algorithm are used to cluster dataset to increase the learning rate of the system and Geographic Information Systems (GIS) provide the system with a digital representation that enables the user to map crimes locations analytically and descriptively. The methodology employed in this project is the object oriented methodology. The system was implemented using the PHP programming language and MySQL database.

Timothy Moses et al [7] created an android crime reporting system that works with the Google Map/Places APIs to notify a nearby police station of a crime. It provides ways to report crimes, including calling other apps, recording audio, and sending pictures of the crime scene. In terms of law enforcement agencies' ability to respond quickly to a crime scene and the anonymity of the individual reporting the incident, it has shown to be effective (when compared to the current method of reporting a crime).

K.Meghana Chowdary et al [8] developed a method for combining computer science and criminal justice to create a data mining strategy that uses K Means Clustering and Decision Tree Classifier to help solve crimes more quickly. Instead of concentrating on variables that contribute to crime, such as the criminal history of the offender or political animosity, we are largely concerned with daily crime elements.

III. SYSTEM ARCHITECTURE

The below figure depicts the flow of the existing system architecture.

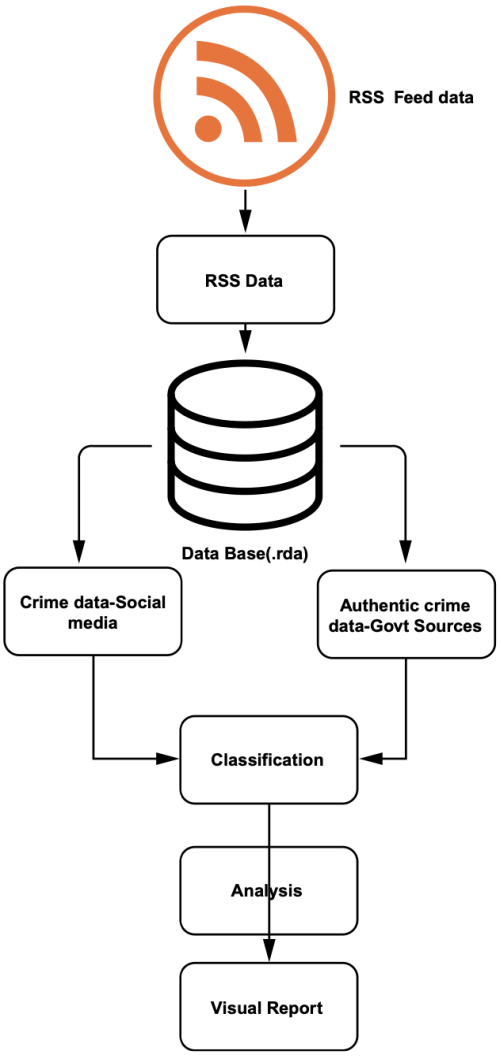

Fig 1. Existing system architecture

A. *LIMITATIONS OF THE EXISTING SYSTEM*

- The report is completely based on the real time news data which provides no state-wise or district-wise analysis
- It classifies the fetched data as Drug-related crimes Violent crimes, Commercial crimes, Property crimes, etc.,
- The end result is not communicated to the beneficiary or the public, rather they reach only the authorities.
- They are limited only to news data and can serve only a small set of requirements.

This work helps in providing visualizations of crime in the specific user location using IPify API and the data is extracted from the NCRB dataset by Rajanand Elangovan.

IV. PROPOSED SYSTEM

Firstly, the dataset is cleaned for irrelevant columns and renamed to one common name across all the datasets. The individual datasets are then merged into one single dataset for easy analysis. The dataset then undergoes outlier analysis and null values handling. The cleaned dataset is considered for further analysis. The dataset is looked upon for categorical values and is label encoded to numerical values for model building. MinMax Scaler is used to scale the dataset. KMeans clustering is then used for clustering the dataset with respect to the 'cases' column. The clustered data is added as a new column to the existing dataset.

Using streamlit, an open-source app framework for Machine Learning and Data Science projects to create web applications, a frontend is designed so as it is appealing to the users. The clustered data is used in the backend and the user location is read using the 'ipify' API that gives the location of a user based on the ISP's IP address. The location is then mapped to the clustered data in the backend to show the user the crime data of the place. The webpage would contain different plots to visually understand the crime history of the place such as bar plots and pie charts. The plots are interactive and responsive so as to facilitate the user to know the numbers when they hover their mouse onto the charts. The below image shows the proposed system architecture.

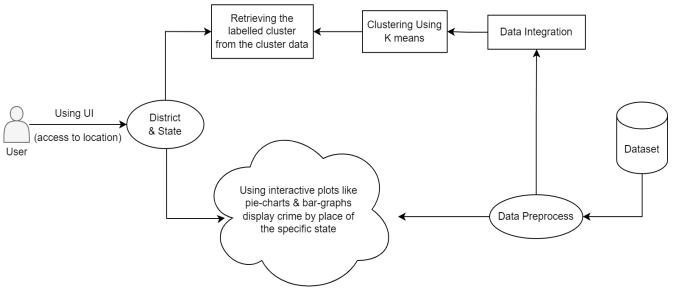

Fig 2. Proposed system architecture

Finally, few safety measures are also included in the webpage to educate the user on the do's and dont's when in a public place so as to avoid becoming a victim of any crime. This implementation would help travelers understand the place they are in currently and be cautious about the surroundings based on the data provided.

NCRB collects, collates, compiles and publishes the police recorded criminal cases only on an annual basis. There are many victims of a crime who may not have reported it to the police, or the police may not have even registered the incident, in such cases the incident is not recorded in the statistics. Under the general supervision of the Ministry of Affairs, the NCRB updated the data gathering proformas in 2014 with input from various Central Ministries and State Governments. The dataset considered is available at data.world in name 'Crime in India'. This dataset is provided by Rajanand Ilangovan, Business Intelligence Developer.

This dataset contains State/UT wise data about various crimes in India. It includes two sub-divisions: District_wise_crimes_committed_IPC and Crime_by_place_of_occurrence. Each division consists of 3 csv files having

crimes that happened from 2001 to 2012, 2013 and 2014. Further, each csv file contains State/UT wise and District wise data about each of the crimes. Each csv file contains various crimes like dacoity, robbery, burglary etc. These crimes are also categorized into places of occurrence like residential, highways, commercial establishments, and etc.

V. IMPLEMENTATION AND RESULT DISCUSSION

The proposed system is implemented using Python and executed in Processor: Intel i7 and above, RAM: 8GB or above, High end GPUs (min. 12 GB)

The Software requirements are Google Colab/Jupyter Notebook, Visual Studio Code, Google Chrome and the Operating System : macOS/Windows/Ubuntu. Framework used is StreamLit

After all the preprocessing, the required data is taken as a separate data frame that contains only the state, district and the cases columns.

```
d.head()
```

	STATE	DISTRICT	CASES
0	A&N ISLANDS	A AND N ISLANDS	807
1	A&N ISLANDS	ANDAMAN	7377
2	A&N ISLANDS	CAR	48
3	A&N ISLANDS	NICOBAR	266
4	A&N ISLANDS	NORTH	271

Fig 3. Overview of the dataset

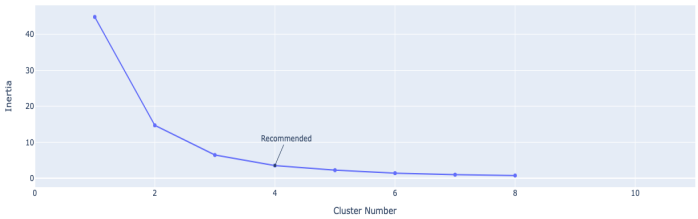

Fig 4. Elbow Method

The resultant data frame is considered for clustering. Using the elbow method, for clustering based on the ‘cases’ column, it is found that k=4 might be one possible optimum value for the number of clusters. Hence the same is chosen and the data frame is clustered.

The clustered data is added as a new column to the existing dataset. The clusters were numbered from 0 to 3.

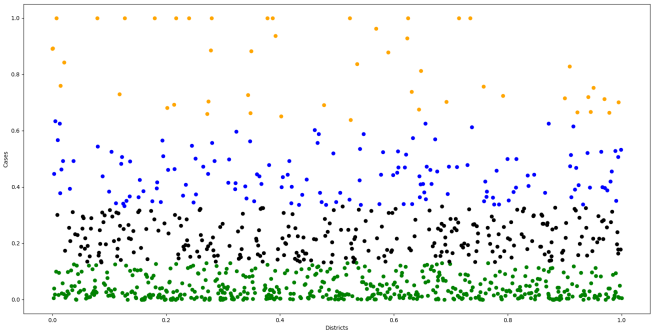

Fig 5. Scatter Plot of the clusters

Fig 6. The landing page

The clusters were then studied for their mean value and labeled as follows:

- 3 as Very High Crime Area
- 1 as High Crime Area
- 2 as Low Crime Area
- 0 as Very Low Crime Area

```
d.sample(10)
```

	STATE	DISTRICT	CASES	cluster
379	JAMMU & KASHMIR	BORDER DISTRICT	1192	0
1061	WEST BENGAL	JHARGRAM POLICE DISTRICT	803	0
914	TELANGANA	NALGONDA	9254	0
183	CHHATTISGARH	KABIRDHAM	16709	0
788	RAJASTHAN	AJMER	95564	1
813	RAJASTHAN	JAIPUR RURAL	66638	1
976	UTTAR PRADESH	HATHRAS	24716	2
636	MAHARASHTRA	SOLAPUR RURAL	58024	1
560	MADHYA PRADESH	KATNI	41389	2
461	KARNATAKA	DAVANAGERE	43645	2

Fig 7. Sample Transformed data

Based on the above data the numerical values are label encoded into categorical values

```
d.cluster = d.cluster.replace(3, "Very high crime area")
d.cluster = d.cluster.replace(1, "High crime area")
d.cluster = d.cluster.replace(2, "Low crime area")
d.cluster = d.cluster.replace(0, "Very low crime area")
```

Fig 8. Label Encoding

Below is the sample of the output obtained after the entire clustering process:

```
d.sample(15)
✓ 0.3s
```

	STATE	DISTRICT	CASES	cluster
857	TAMIL NADU	CUDDALORE	105767	Very high crime area
334	HARYANA	IRRIGATION & POWER	192	Very low crime area
504	KERALA	KASARGOD	51527	High crime area
171	CHHATTISGARH	BEMETRA	1645	Very low crime area
570	MADHYA PRADESH	RAISEN	53522	High crime area
389	JAMMU & KASHMIR	KATHUA	15119	Very low crime area
22	ANDHRA PRADESH	KHAMMAM	79510	High crime area
945	UTTAR PRADESH	BAHRAICH	23522	Low crime area
597	MAHARASHTRA	AURANGABAD RURAL	44283	Low crime area
291	GUJARAT	KUTCH (EAST(G))	8466	Very low crime area
1035	UTTARAKHAND	UTTARKASHI	2290	Very low crime area
41	ANDHRA PRADESH	VISAKHA RURAL	42178	Low crime area
7	A&N ISLANDS	SOUTH ANDAMAN	556	Very low crime area
369	HIMACHAL PRADESH	SIRMAUR	11101	Very low crime area
792	RAJASTHAN	BARMER	42161	Low crime area

Fig 9. Label Encoded sample data

In case the above location is incorrect or if the user wants to know about the crimes in a specific location in India, a drop-down is provided that takes in the user input and computes for results in real time.

The user-entered data or the automatic tracking of the user location is stored in a variable and the coordinates of that particular location is retrieved using the geopy python package. Then those coordinates are plotted in a map using the streamlit.map() method which accurately plots the location in a responsive map

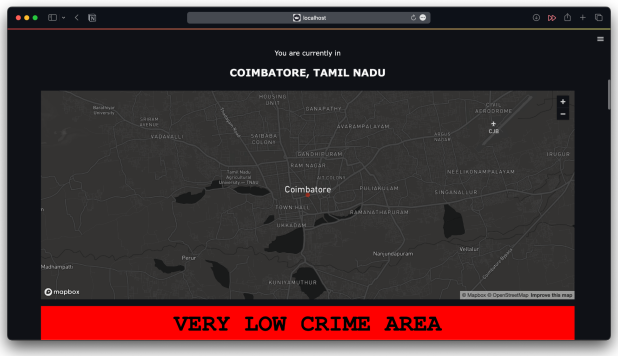

Fig 10. Map that locates the user

Once the user is notified of the location on a general basis, a deep dive into the type of crimes and a short analysis of crime in that place is provided to the user which educates the user on the type of crime, place of crime and the most frequently occurring crimes of the place.

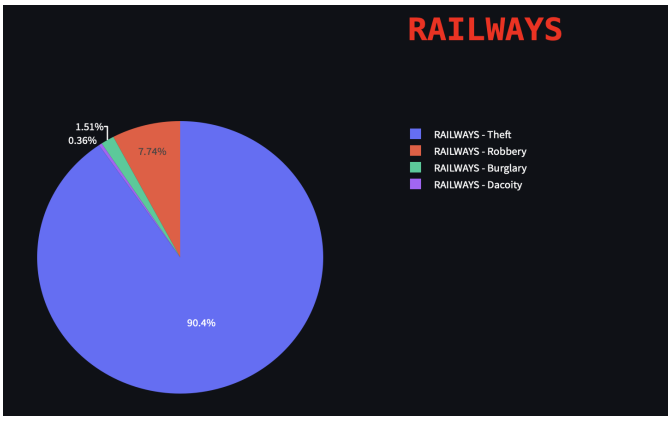

Fig 11. Visualization of the crime in Railway station in the user locality

Finally, there are few general safety recommendations one must follow in order to prevent themselves from being a victim to crimes of any sort. These instructions warn the user and guide them on the do's and dont's when in a public place, when alone and while traveling.

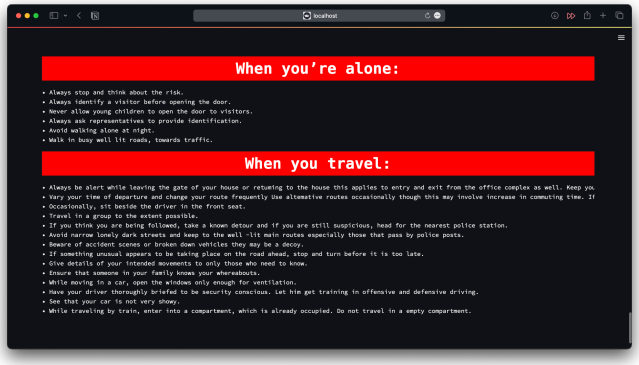

Fig 12. Safety Recommendations

VI. CONCLUSION AND FUTURE WORK

In today's world, when the average amount of data encountered by a person has expanded by leaps and bounds in recent years, the use of data mining techniques to extract meaningful information from large volumes of raw data has become crucial. The results of using the different data mining techniques, algorithms, and models described above on datasets can be quite useful, especially for law enforcement groups.

Future work includes reporting a crime to administrators through App-to-App call, sharing crime scene photographs, or utilizing the smartphone microphone to record the event and then forward it to the station. Future work also includes pop-up notifications using Web push notifications. Anticipate that the findings of this work will be utilized to enhance crime prediction models, spark debate on the integration of data accumulation and feature design for characterization learning, and advance the modeling of law enforcement experience and criminology theory.

At present, the crimes in different states of India are increasing day by day because of this, people feel insecure and find their society inappropriate. Due to different types of crimes, the security personnel face difficulty in handling those. This system will identify and focus on the highest committed crime at the location. To identify the accurate location of the user based on GPS. To automatically notify the user of the crime history as the location changes during a travel. Plotting all the nearby police stations and phone numbers. To improve the system by using the latest data when available.

References

[1] Vishan Kumar Gupta, "Crime Tracking System and People's Safety in India using Machine Learning Approaches" International Journal of Modern Research, 2022.

[2] Jadhav Payal, "Crime Detection System using Data Mining ", International Journal of Computer Science and Mobile Computing, 2021.

[3] Neil Shah,Nandish Bhagat, "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention" al. Visual Computing for Industry, Biomedicine, and Art, 2021.

[4] Ashokkumar P, "An Optimized Machine Learning and Big Data Approach to Crime Detection" Wireless Communications and Mobile Computing, 2021.

[5] T. Chitra, S. Karunanidhi, The impact of resilience training on occupational stress, resilience, job satisfaction, and psychological well-being of female police officers, J. Police Crim. Psychol. 36 (2021) 8e23, http://dx.doi.org/10.1007/s11896- 018-9294-9.

[6] Boppuru Rudra Prathap, K. Ramesha, "Geospatial crime analysis and forecasting with machine learning techniques", Journal of Computational and Theoretical Nanoscience Vol. 17, 74–86, 2020.

[7] V. Srinidhi, P. Saranya, M. Ashok, An affirmative learning techniques to analyse the crime scene in jewel theft murder, Int. Res. J. Multidiscip. Tech. 2 (2020) 1e7, http://dx.doi.org/ 10.34256/irjmt2051.

[8] P. B. Fajemiseye, "An Intelligent Crime Prediction System Using Artificial Neural Network And geographical Information System", International Journal of Scientific & Engineering Research Volume 11, Issue 9,2020.

[9] Timothy Moses, "An Android Location-Based Crime Reporting System Using The Google Map Api", google map for location based crime identification, 2020.

[10] K.Meghana Chowdary, P.Samyuktha,V.Ganesh Kumar,T.Y.Seshadri Rao, "An Approach For Crime Analysis Using Clustering Algorithm" under Anil Neerukonda Institute of Technology & Sciences 2020.

[11] A. Onan, M.A. Toçoglu, Satire identification in Turkish news articles based on ensemble of classifiers, Turk. J. Electr. Eng. Comp. Sci. 28 (2020) 1086e1106, http://dx.doi.org/10.3906/ elk-1907-11.

[12] Neil Shah , Nandish Bhagat and Manan Shah, Shah et al. "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention", Visual Computing for Industry, Biomedicine, and Art (2021) 4:9 pages.

[13] M.K. Anser, Z. Yousaf, A.A. Nassani, S.M. Alotaibi, A. Kabbani, K. Zaman, Dynamic linkages between pov-erty, inequality, crime, and social expenditures in a panel of 16 countries: two-step GMM estimates, J. Econ. Struct. 9 (2020) 1e25, http://dx.doi.org/10.1186/s40008-020-00220-6.

[14] Reid, J. A. "Crime and Personality: Personality Theory and Criminality Examined", Inquiries Journal/Student Pulse, 3(01), 2011.Accessed 28.06.2019: http://www.inquiriesjournal.com/a?id=1690

[15] Ying-Lung Lin, Meng-Feng Yen and Liang-Chih Yu, "Grid-Based Crime Prediction Using Geographical Features" ISPRS Int. J. Geo-Inf. 2018.

[16] H.Benjamin Fredrick David and A.Suruliandi, "Survey On Crime Analysis And Prediction Using Data Mining Techniques" ICTACT JOURNAL ON SOFT COMPUTING, 2017.

[17] Vineet Pande and Viraj Samant and Sindhu Nair, "Crime Detection using Data Mining" International Journal of Engineering Research & Technology (IJERT), 2016.

[18] Wajiha Safat, Sohail Asghar , (Member, IEEE), and Saira Andleeb Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques" in IEEE Access, Received April 24, 2021, accepted May 2, 2021, date of publication May 6, 2021, date of current version May 17, 2021. Digital Object Identifier 10.1109/ACCESS.2021.3078117.

[19] J Vimala Devi and Dr K S Kavitha, "Time Series Analysis and Forecasting on Crime data" in Press.

[20] Clark, N. J. and Dixon, P. M. (2020). Spatially correlated self-exciting statistical models with applications to modeling criminal activity, Statistical Modelling Forthcoming

[21] https://data.world/rajanand/crime-in-india

[22] https://ncrb.gov.in/en/crime-india

[23] https://api64.ipify.org