Crime Prediction using Machine Learning

Sridharan S[1], Srish N[2], Vigneswaran S[3] and Santhi P[4]

[1, 2, 3, 4]Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, Chennai, India

Abstract

The process of researching crime patterns and trends in order to find underlying issues and potential solutions to crime prevention is known as crime analysis. This includes using statistical analysis, geographic mapping, and other approaches of type and scope of crime in their areas. Crime analysis can also entail the creation of predictive models that use previous data to anticipate future crime tendencies. Law enforcement authorities can more efficiently allocate resources and target initiatives to reduce crime and increase public safety by evaluating crime data and finding trends. For prediction, this data was fed into algorithms such as Linear Regression and Random Forest. Using data from 2001 to 2016, crime-type projections are made for each state as well as all states in India. Simple visualisation charts are used to represent these predictions. One critical feature of these algorithms is identifying the trend-changing year in order to boost the accuracy of the predictions. The main aim is to predict crime cases from 2017 to 2020 by using the dataset from 2001 to 2016.

*Corresponding author. Email: ch.en.u4cys21080@ch.students.amrita.edu

1. Introduction

This paper focuses on the problem of predicting crime in the location of a particular year and the type of crime that occurred in the past. Due to a population outburst, there are lack of unemployment, and drug abuse among others. Crimes are of two types violent and passive. Violent crime, it is murder, forcible Rape, Robbery, etc… that lead to injury. The challenge faced in violent crime is it cannot be predicted with great certainty since it is systematic or random [1]. According to National Crime Research Center, NCRC, the crime like burglary, and arson are said to have reduced meanwhile others like murder, sex abuse, and rape have been reported. Traditional methods of crime prediction are often based on analyzing past incidents manually so they may not adapt well to changing patterns and trends. Nevertheless, thanks to current technology and a rise in the usage of artificial intelligence, machine learning, and quantitative statistics, now have some excellent instruments at our disposal for researching and, eventually, containing crime [2]. Even while it can be challenging to predict who will commit a crime and who will be the victims, it is possible to foresee the location where it is most likely to happen within a certain time frame.

The urgent demand to preserve the data and investigate the numerous crimes that occur in our modern cities is the need for data prediction, which entails data mining and manipulation. To identify criminals that are well-coordinated and well-equipped to exploit contemporary technology for good communication and to pose significant hazards to the country's security, extensive information, and potent modern procedures are needed. Classification, clustering, and regression techniques, as well as data mining technology, can be used to identify trends and conduct criminal investigations [3]. Data mining is an effective method for evaluating huge amounts of data, but it requires pre-processing processes to extract the needed data rapidly.

1.1. Problem Statement

The problem that crime analysts face is determining how to extrapolate previous criminal activity data into the chances of future occurrences occurring at specific points in space and time. Descriptive analysis deals with identifying temporal and spatial relationships in crime data Predictive

analytics techniques are typically used to foretell the type of crime that could happen at any place at a certain time. The analyst would want a visual map that depicted the degree of probable criminal activity at each place inside their jurisdictional limits. This happens by merging crime and population statistics and feeding them to machine learning algorithms. The prescriptive analyzer offers process re-engineering measures that effectively deploy police resources with the goal of reducing crime and its impact on the general public. Points of view or opinions are clearly important in preventing crime since they enable the police to direct resources to high-risk areas.

1.2. Objective

Using analytical and predictive data analytics methodologies, provide a platform for assessing crime data. Analyse the spatial and temporal (time of day, day of week, and seasons) connections in crime data using the suggested platform. Analyse the relationship between crime data and census data.

The model examines the crime patterns and identifies shifts in the general crime ratio depending on population or demographic ratios, which will make it easier to predict future crime instances. Additionally, it is forecastable how many security measures will be required and how many criminal activities will need to be controlled to stop or lessen the occurrence of any low-level to treacherous criminal activity.

1.3. Scope

The suggested system will be in charge of criminal management, detection, and prevention. The system will use time series, clustering, and data mining techniques to generate predictions of future crime rates. This will be accomplished by displaying crime patterns graphically and by using geographic heat maps to show data concentration and hotspots in real time.

1.4. Dataset

Data for our dataset go from 2001 to 2016. Fieldwork-based primary data collection is used to collect the initial data set for crimes. More than 500 more than ten rows' worth of information make up this set. The primary elements are Name, Years, Months, Crime Types, Crime Areas, Victims Genders, Victim Ages, Victim Areas, and Year. Our dataset is divided into three categories: cases involving women, kids, and IPC at the state level. We used this dataset to forecast crime at the state level.

2. Literature Survey

When a suspected list of criminals is merged with criminal data generated synthetically using the Gaussian Mehmet Sait and Mustafa Gök came up with the Mixture Model. *Tayebi et al.* chose the clustering technique above any other

supervised technique, such as classification, because crimes vary considerably in nature and criminal databases are frequently overflowing with unresolved crimes. The model was then examined, created, and used to sample the data set and train the algorithm [4]. More than 75% were provided by the K-Means Clustering technique. *Akash et al.* have applied the theory of broken windows. The model was then analysed, pre-processed then put into practice to train the algorithm and taste the collection of data. The K-Means Clustering algorithm returned over 75% [5].

The model was then examined, prepared, and put into use to sample the data set and train the algorithm. More than 75% came from the method of K-Means Clustering. The authors frequently generated item sets using a priori techniques, which were also possible for criminals. A crime prediction approach to finding the most likely perpetrators of a given crime. The authors compared the performance of the Nave Bayes Classifier and Decision Tree offender prediction methods. Data sets are transformed into clusters using clustering algorithms, which are then investigated to identify crime-prone locations [6]. *Rajesh Kanna et al.* presented a CNN-based deep learning model based on long short-term memory for detecting crime.

A deep learning model based on Map Reduce is also used to detect intrusions using spatiotemporal characteristics. To improve feature selection accuracy, the black widow optimised method is utilised. [7]. Researchers have published a variety of data mining strategies that enable crime data analysis, crime forecasting, criminal identification, and location of crime hotspots [8]. Utilizing a stacked sparse autoencoder network for the identification of malicious modules for the reliability of VLSI circuits [9]. The Naive Bayes Classifier required less time to execute and achieved a higher accuracy of 78.05%. He investigated various offenses committed by offenders and predicted the likelihood of each offence being committed by that offender again [10].

Sivaranjani et al. [13] present a crime study of six cities in Tamil Nadu, India, using clustering methods k means, DBSCAN, and Agglomerative clustering to group similar patterns for crime detection. Conclude that is superior. *Kansara, Chirag, E.* [14] The authors compared the performance of random Forests, Nave Bayes, and linear regression in identifying factors influencing high crime rates. The authors conclude that the Random Forest performs better based on the comparison with 81.35% accuracy.

3. Proposed Methodology

The answer is provided as a statistical and machine learning model that employs classification, clustering, and regression algorithms; K-NN algorithms, Bayes Nave algorithms, and Regression algorithms that can be used to describe the functional relationships among demographic, economic,

social, victim, and geographic variables. through analysing patterns in criminal data sets. temporal series techniques have been proven to be beneficial in conjunction with the algorithms mentioned previously in allowing the model to forecast criminal incidents with high accuracy according to temporal growth and changing features.

3.1. Algorithm

Clustering algorithms are included in the domain. The K-means partitioning method is widely employed and accepted. Instead of the K-means method, this linear regression is utilized to access consumers to decide the number of clusters on the bases of values however Navies Bayes gives a first-rate outcome, with the two algorithms which get a high accuracy rate. Linear and multi-linear regression which shows the relation of dependent data or variables (like age, gender, etc) and a collection of independent variables discovered at the site of the crime. This method calculates the age values for the victims based on the input criteria mentioned in the metadata column. Given the crime locations, linear regression is used in the crime prediction scenario to determine the age of the most probable offender. Analyzing historical data demonstrates that the ratio of female victims to male victims is steadily rising. This statistic depicts the victim rate for men and women. Due to the numerous crime data sets and the intricate connections between these different forms of data, criminology is a suitable subject for the application of data mining techniques.

Figure 1 illustrates the proposed method. Data sets are transformed into clusters using clustering algorithms, which are then investigated to identify crime-prone locations. These clusters graphically depict a collection of crimes superimposed on a map of the police jurisdiction. combines the location of crimes in stores with details on the type and timing of the incidents. The members of these clusters are used to categorize them. Clusters with a high density of people become crime hotspots, whereas clusters with less people are disregarded Depending on the type of offense, preventive measures are implemented in areas where crime is a problem. The simplest and most popular clustering technique in research and commercial applications is K-means. Large data sets can be clustered using this method because of its lower processing complexity. Given that crimes vary greatly in character and that crime databases are frequently filled with unsolved crimes, we preferred the clustering technique over any other supervised technique, such as classification. The model was then examined, created, and used to sample the data set and train the algorithm. More than 75% were provided by the K-Means Clustering technique. The author applied the theory of broken windows.

The model was then analysed, pre-processed then put into practice to train the algorithm and taste the collection of data. The K-Means Clustering algorithm returned over 75%

[2]. The model was then examined, prepared, and put into use to sample the data set and train the algorithm. More than 75% came from the method of K-means clustering. Using the broken window theory, random forest, and naive Bayes, crime was reduced, and the crime area was located. Create the data frame required to train the model for image recognition, information preprocessing, and identifying criminal hotspots. 0.87% of the best accuracy is provided by the deep learning-tuned model. It is possible to predict crime rates using machine learning's classification and regression techniques. To establish a link between both the dependent and independent variables, multi-linear regression is performed. For single-class and multi-class variable classification, K-Nearest Neighbours is utilized. When we need to divide the target variable into more than two groups, that use the K-nearest neighbours' method. This dataset has three groups of people based on their gender: men, women, and those whose gender is not known. Age can be classified into three categories: young, old, and young. K-nearest Neighbours Classifier is used to group or classify the target variables.
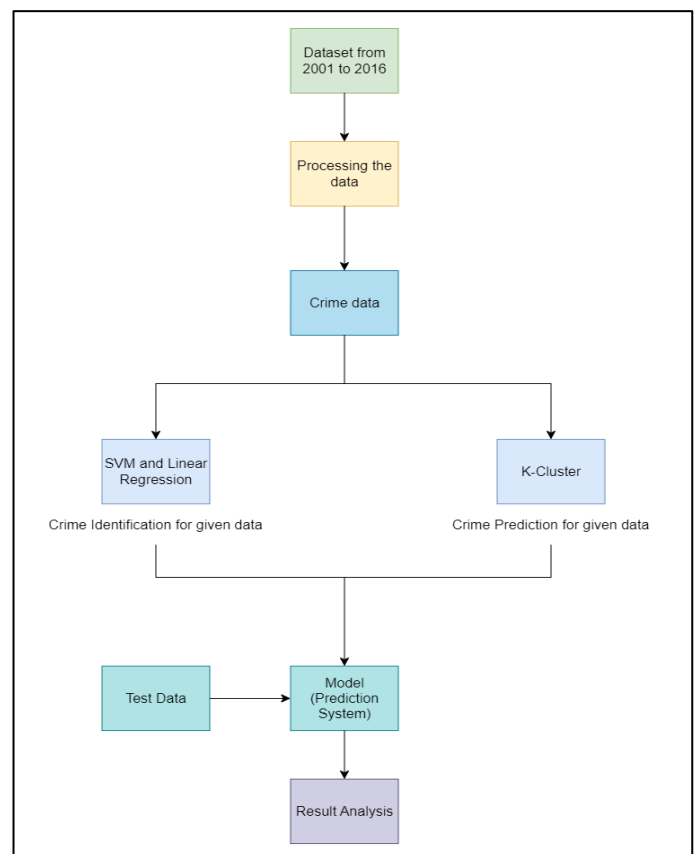

Figure 1. System architecture

3.2. Outcome

The System "Indian Crime Analysis" has Software is currently available and has been designed specifically for criminal investigation to perform tasks that no other method

can. Thus, it is clear that despite the fact that several answers to the issue have been put up, a perfect solution has been produced for every city, state, and nation for every kind of user. The System is precise and would present the analysis in the form of animate visuals and predict the crime ratio precisely, if the system is unable to provide accurate results, then it would notify about the unavailability of data or the proximate cause.

An overview of forthcoming crime data sets and algorithmic crime bases may be found in this section. They assess the crime rate based on a variety of factors, including age, gender, location, and monthly ratios. A range of data sources and methods are used to make predictions, including literature reviews, surveys of common personal data, and statistical models that predict future crime trends. Because some minority groups were included in the classification of crimes, which caused data imbalance, the prediction model had a significant miss rate. So, in order to solve the problem, we employed random oversampling. By extrapolating using a time series analysis of existing crime trends, predict future crime trends., algorithm forecast future crime patterns using a time series study of current crime trends., algorithm Using algorithms that extrapolate using a time series study of current crime trends, anticipate future crime trends. The behaviour of previously recorded data can be used to forecast future patterns in crime.

Every predictive model aims to demonstrate the relationship between a certain predictor and a dependent variable. In order for these models to be more accurate, they must be able to recognize and foresee the variety of circumstances that may in the future affect victimization and crime. Future crime rates are projected in this study in a much more thorough and precise manner.

4. Test Cases

The output of the proposed model is presented in figure 2 to 5, Figure 2 illustrates the visualization of the data set from 2001 to 2015.
Case 1:
Prediction for Prohibition of Child Marriage 2020 in Haryana as shown in figure 3.

Case 2:
Prediction for murder cases up to 2020 in Tamil Nadu as shown in figure 4.

Case 3:
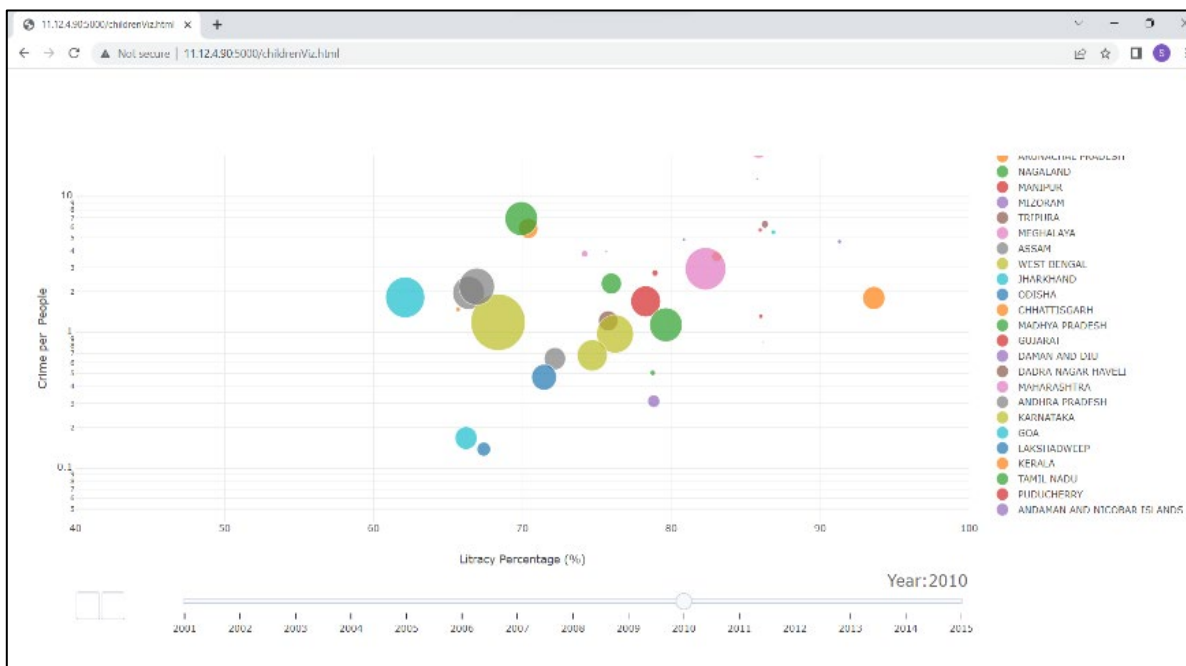Prediction for ARMS ACT cases up to 2020 in Kerala as shown in figure 5.

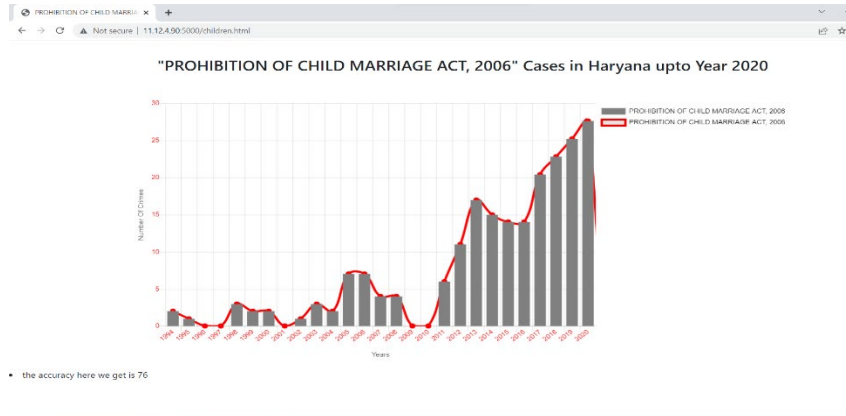**Figure 2.** Visualization of the data set from 2001 to 2015

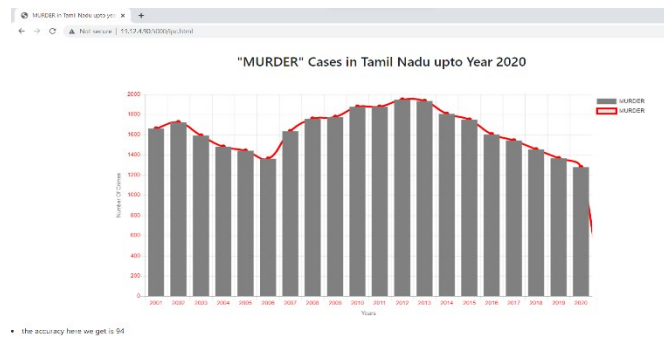**Figure 3.** Predicted graph of Child Marriage in Haryana

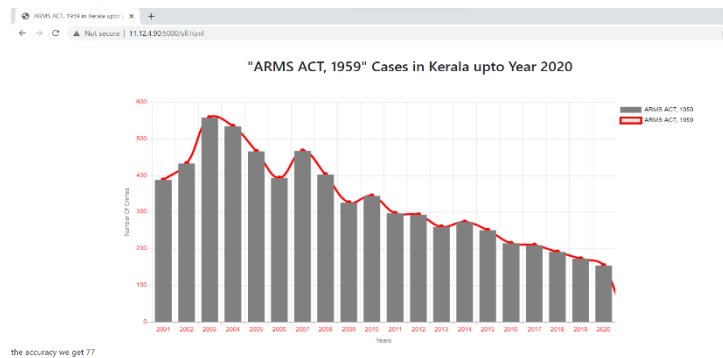**Figure 4.** Predicted graph of Murder case in Tamil Nadu
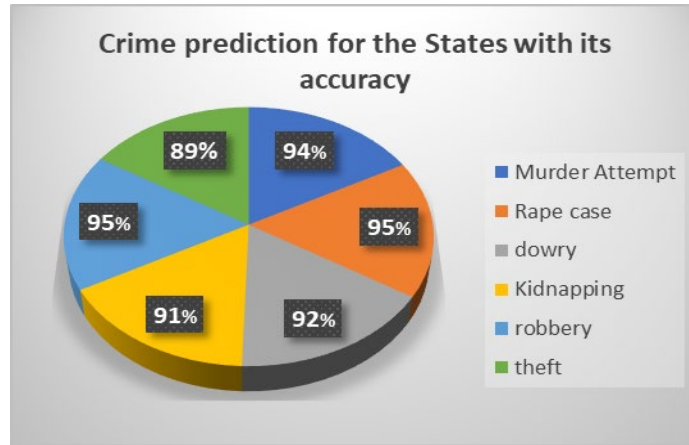

Figure 5. Predicted graph of ARMS act (2020) in Kerala

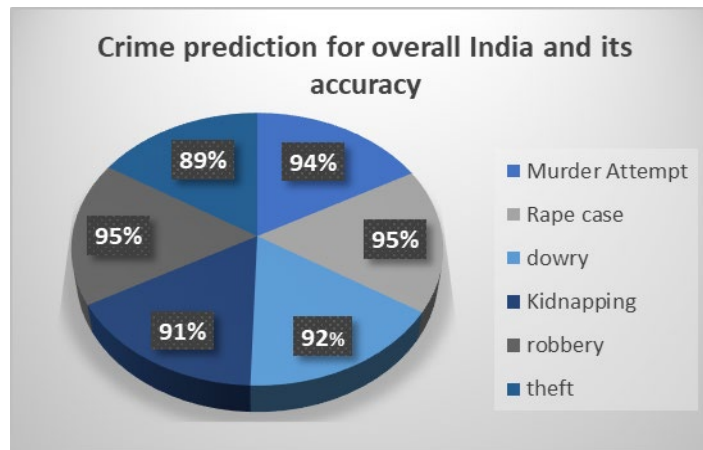**Figure 6.** Crime prediction for the state with accuracy

Figure 7. crime prediction for overall India and it's accuracy

5. Conclusion

It is difficult to use the prediction rate area-specific modelling since crime is rare in many places. In that study, a machine learning algorithm was used to construct and test a model that forecasts crime by age, sex, year, and month. In that study, three distinct machine learning methods are employed. To assess the efficacy of K-nearest neighbour, Naive Bayes, and linear regression in diverse contexts Figures 1 and 2 show how accurate the crime is. While certain linear systems work well and provide greater precision, the overall scenario model uses K-nearest neighbour as our crime prediction approach since it also provides the desired accuracy. We can determine the stronger with the use of these prediction techniques. The algorithm will also show superior accuracy in identifying and locating the places with the highest incidence of crime. Finally, it makes use of the CNN algorithm to analyze the photo data and the Google API to detect the heated zone.

References

[1] Zakir Hussain, K, Durairaj, M, Farzana, G.R.J.: Criminal behaviour analysis by using data mining techniques,30-31 March 2012, Nagapattinam, India, Proceedings of the International Conference On Advances In Engineering Science And Management, IEEE, 2012 pp. 1 – 8.

[2] Kavitha, M, Roobini, S, Systematic View and Impact of Artificial Intelligence in Smart Healthcare Systems, Principles, Challenges and Applications, Machine Learning and Artificial Intelligence in Healthcare Systems. 2023; 25-56.

[3] Sathya, R, Ananthi S, Vaidehi K, A Hybrid Location-dependent Ultra Convolutional Neural Network-based Vehicle Number Plate Recognition Approach for Intelligent Transportation Systems, Concurrency and Computation: Practice and Experience, 2023; 35:1-25.

[4] Tayebi, M.A, Gla, U, Brantingham, P. L. Learning where to inspect: location learning for crime prediction, 27-29 May 2015, Baltimore, MD, USA, Proceedings of the IEEE International Conference on

Intelligence and Security Informatics (ISI), IEEE, 2015, pp. 25-30.

[5] Akash, S, Prabaharan Poornachandran, Vijay Krishna Menon, Soman, K.P. Cybersecurity and Secure Information Systems, Springer Cham, 2019, Chapter number:12, A Detailed Investigation and Analysis of Deep Learning Architectures and Visualization Techniques for Malware Family Identification, Cybersecurity and Secure Information Systems, pp. 24-46.

[6] Rajesh Kanna, P, Santhi, P, Hybrid Intrusion Detection using Map Reduce based Black Widow Optimized Convolutional Long Short-Term Memory Neural Networks, Expert Systems with Applications, 2022, Vol. 194:(116545).

[7] Rajesh Kanna, P, Santhi, P, Unified Deep Learning approach for Efficient Intrusion Detection System using Integrated Spatial–Temporal Features, 2021, Knowledge-Based Systems Vol. 226:(107132).

[8] Sathyadevan, S., Gangadharan, S.: Crime analysis and prediction using data mining, Date of conference: 19-20 August 2014, Location of conference: Guntur, India, Proceedings of International Conference on Networks & Soft Computing, IEEE, 2016, pp. 406-412.

[9] Nath, S. V.: Crime pattern detection using data mining, 18-22 December 2006, Hong Kong, China, Proceedings of the International Conference in Web intelligence and intelligent agent technology, IEEE, 2007, pp. 41-44.

[10] Zhao, X, Tang, J.: Exploring Transfer Learning for Crime Prediction, 18-21 November 2017, New Orleans, LA, USA, Proceedings of the International Conference on Data Mining Workshops, IEEE, 2017, pp. 1158-1159.

[11] Priyatharishini, M, Nirmala Devi, M.: A deep learning based malicious module identification using stacked sparse autoencoder network for VLSI circuit reliability, Measurement, 2022, Vol. 194(111055).

[12] Shamsuddin, N. H. M., Ali, N. A., Alwee, R.: An overview on crime prediction method, 23-24 May 2017, Johor, Malaysia, 2017 6th ICT International Student Project Conference, IEEE, 2017, pp. 1-5.

[13] Sivaranjani, S, Sivakumari, S, Aasha, M.: Crime prediction and forecasting in Tamil Nadu using clustering approaches, 21-22 October 2016, Kollam, India, Proceedings of the International Conference on Emerging Technological Trends, IEEE, 2016, pp. 1-6.

[14] Kansara, Chirag, E.: Crime mitigation at Twitter using Big Data analytics and risk modelling, 23-25 December 2016, Jaipur, India, Proceedings of the IEEE International Conference on Recent Advances and Innovations in Engineering, IEEE, 2017, pp. 1-8.