International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Survey Paper Volume-10, Issue-4 ,2023 Regular Edition E-ISSN: 2349-7084

A Study on Crime Data Visualization and Trends Forecasting

Bhargavi Akula[1*], V. Saritha[2], N. Sai Lohitha [3]

[1*] Computer Science and Engineering, Sri Padmavati Mahila Visva Vidyalayam, SPMVV, Tirupati, India
[2] Computer Science and Engineering, Sri Padmavati Mahila Visva Vidyalayam, SPMVV, Tirupati, India
[3] Computer Science and Engineering, Sri Padmavati Mahila Visva Vidyalayam, SPMVV, Tirupati, India
e-mail: bhargaviakula20@gmail.com , psarithakrishna@gmail.com, sailohitha.n@gmail.com

*Corresponding Author: bhargaviakula20@gmail.com,

https://doi.org/10.22362/ijcert/2023/v10/i04/v10i0402

Received: 20/03/2023, Revised:12/04/2023, Accepted: 23/04/2023 Published:28/04/2023

Abstract:- Criminal activity is one of our society's most serious and ubiquitous challenges, making crime prevention an essential endeavour. There are several types of crimes. The evolution of both technologies and society has had a positive impact on both crime and the harm it causes. Although many cities have adopted various crime prevention measures, it has become much more challenging to prevent in places with higher population density and rapid change. But criminal activity can happen everywhere. The key goal is to figure out how often they occur so that we can reduce the crime rate. Crime investigation or analysis is seen as a significant application for improving society's standard of living. The goal of this study is to gather data on crimes that have already occurred and the various algorithms that have been used to achieve improved accuracy. Criminal case resolution can be sped up with the help of crime data analysis, according to law enforcement officials.

Keywords: Criminal activities, Visualization, Random forest algorithm, Decision tree, Support Vector Machine algorithm, Naive Bayes algorithm, K-Nearest Neighbous Algorithm.

1. Introduction

The main danger to humanity is crime. Many different crimes happen frequently. Crimes might be growing, spreading and expanding swiftly and broadly. Crime activities occurs in all types of communities, including small towns, big cities, and rural areas. There are many different forms of crimes, such as robbery, homicide, rape, false imprisonment, assault, and battery [2][3]. When the personal or professional space of the perpetrator and the target intersect at one location, crime has occurred [1].

A single person, a group of people, or even an entire territory could be the target. Crime may occur accidentally or on purpose. Accidental crime is regrettable and unplanned. Accidental crime can happen anywhere. A group of people argue with one another over a trivial issue that could hurt those who have no connection to it. A planned crime is one that is carried out on purpose. The individual who intends to commit a crime first investigates the target or the target region and studies it in order to carry out the crime. Crime is more likely to happen in isolated areas since there are fewer police patrols there [1].

Overall, there are many reasons why crimes occur, including individual motives, human nature and behaviour, difficult conditions, and poverty. A rise in violent crimes may also be influenced by a number of other problems, such as illiteracy, gender inequality, high rates of unemployment, and a dense population. The increased crime rates linked to many surroundings, including commercial structures and municipal dwelling areas, are strongly correlated with the

expanding and populous cities [6]. Due to the rise in crimes, cases must be resolved significantly more quickly [2][3].

The police department's task is to manage and lessen the crime activities, which have risen in number and intensity at an increasingly rapid rate. The police department's main issues are Due to the abundance of available crime data, it is possible to forecast crimes and identify criminals. A technological solution is required in order to speed up case solving. I decided to do some study on how to solve a crime case as a result of the aforementioned issue [2][3].

Random Forest is one of the Machine Learning Algorithm used for both Classification and regression problem, algorithm is proposed for Classifying the different types of crimes which Separate Crime Categories with the Decision tree. Decision trees compares the crimes, if it is belonging to specific category crime. Then assigns to that category. Hence, Random Forest can be helpful in identifying, classifying the crimes into categories [5].

Decision Tree is a part of Random Forest, although the actual Random Forest is a collection of several Decision Trees. used for larger data sets. But a Decision Tree is a straightforward decision-making diagram for a single, smaller data set. With Single Decision tree, overfitting is a possibility. Accuracy rate is lower than that of the Random Forest algorithm [1].

Logistic Regression is one of the most used machine learning algorithm. Which is used to forecast categorical

dependent variables. It provides probabilistic values in the range of 0 and 1, which are used to address classification issues[5].

Support Vector Machine used for both regressions and classification problems. It is utilised primarily for classification issues. Every data point in SVM is plotted with the value of the specific coordinates. With that plotting's the classification can be carried out using the hyper plane that also distinguished the two classes [3].

The K-Nearest Neighbours technique groups new instances in the categories that are most similar to the existing categories, presuming that the new case/data and previous cases are comparable. Each piece of data is stored, and new data is categorised according to how similar it is to the current data. This indicates that by using the K-NN technique, new data can be quickly categorised using the underlying data. The fact that information is saved and retrieved later rather than immediately learning from the training set gives rise to the term "lazy learner algorithm" for this technique [3].

Gradient Boosting Algorithm is one of the machine learning algorithms to minimise bias mistakes; it is used for error prediction in both regression and classification issues. For regression problems, mean square error is used as the cost function, and for classification problems, log loss is utilised.

The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms available at the moment. It facilitates the quick development of machine learning models with accurate prediction capabilities. As a probabilistic classifier, it bases its forecasts on the probability that a specific event will occur [1].

Long Short Term Memory is a specific branch of Recurrent Neural Network (RNN) architecture, which is more dependable and effective in addressing both long-term and short-term dependency issues. When there is a significant difference between current values and those that must exist in the future, it is highly helpful. LSTM contains feedback connections in contrast to conventional feed-forward neural networks. Both individual data points and complete data sequences can be processed by it.

The main contribution of the paper is in this study largely concentrates on predicting the type of crime that may occur given the location where it has previously occurred.

Rest of the paper is organized as follows, Section I contains the introduction of A Study on Crime Data Visualization and Trends Forecasting, Section II contain the Literature Survey on A Study on Crime Data Visualization and Trends Forecasting, Section III contain Survey Analysis on A Study on Crime Data Visualization and Trends Forecasting, Section IV contain Related Work on A Study on Crime Data Visualization and Trends Forecasting, section V contains the Conclusion and Future scope of A Study on Crime Data Visualization and Trends Forecasting, Section VI contains the References.

2. Literature Survey

The complexity of the crime dataset has recently been shown by crime analysis. The parties involved in law enforcement will benefit from this procedure in terms of apprehending criminals and directing crime prevention measures. From a tactical or strategic standpoint, having the ability to anticipate future crimes based on their location, pattern, and timing might be quite useful. The rising frequency of crime in modern society makes it difficult to reliably predict future crime with higher performance. As a result, crime prediction methods are crucial for predicting future crimes and lowering crime rates. A study to predict crime based on certain inputs is now being done by several researchers.

The systematic process of discovering and analysing patterns and trends in crime is known as "crime analysis and prevention." Due to the growing use of digital technologies, crime data analysts can help law enforcement officers expedite the process of conducting criminal investigations.

The authors, S. Ayeaha Tanveer, Bonthala Khyathi, Vaedhamsetty Guru Sri Latha, Thalla Jagadeeswar et al, in their paper "Crime Rate Prediction Using Machine Learning" have presented a system to identify the potential type of crime and the location where it occurred. For prediction, Random Forest classification algorithm and Decision Tress

Algorithm is used to find the patterns of the criminal activities in a particular area. Random Forest Algorithm is one of the Machine Learning Algorithm used for both Classification and regression problem, proposed for Classifying the different types of crimes which Separate categories with the Decision tree. Decision trees compares the crimes, if it is belonging to specific category crimes. Then assigns to that category. Hence, Random Forest can be helpful in identifying, classifying the crimes into categories Analysing, Data visualisation makes a data set easier to understand. The graph include bar, pie, line, and scatter diagrams. The classifier accuracy observed is 78.9% [7].

The authors, Vineet Jain, Ayush Bhatia, Ayush Bhatia, Yogesh Sharma et al. in their paper "Crime Prediction using K-Means Algorithm's have used K-means clustering algorithm to find out patterns from the crime dataset. K-Means clustering algorithm is distance based algorithm. Finding a point's distance from the closest centres using the

Euclidean distance metric determines whether or not it should be included in the cluster.

At the beginning of the procedure, the number of clusters is unknown. K-Means must therefore be applied in numerous iterations. The authors have used Rapid Miner tool for analysis because of its flexibility and scalability . The main aim of the analysis was to understand which year was the crime rate highest and lowest. Supporting this information, bar graphs are plotted for each cluster [9].

The authors Devendra Kumar Tayal, Arti Jain, Surbhi Arora , Surbhi Agarwal , Tushar Gupta, Nikhil Tyagi et al, in their paper "Crime Detection and Crime Identification in India using Data Mining technique" have used data mining using the WEKA® tool for applying K means on the crime dataset. After the cluster identification, Crime prediction is also done. For prediction, KNN classification is used. The classifier accuracy observed is 93.62 % and 93.99 %. Google maps 3.r3.epresentation is done to show the clusters over a map. However, the representation is not useful enough, since only markers over the map give information about the number of crimes took place in a vast area around the marker, not giving much information about the exact location of crime [10].

3. Survey Analysis

Table 1. Summary of the literature analysis

S.no	Title	Method	Accuracy	Conclusion
1.	Using Big Data Analysis for Developing Crime Prediction Model [2016]	Big Data Analytics, Artificial Neural Networks	-	This concept aids security officials in lowering crimes. The use of this prediction model by law enforcement in criminal investigations. Retrieving buried information through categorization is made easier and less time-consuming by combining big data analytics with a clustering approach.
2.	Crime prediction and analysis using Machine Learning [2018]	K-Nearest Neighbour, Support Vector Machine, Decision tree, Gaussian Naïve Bayes, Multinomial Naïve Bayes, Bernoulli Naïve Bayes	0.7873, 0.3135, 0.7860, 0.6460, 0.4562, 0.3135	This model predicts the crime type which may happen if we know the area of where it has occurred. The model has a 0.789% accuracy rate when predicting the crime kind. Compared to other algorithms K-Nearest Neighbour achieves high accuracy.
3.	Prediction of Crime rate analysis using	Decision tree, Random Forest Algorithm, K-	0.9800, 0.9800, 0.8673, 0.7971,	If we know the location of the incident, this model has the ability to predict the

	Supervised Classified Machine Learning approach [2019]	Nearest Neighbour, Logistic Regression, Support Vector Machine	0.5997	kind of crime that would happen there. The model has a 0.9800 accuracy rate for predicting the kind of crime. Compared to other algorithms Decision tree, Random Forest Algorithm achieves high accuracy.
4.	Big Data Analytics and Mining for effective visualization and trends forecasting of crime data [2019]	Long short term memory, Artificial Neural Networks, Prophet model	San-Francisco: 0.42, 0.62, 0.72; Chicago: 0.29, 0.50, 0.53; Philadelphia: 0.39, 0.65, 0.72	We examined crime statistics from three US cities (San Francisco, Chicago, and Philadelphia) using big data analytics and visualization tools, which allowed us to spot trends and patterns. Found that when compared to each other, a neural network model, and the Prophet model, the LSTM deep learning method performs better than conventional neural network models.
5.	Analysis and Prediction of Crime against Woman Using Machine Learning Techniques [2021]	Logistic Regression, Naïve Bayes, Decision tree, Random Forest Algorithm	0.807, 0.807, 0.769, 0.769	This model analyses crime data to predict the likelihood of a crime occurring in a city. The four implemented prediction models—Logistic Regression, Naive Bayes, Decision Tree method, and Random Forest Classifier—are compared for accuracy. Two algorithms receive accuracy scores of 0.769 and 0.807, respectively, on a scale from 0 to 1. Logistic Regression, Naive Bayes having the same accuracy of 0.807. Decision Tree algorithm, Random Forest Classifier are having the same accuracy of 0.769.
6.	Emperical Analysis for Crime Prediction and Forecasting using Machine Learning and Deep Learning Techniques. [2021]	Extreme Gradient Boosting (XGBoost), Decision tree, Logistic Regression, Random Forest Algorithm, Multi-layer Perceptron, Naïve Bayes, Support Vector Machine, K-Nearest Neighbour	C=Chicago: 94, 88, 66, 60, 90, 48, 77, 43, 87, 84, 73, 71, 66, 60, 88; L=Los Angeles: 89, 73, 63, 46, 85, 45, 78, 50, 89, 73, 63, 47, 60, 36, 64	The application of several machine learning methods, such as LSTM and ARIMA modelling, was investigated. The efficacy of different machine learning techniques, such as logistic regression, SVM, Naive Bayes, KNN, decision tree, MLP, random forest, and XGBoost, was first evaluated on datasets for Chicago and Los Angeles. Machine learning algorithm performance is more consistent for the Chicago dataset when compared to the Los Angeles dataset. XGBoost improves prediction accuracy efficiency (between 94% and 88%). KNN on both crime datasets (between 88% and 89%). The efficiency of prediction accuracy attained by various algorithms is comparable to that published earlier and predicts greater performance. According to this study, Chicago and Los Angeles had superior forecast accuracy, at 94% and 88%, respectively. XGBoost is best for Chicago. K-Nearest Neighbour is best for Los Angeles. The time series analysis in this

				makes use of LSTM and a deep learning architecture.
7	Crime rate prediction using Machine Learning. [2022]	Random Forest Algorithm, Decision Tree	0.789	This model predicts the potential crime's type and location. The model can classify the type of crime with an accuracy of 0.78.
8	A Survey on analysis of crime detection techniques using Machine Learning. [2022]	K-Nearest Neighbour, Decision Tree	-	In this work, a survey on crime analysis and prediction utilising machine learning has been undertaken. The survey discussed 5 types of crimes: detection of frauds (credit card, internet, insurance, cheque), violence of crimes (rape, murder), violence of traffic, cybercrimes, and sexual offenses (straightforward, with the use of a weapon, verbal assault, aggravated sexual assault). The performance of two machine learning algorithms, K-Nearest Neighbour and Decision Tree, for crime detection was evaluated.

4. Related Work

Crime is defined as "acts against the law which inflict harm on innocent persons and result in the imposition of penalties by the legal authorities, such as the law enforcement or judicial power of the government.". The primary types of crime include traffic infractions, fraud, sex crimes, arson, drug offences, violent crimes like murders and robberies, as well as property damage, theft, and cybercrime.

It is suggested to develop a system to identify the potential type of crime and the location where it occurred [2]. Most of the parameters are observed while studying the proposed system of the paper. The proposed architecture analysed crimes under different categories as well as different parameters. The listed are the parameters:

- Crime Types vs. Crimes Committed.
- The number of offences committed during the course of a year.
- The number of offences committed in an hour.
- The percentage of those arrested.
- Crimes perpetrated in various places.
- Information on serious crimes committed.

The results showed the percentage of most occurred crimes in the city by denoting the types of crimes and no. of crimes committed. According to that the arrest ratio in the city is 32.8%. Almost 800 crimes are committed in the month of march, at 12:00 noon. Highly occurred crimes are thefts and narcotics.

A proposed architecture to identify the potential type of crime and the location where it has occurred [3]. Most of the parameters are observed while studying the proposed system of the paper. The proposed architecture analysed crimes under different categories as well as different parameters. The listed are the parameters:

- Crimes categorised by total areas.
- Percentages of different crimes happening in the city.
- Classify the crime rate by PSA (Public Service Announcements).
- Classify the crime rate by year.

The results showed the percentage of most occurred crimes in the city by denoting the types of crimes and its area code. According to it, the Highly occurred crimes are Vandalism, thefts and their percentage are 17.37, 16.38 respectively. Averagely occurred crimes are Robbery, Weapon violation and their percentage are 4.96, 4.82 respectively. Least occurred crimes are Gambling, Homicide and their percentage are 0.92, 0.86 respectively. Crime rate is shown by PSA which shows the high happened crimes and low happened crimes (%) and crime rate by year.

It is suggested to develop an architecture to identify the potential type of crime and the location where it occurred [4]. In this the study three different states in United States are considered. The proposed architecture analysed crimes under different categories as well as different parameters. The listed are the parameters:

- Visualization of crime activity for the cities of (a) San Francisco, (b) Chicago, and (c) Philadelphia on open street maps.
- Crimes committed over time.
- Top 10 crimes of cities.
- The hourly crime trend in each city.

The results showed that the percentage of most occurred crimes in the city by denoting the types of crimes and no. of crimes committed. The top ten offences that happened most frequently were the system's major emphasis. Each city faces the serious threat of theft. All three cities exhibit comparable patterns in crimes on an hourly basis. The safest time of day is between 5 and 6 pm, whereas the hour with the most recorded crimes is between midnight and two in the morning. 12:00 PM is also very dangerous.

It is suggested that an architecture be used to identify the potential type of crime and the location where it has already occurred. [6]. Most of the parameters are observed while studying the proposed system of the paper. The proposed architecture analysed crimes under different categories as well as different parameters. They listed are the parameters:

- Time series analysis for the daily, weekly, monthly, quarterly, and yearly mean crime density areas, quarterly, and yearly.
- Analysis of the daily, monthly, and yearly trends in crime.
- Locations with high crime rates and crime counts.

The results showed percentage of most occurred crimes in the city by denoting the crime count and its location. The system focused mainly on occurred crimes in two states. Harrison in Chicago has highest crime count and Los Angeles has highest crime rate. Mean crime type resampled is reducing over day by day as months and years passed. As compared to Los Angeles and Chicago, Los Angeles have the highest crime rate.

5. Conclusion and Future Scope

Crimes pose significant risks to human society. Control is intended for both sustainability and safety. To further strengthen city security and safety and aid in crime prevention, investigation authorities frequently request computational forecasts and predictive systems that increase crime analytics. With the use of machine learning technologies, identifying links and patterns among various types of data has gotten considerably simpler. This study largely concentrates on predicting the type of crime that may occur given the location where it has previously occurred. Every algorithm has its own pros and cons. In order to get the better accuracy levels, a novel method is proposed by considering few of the above mentioned algorithms.

References

[1] Tirthraj Chauhan , Rajanikanth Aluvaiu , " Using Big Data Analysis for Developing Crime Prediction Model",RK University, ISBN: 978-93-5254-061-7, Jan. 5,6 2016.

[2] Alkesh Bharati, Dr Sarvanaguru RA . K, " Crime Prediction analysis using Machine Learning " IRJET, Vol. 05, Issue: 09, e- ISSN: 2295-0056, p- ISSN: 2395-0072, Sep 2018.

[3] Kirthika V , Krithika Padmanabhan A, Lavanya M, Lalitha S D, "Prediction of Crime rate analysis using Supervised Classified Machine Learning approach" IRJET, Vol. 06, ISSN: 2295-0056,p- ISSN: 2395-0072, march 2019.

[4] Mingchen Feng; Jiangbin Zheng; Jinchang Ren; Amir Hussain; Xiuxiu Li; Yue Xi; Qiaoyuan Liu, " Big Data Analytics and Mining for effective visualization and trends forecasting of crime data ", Date of Publication **:** 22 July 2019 , Electronic ISSN: 2169-3536 , INSPEC Accession Number : 18897693 , August 15, 2019.

[5] Prasad DS, Rachit Sharma, Dr. V. Anbarasu, " Analysis and Prediction of Crime Agains Women Using Machine Learning Techniques", vol. 25, Issue 6, 2021.

[6] Wajiha Safat, Soahail Asghar, (Member,IEEE), and Saira Andleeb Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques". May 17, 2021. DOI:10.1109/ACCESS.2021.3078117.

[7] S. Ayeaha Tanveer, Bonthala Khyathi, Vaedhamsetty Guru Sri Latha, Thalla Jagadeeswar "Crime Rate Prediction Using Machine Learning ", IJRPR, ISSN- 2582-7421, Vol. 3, no. 6, pp 3089-3093, June 2022.

[8] Ashish Kumar, Kaptan Singh, Amit Saxena, "A Survey on analysis of crime detection techniques using Machine Learning". Vol. 10, Issue: 02, e-ISSN: 2347-2693. Feb 2022.

[9] Vineet Jain, Ayush Bhatia, Ayush Bhatia, Yogesh Sharma " Crime Prediction using K-Means Algorithm's" GRD Journal for Engineering, Volume 2, Issue 5, April 2017, ISSN: 2455-5703.

[10] Devendra Kumar Tayal, Arti Jain, Surbhi Arora ,Surbhi Agarwal , Tushar Gupta, Nikhil Tyagi " Crime Detection and Crime Identification in India using Data Mining technique ". DOI 10.1007/s00146-014-0539-6.

