

Crime Type and Occurrence Prediction Using Machine Learning Algorithm

P. Kalyan Chakravarthi^{1*}, M. Sai Sumanth², G. Sunil Kumar³, P. Ajay Kumar⁴, M. Viswa Pranay⁵

Abstract:

As of late, crime has turned into an unmistakable way for individuals and society to experience difficulties. An irregularity in a country's populace happens when crime goes up. To assess and manage this sort of crime, it's vital to know how crime designs change over the long time. This study utilizes crime information from Kaggle open source to do a sort of crime design examination. The information is then used to think about what crime will occur straightaway. The central matter of this study is to figure out what sort of crime has the greatest effect, as well as when and where it worked out. In this work, machine learning techniques like Nave Bayes are utilized to bunch different criminal patterns into gatherings. The outcomes were exact contrasted with other comparative works.

Keywords - Crime, Analyse, Crime patterns, Kaggle, Estimate, Naïve Bayes, Accuracy.

1. INTRODUCTION

Crime has turned into a significant constrained string that is probably going to worse. As per the definition, conduct is a crime when it disrupts the norm, is illegal, and is extremely terrible. For the Crime design examination to work, analysts need to investigate numerous areas of criminal science and track down similitudes. The public authority needs to invest a great deal of energy and exertion into utilizing innovation to stop a portion of these unlawful activities. Along these lines, machine learning frameworks and their information are expected to anticipate the kinds of Crime that will occur and how they will occur. It utilizes something like date crime information to figure the sort of Crime that will occur and how frequently it will happen in light of overall setting. Heaps of studies have been finished by specialists to assist them with sorting out how crime percentages change after some time and how they connect with one another in a specific spot. A portion of the spots that were taken a gander at have made it more straightforward to bunch wrongdoing patterns into various gatherings. This makes it more straightforward for the public authority to take care of issues quicker. For this technique, a dataset from Kaggle open source is utilized in view of various elements, including the general setting where the occasion happens during a specific time span. We proposed a method for gathering things into bunches that would assist with peopling sort out what sort of wrongdoing is occurring and where it's going on most frequently at specific times. In this proposed one, machine learning strategies are utilized to find criminal patterns that match and assist with placing them into bunches utilizing the given geographical and temporal data.

The fact that a crime is makes an exhibit all. It against the law against the law to do that. It's illegal. It is exceptionally difficult for the police to find and concentrate on crime that are occurring on display. There is likewise a great deal of data about the crime . In this way, there ought to be a few strategies that can assist with the exploration. The technique ought to assist with viewing as the criminal.

The objective of this study is to take a gander at ongoing crime patterns utilizing Kaggle open source crime information and track down the most widely recognized kinds of crime , when they occur, and where they happen most frequently. Utilizing machine learning techniques, chiefly Naive Bayes, we need to precisely bunch various sorts of unlawful way of behaving into bunches that will help individuals comprehend and manage crime better.

Corresponding Author: P. Kalyan Chakravarthi

1. Assistant professor, Department of CSE (AI/ML &CS), Godavari institute of engineering and technology.

Email: viitchakri@gmail.com

2. Department of CSE (AI/ML &CS), Godavari institute of engineering and technology.

Email: sumanthchowdary647@gmail.com

3. Department of CSE (AI/ML &CS), Godavari institute of engineering and technology.

Email: sunilgolla123@gmail.com

4. Department of CSE (AI/ML &CS), Godavari institute of engineering and technology.

Email: ajaykumarpasupula7@gmail.com

5. Department of CSE (AI/ML &CS), Godavari institute of engineering and technology.

Email: pranay.muddana@gmail.com

An ever-increasing number of violations are going on, which is a major issue for the two individuals and society in general. Kaggle's open-source crime information is utilized in this review to check out and think about what violations will occur sooner rather than later. The primary objective is to find the sort of crime that happens most frequently, when it works out, and where it works out. Machine Learning, like Naive Bayes, is utilized to bunch different criminal ways of behaving into bunches precisely.

2. LITERATURE REVIEW

Crime Analysis Through Machine Learning:

This paper investigates how machine learning can be utilized to predict crimes. This paper utilizes two distinct ways of seeing crimes figures from Vancouver throughout recent years. Two machine learning models — K-nearest-neighbor and supported choice tree — are utilized to predict crime in Vancouver. The accuracy of these models' accuracy goes from 39% to 44%.

SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES:

The course of data mining includes glancing through enormous data sets that as of now exist to find new data that could be valuable to the business. Utilizing the accessible datasets, we can think about how to get new data. Many approaches to dissecting and causing expectations to have been utilized in data mining. However, not much of the work has been finished in the space of criminology. A tiny number of individuals have attempted to look at the data that these strategies give. Police stations and other comparative law enforcement offices keep a ton of information in huge data sets that can be utilized to suppose or concentrate on how crooks move around and how they communicate with society. The crime records can likewise be utilized to think about who the criminals are. The primary objective of this work is to do a survey of the managed and solo learning techniques that have been utilized to distinguish criminals. This exposition gives an outline of how to utilize various Data Mining strategies to study and predict crime.

A Survey on Crime Occurrence Detection and prediction Techniques:

There should be a framework that can find and recognize crimes progressively on the grounds that the quantity of crimes is going up. The objective of this overview is to investigate Data Mining strategies that can help find and distinguish crimes. These techniques incorporate affiliation rule mining, k-means clustering, decision trees, naive bayes, and machine learning strategies like deep neural networks and artificial neural networks. The survey results showed that pre-processing is vital when there is a great deal of missing qualities in the dataset and that crime doesn't occur equally across urban communities yet rather centers in certain areas. Thus, sorting out where crimes are probably going to happen is vital work. Utilizing post-processing will likewise assist with bringing down the crime rate.

SMART RESOURCE USAGE PREDICTION USING CLOUD COMPUTING FOR MASSIVE DATA PROCESSING SYSTEMS:

Asset the executives is a significant piece of cloud computing on the grounds that the requirement for Resource management of data processing is filling rapidly in regions like healthcare, business solutions, and the internet of things. Assigning the right and amazing assets is as yet the principal justification for why applications can be figured effectively. Traditional methods for overseeing assets, then again, which depend on straightforward heuristics, don't work since they don't create solid outcomes. Predicting the perfect proportion of work to do and assigning assets accurately are both connected. Assigning assets accurately will assist with reducing down on expenses, time, power, and memory use. The paper recommends a method for utilizing machine learning to oversee how assets are designated in distributed computing for an exceptionally enormous information processing framework. By recreating the proposed model utilizing the organization test system 2, it is feasible to get better execution and asset use while utilizing less cash, time, power, and memory.

Crime Pattern Detection, Analysis & Prediction using Machine Learning:

The orderly course of criminal analysis is to find and concentrate on examples and patterns in crime. As computers become more normal, crime information specialists can assist police with addressing crimes quicker. Utilizing the possibility of "data mining," a framework can glance through unorganized data and find valuable data that wasn't there previously. Utilizing scientific and predictive strategies to find lawbreakers is what's really going on with prescient police, and functioning admirably has been shown. Since the crime percentage has gone up throughout the long term, the framework should manage an enormous measure of crime information put away in distribution centers, which

would be difficult to break down the hard way. Additionally, criminals are getting better at utilizing innovation, so police need to utilize new advances to remain in front of them. The central matter of this study is to take a gander at the calculations and strategies that are utilized to track down criminals.

3. METHODOLOGY

Existing System:

During pre-work, the open-source dataset is first pre-processed to eliminate copy values and elements. A decision tree has been utilized to remove qualities from massive volumes of information and to recognize criminal patterns. It gives the fundamental construction to the course of later classification. Highlights are removed from the sorted crime designs utilizing a deep neural network. The prediction decides the presentation for both preparation and test values. Crime prediction assists specialists with expecting future occurrences of criminals and facilitates the goal of such episodes.

Disadvantages:

1. The ongoing framework makes reference to pre-processing's evacuation of repetitive elements and values, yet it doesn't frame the techniques or calculations utilized. It's obscured the way that successful the pre-processing is a direct result of this absence of data.
2. Although the ongoing framework examines crime prediction, it overlooks data with respect to how it predicts criminal activity, which is a fundamental part of crime prediction frameworks.
3. It is trying to assess the similarity and versatility of the ongoing framework since it makes no reference of the innovative stack that was used for execution.

Proposed System:

The procured information is first pre-processed utilizing a machine learning calculation channel and covering to eliminate repetitive and pointless data factors. Moreover, it diminishes dimensionality, an indication of cleaned information. The information then, at that point, goes through a parting cycle. The preparation and test informational collections are isolated into two classifications. The model is prepared utilizing both preparation and testing datasets. Then comes the planning. To work with arrangement, numbers are allotted to the crime type, year, month, time, date, and area. In the first place, Nave Bayes is utilized to analyze the characteristics' free impact. The autonomous traits that were recovered are characterized utilizing Bernouille Nave Bayes. By recognizing the criminal components, examining the commonness of wrongdoing in a specific area and timeframe is conceivable. Finally, the most incessant offenses are found alongside spatial and worldly information. The forecast model's presentation is surveyed utilizing the exactness rate. Python was utilized to foster the prediction model, which was then executed on Colab, an internet-based compiler for machine learning and information examination.

Advantages:

1. The proposed framework pre-processes information utilizing a machine learning approach, which can possibly be more productive than conventional methods in taking care of complicated information and recognizing repetitive qualities.
2. The framework expresses that by zeroing in on pertinent elements during pre-processing, it limits dimensionality, which can expand the viability of later examination and modeling.
3. The utilization of the notable machine learning strategy Naive Bayes for order undertakings can offer light on the many impacts that every trademark has on crime prediction autonomously.

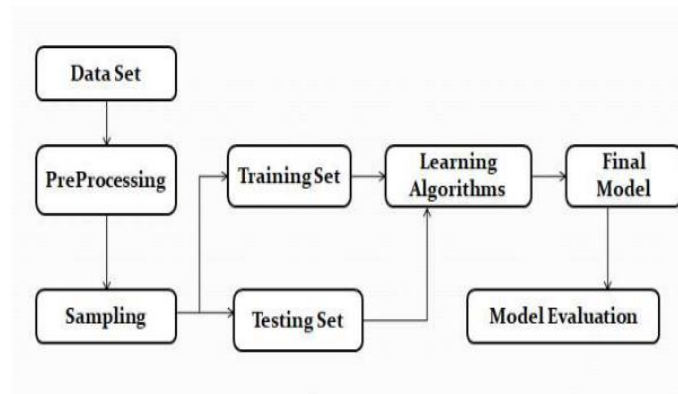


Fig 1 System Architecture

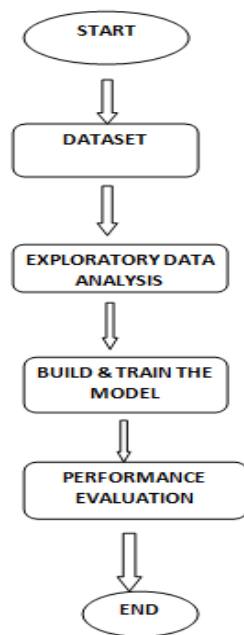


Fig 2 Block Diagram

Modules:

We have used the accompanying modules to complete the previously mentioned project.

- Data exploration: this module will be utilized to stack information into the framework.
- Processing: information will peruse for process.
- Splitting data into train & test: information will be separated into train and test utilizing this module.
- Model generation: Model building - Naive Bayes - Gaussian NB - Multinomial NB - Bernoulli NB - Decision Tree - KNN - DL (CNN) - Gradient Boosting - Voting Classifier (RF + AdaBoost)
- Users sign up and login: This module gathers enrollment and login information.
- User input: This module gathers client information for expectation.
- Prediction: the last expectation is shown.
- As an expansion, we utilized an outfit approach that joined the predictions of a few unique models to give a last prediction that was more reliable and accurate. We may, nonetheless, further develop execution much further by exploring more outfit procedures, for example, Voting Classifier, which yielded test exactness, accuracy, precision, and F1Score scores of 94%, 97%, 95%, and 94%, separately.

4. IMPLEMENTATION

Here in this project, we are using the following algorithms.

KNN: The calculation for K-Nearest Neighbors. The k-nearest neighbors' calculation, in some cases alluded to as KNN or k-NN, is a non-parametric supervised learning classifier that bunches individual information direct in view of closeness all together toward characterize or foresee them.

Voting Classifier (RF + AdaBoost): or Voting Classifier (Stomach muscle + RF): A Voting classifier is an machine learning assessor that totals the consequences of each base assessor to make expectations in the wake of preparing various base models or assessors. Voting a ballot decision for every assessor result can be converted to shape the aggregating criteria.

Naive Bayes: The Bayes hypothesis fills in as the establishment for this clear probabilistic characterization calculation. It makes the straightforward however fundamental assumption that the highlights utilized for arrangement are free. In view of the probabilities of individual elements, it decides the most probable class and figures the likelihood of a particular occasion or class.

Gaussian NB: A variety of the Naive Bayes characterization method is called Gaussian Naive Bayes (GNB). Predicated on the thought highlights have a Gaussian (normally) circulation and is planned for constant information. For characterization issues, GNB utilizes the assessed mean and difference of each element for each class to produce probabilistic predictions.

Multinomial NB: Explicitly intended for text investigation applications like spam distinguishing proof and record order, Multinomial Naive Bayes (MNB) is a grouping strategy that functions admirably with discrete information. Expecting a multinomial conveyance, it reenacts the probability of tracking down each term's recurrence in records. At the point when text information is displayed as term frequencies or TF-IDF values, MNB is often used.

Bernoulli NB: A famous grouping calculation for parallel information, Bernoulli Naive Bayes (BNB) is utilized for message characterization in feeling examination and spam discovery, among different applications. It addresses qualities as parallel factors (like a word's presence or nonattendance). BNB gauges class probabilities utilizing Bernoulli appropriations under the presumption that every trademark is restrictively free.

Decision Tree: A supervised machine learning approach for grouping and relapse issues is known as a decision tree. Recursively partitioning the dataset as per include properties, it builds a tree-like model determined to boost data gain or limiting pollutant at every hub. It offers a fathomable, straightforward structure for simply deciding.

DL (CNN): Convolutional neural networks, or CNNs, are an intense machine learning innovation that are generally used for picture examination and acknowledgment. CNNs are excellent at errands like article location, picture arrangement, and facial acknowledgment since they naturally separate various levelled highlights from pictures utilizing many layers of convolution and pooling processes.

Gradient Boosting: For relapse and order, slope helping is a machine learning gathering strategy. It continuously consolidates a few frail students (normally decision trees), every one of which fixes the slip-ups of the one preceding it. By shifting the loads of preparing cases, it streamlines a misfortune capability and produces a strong forecast model. XGBoost and LightGBM are popular methodologies.

5. EXPERIMENTAL RESULTS

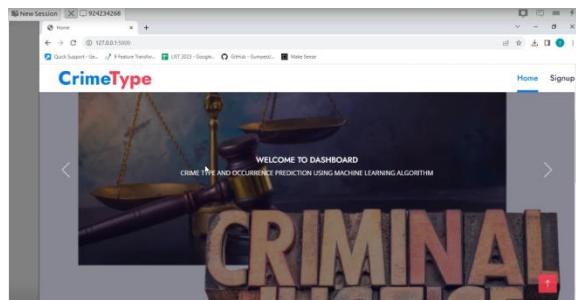


Fig 3 Home Page

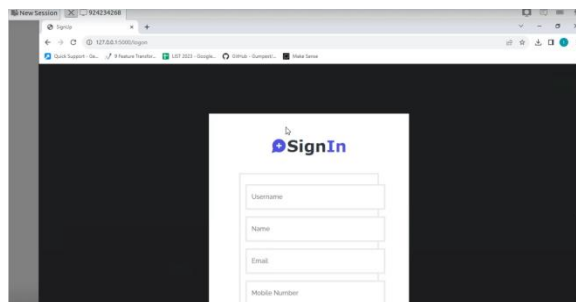


Fig 4 Signup Page

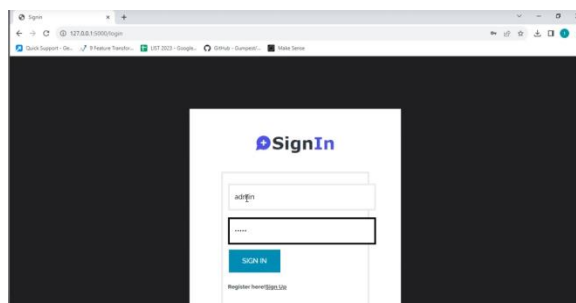


Fig 5 Signin Page

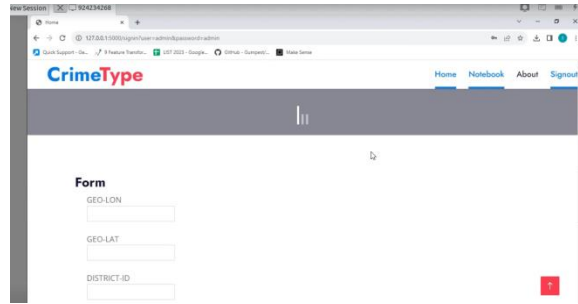


Fig 6 Main Page

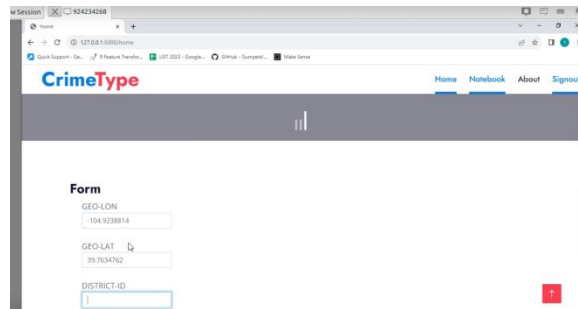


Fig 7 Upload Input Values

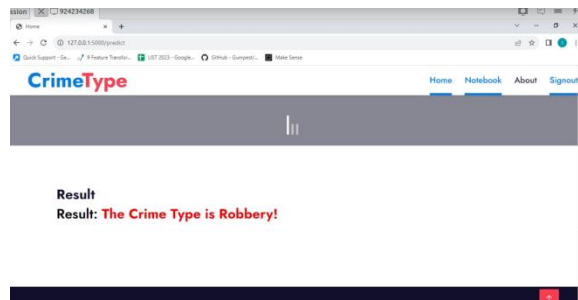


Fig 8 Prediction Result

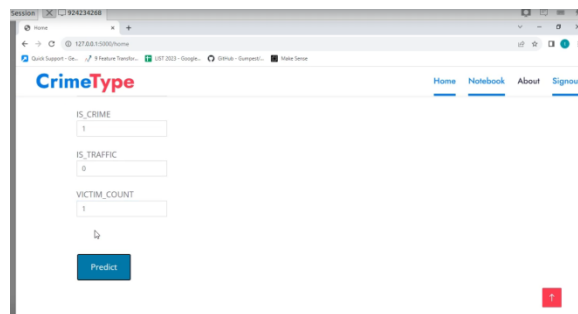


Fig 9 Upload another input values

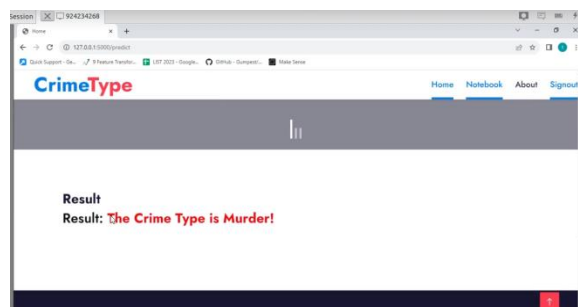


Fig 10 Prediction Result

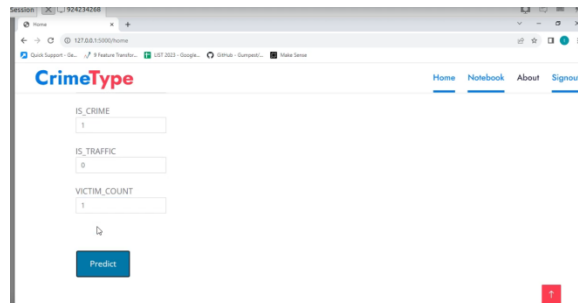


Fig 11 Upload input values

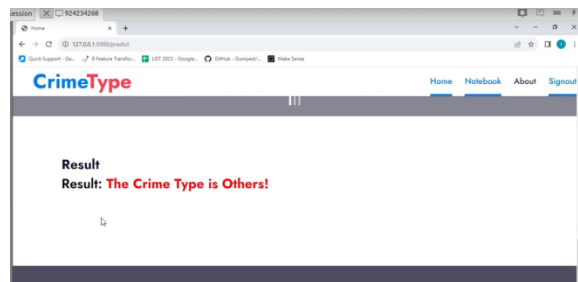


Fig 12 Prediction Result

6. CONCLUSION

In this review, two classifiers — Multinomial NB and Gaussian NB — are utilized to resolve the issue of taking care of the ostensible conveyance and genuine esteemed properties. It is ideal for ongoing estimating in light of the fact that it doesn't need a ton of preparing time. Moreover, it settle the test of managing a persistent goal set of factors, which the past work couldn't do. In this way, Nave Bayesian Characterization could be utilized to expect and distinguish the most well-known violations. Moreover, the calculation's presentation is processed utilizing various standard measurements. The actions utilized in the calculation evaluation are exactness, F1 score, normal precision, and review. Accuracy could be extraordinarily expanded by incorporating ML strategies.

Despite the fact that it settle the issue from the previous work, it has a ton of impediments. The gauge has no likelihood of being precise without even a trace of class assignments. Utilizing more ML order models upgrades by and large execution and raises wrongdoing expectation precision as a likely expansion of the proposed work. By taking into account neighborhood area pay measurements to decide if there is a connection between a local's pay levels and crime percentage, it assists with giving better review to future upgrades.

REFERENCES

1. Suhong Kim, Param Joshi, Parminder Singh Kalsi, Pooya Taheri, "Crime Analysis Through Machine Learning", IEEE Transactions on November 2018.
2. Benjamin Fredrick David. H and A. Suruliandi, "Survey on Crime Analysis and Prediction using Data mining techniques", ICTACT Journal on Soft Computing on April 2012.
3. Shruti S. Gosavi and Shraddha S. Kavathekar, "A Survey on Crime Occurrence Detection and prediction Techniques", International Journal of Management, Technology And Engineering, Volume 8, Issue XII, December 2018.
4. Chandy, Abraham, "Smart resource usage prediction using cloud computing for massive data processing systems" Journal of Information Technology 1, no. 02 (2019): 108-118.
5. Learning Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimparia, "Crime Pattern Detection, Analysis & Prediction using Machine", International Research Journal of Engineering and Technology, (IRJET) e-ISSN: 2395-0056, Volume: 07, Issue: 06, June 2020
6. Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi, "Spatio-Temporal Crime Hotspot Detection and Prediction: A Systematic Literature Review", IEEE Transactions on September 2020.
7. Nasiri, Zakikhani, Kimiya and Tarek Zayed, "A failure prediction model for corrosion in gas transmission pipelines", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, (2020).
8. Nikhil Dubey and Setu K. Chaturvedi, "A Survey Paper on Crime Prediction Technique Using Data Mining", Corpus ID: 7997627, Published on 2014.

9. Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, "Computational System to Classify Cyber Crime Offenses using Machine Learning", Sustainability Journals, Volume 12, Issue 10, Published on May 2020.
10. Hyeon-Woo Kang and Hang-Bong Kang, "Prediction of crime occurrence from multimodal data using deep learning", Peerreviewed journal, published on April 2017.