



**ISSN: 2454-9940**



**INTERNATIONAL JOURNAL OF APPLIED  
SCIENCE ENGINEERING AND MANAGEMENT**

**E-Mail :**  
**editor.ijasem@gmail.com**  
**editor@ijasem.org**

**[www.ijasem.org](http://www.ijasem.org)**

# CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING

---

<sup>1</sup>MR.A N L KUMAR, <sup>2</sup>ANANTHA SRI SANTOSH KUMAR

<sup>1</sup>(Head of the Department MCA), MCA, Swarnandhra College

<sup>2</sup>MCA, Scholar, Swarnandhra College

## ABSTRACT

Crime has become a clear sign that individuals and society are in crisis in our modern day. There will be a shift in a country's demographics as a result of rising crime rates. It's important to comprehend crime trends in order to analyze and respond to this kind of criminal activity in advance. This research implements a crime pattern analysis by making use of open-source crime data from Kaggle to forecast future criminal activity. Estimating the most significant form of crime, as well as the time period and place in which it occurred, is the primary focus of this study. This study makes use of machine learning methods like Naïve Bayes to categorize different types of crime, and the results were rather accurate when compared to previously written works.

## 1. INTRODUCTION

A key thread that is thought to be growing rather intensely is crime, which has just emerged. When something is both very offensive and in violation of government laws, we say that it is a crime. Research into the several subfields of criminology and the art of pattern recognition are prerequisites for any crime pattern analysis. In order to regulate some of these illegal activities, the government must invest significant time and effort on proposing technological solutions. Therefore, in order to forecast the kind and frequency of crimes, machine learning methods and their associated data are necessary. It makes use of preexisting crime data to forecast the kind and frequency of crimes based on time and place. Researchers have conducted a plethora of research that shed light on crime trends and their

connections in a given area. An simpler method of categorizing criminal trends has emerged in several of the examined regions. Helping the authorities to resolve things quickly is the result of this. This method makes use of a dataset retrieved from the open-source Kaggle platform, which is dependent on a number of variables, including the time and location of occurrence over a certain duration. In order to determine the nature of the crime and the locations most prone to criminal activity at a given time and day, we included a categorization algorithm. This one suggests using machine learning algorithms to scour available geographical and temporal data for crime trends that satisfy the criteria for each category.

## 2. LITERATURE SURVEY

**Suhong Kim and Param Joshi [1]** two distinct machine learning models that are used for prediction, the K-nearest-neighbor method (KNN) 978-1-7281-95377/21/\$31.00 ©The year 2021 This is the IEEE 266th Annual International Conference on Intelligent Systems and Computing, published on April 12, 2021. and a method based on decision trees. When

it comes to forecasting crime trends and determining the sort of crime, the acquired accuracy falls anywhere between 39 and 44 percent.

**Benjamin Fredrick David. H [2]** forced a data mining method that examines and evaluates massive existing datasets in order to provide additional information. It is common practice to compare newly extracted patterns to existing, predetermined datasets. Shraddha S.

**Kavathekar [3]** included association rule mining into crime prediction. Deep Neural Networks (DNNs) and Artificial Neural Networks (ANNs) are two of the machine learning techniques that have been mentioned. When fed a dataset at the feature level, a deep neural network performs better. For multi-labeled data categorization in particular, the prediction model has been constructed using DNN using fully linked convolution layers. Tenserflow, an API primarily for deep learning techniques with dropout layers, was used to construct it. The foundation of Artificial Neural Networks (ANNs) for problem-solving is trend-analysis forecasting. As a whole, its vast

processing components contribute to the development of a model.

**Chandy and Abraham [4]** provided a random forest classifier that could be used to extract features for data processing in the cloud. Request number, user ID, expiration time, arrival time, and memory need are the characteristics that have been retrieved. Following feature extraction, the trained data observed during the learning step is used for work load prediction. This data helps the system to understand the specifics of the characteristics that were extracted from the user's request.

### **3.EXISTING SYSTEM**

The initial step in preprocessing a dataset is to identify and delete duplicate values and characteristics. This is done with datasets that are received from free sources. Among the many applications of decision trees is the identification of criminal trends and the extraction of characteristics from massive datasets. It lays the groundwork for further categorization. A Deep Neural Network is used to extract features from the identified criminal patterns. Both the training

and test values; performance is computed using the prediction. Officials may use crime prediction to get a head start on solving any kind of crime by seeing it coming up in the future.

### **DISADVANTAGES**

1. The classifier in previous studies employs categorical values, which leads to a biased conclusion

for the nominal qualities with higher value, which explains why the accuracy is poor.

2. Areas with unsuitable data and actual valued characteristics are not suitable for the classification

procedures.

3. The classifier's value has to be adjusted, thus it's important to choose the best one.

### **3.1 PROPOSED SYSTEM**

❖ In order to eliminate unnecessary and repetitive data values, the collected data is first pre-

processed using machine learning techniques such as filter and wrapper.

Additionally, it cleans up the data by reducing the dimensionality. A further splitting technique is applied to the data. The data is divided into two categories: test and training. Both the training and testing datasets are used to train the model. After that, mapping comes next. In order to facilitate categorization, the following fields are translated to integers: crime category, year, month, time, date, and location.

❖ The first step is to use Naïve Bayes to analyze the independent influence between the characteristics. To categorize the retrieved independent characteristics, Bernoulli Naïve Bayes is

used. By labeling the criminal aspects, it is possible to analyze the crime rate at a certain time and place. In the end, we learn about the most common crimes, along with their geographical and temporal occurrences. The accuracy rate is a measure of the prediction model's performance. Colab is an online compiler for data analysis and machine learning models. The prediction model was designed using the language Python and ran on it.

## THE BENEFITS

1. Because most of the included properties rely on the time and place, the proposed method is

highly suited for crime pattern identification.

2. It also gets around the issue of evaluating the qualities' independent effects.

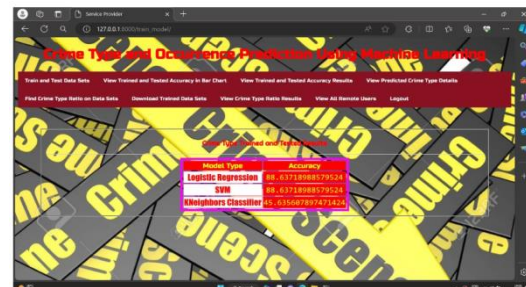
3. We don't need to initialize the ideal value since it takes into consideration real and nominal

values and also the area with incomplete data.

4 In comparison to other models of machine learning prediction, the accuracy has been comparatively good.

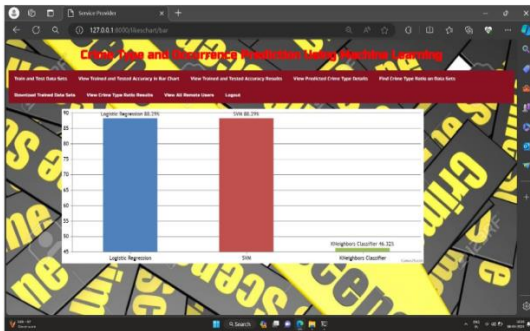
## 4. OUTPUT SCREENS

### Train and Test Data Sets



### View Trained and Tested Accuracy in Bar Chart

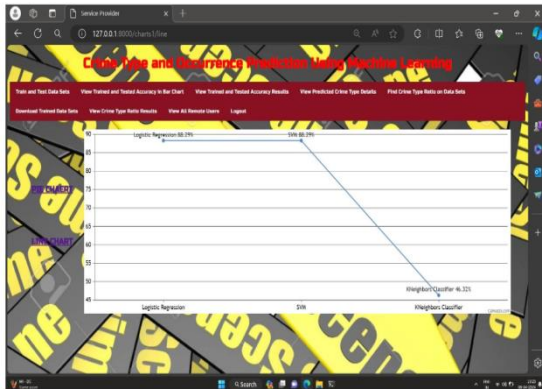




**View Trained and Tested Accuracy Results**

ID	Location	Crime Type	Date	Status
100290405	Loronek	LOMBONG GAS STATION	2018-09-03 19:00:00	0
100290403	Bandarlaha	WANGKATUN	2018-09-21 00:00:00	0
100290401	Trusmi	TERANG BAYAN WANGKATUN	2018-09-03 19:00:00	0
100290400	Investigative Property	INVESTIGATIVE PROPERTY	2018-09-03 19:00:00	0
100290399	Investigative Property	INVESTIGATIVE PROPERTY	2018-09-03 19:00:00	0

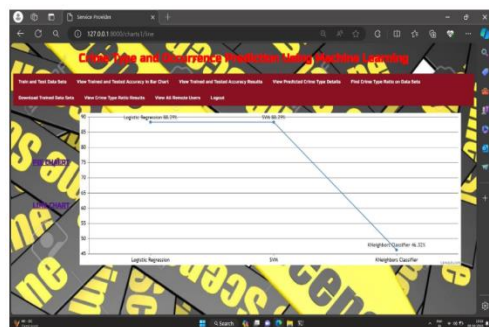
**Find Crime Type Ratio Data Sets**



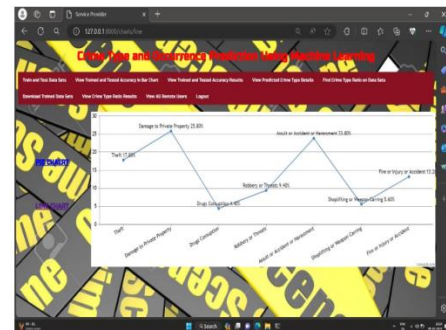
**View Predicted Crime Type Details**

Crime Type	Ratio
Damage to Private Property	21.4
Drug Consumption	21.4
Robbery or Stealing	21.4
Abuse or Assault or Harassment	21.4
Identification or Weapons Carrying	21.4
Fire or Explosion or Explosion	21.4

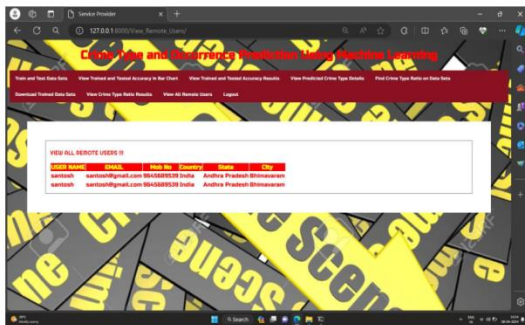
**View Crime Type Ratio Results**



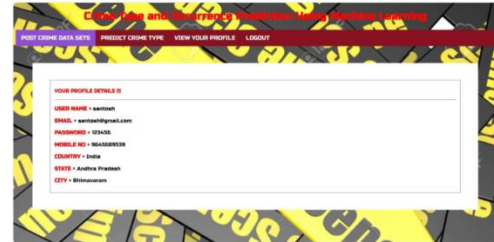
**View Predicted Crime Type Details**



**View All Remote Users**



### View your profile



### Remote User

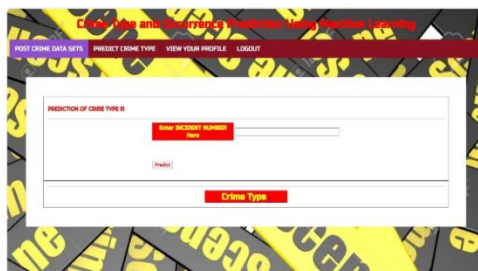
### Register



### Post crime data sets

INCIDENT NUMBER	NUMBER OFFENSE CODE	OFFENSE CODE	OFFENSE DESCRIPTION	OFFENSE LOCATION	AREA/CATEGORY	DATE	YEAR	MONTH	DAY
IN0070915	819	Larceny	LARCENY ALL OTHERS	01A	808	2018-09-02	2018	9	Sunday
IN0070913	1622	Vandalism	VANDALISM	CT1	347	2018-02-21	2018	2	Tuesday
IN0070914	3410	Towed	TOWED MOTOR VEHICLE	04A	101	2018-09-03	2018	9	Monday
IN0070910	376	Investigate Property	INVESTIGATE PROPERTY	04A	272	2018-09-03	2018	9	Monday
IN0070918	376	Investigate Property	INVESTIGATE PROPERTY	03	421	2018-09-03	2018	9	Monday
IN0070916	3820	Motor Vehicle Accident/Impairment	MV ACCIDENT/IMPAIRMENT	01	388	2018-09-03	2018	9	Monday
IN0070916	724	Auto Theft	AUTO THEFT	02	330	2018-09-03	2018	9	Monday
IN0070912	3301	Verbal Disputes	VERBAL DISPUTE	02	584	2018-09-03	2018	9	Monday

### Predict Crime Type



## 5.CONCLUSION

Utilizing two classifiers, in particular Multi ostensible NB and Gaussian NB, this examination effectively handles ostensible dissemination and genuine esteemed attributes. It is great for making expectations progressively and needn't both with a ton of preparing time. It additionally tackles the issue of managing ceaseless objective arrangements of factors, which past work neglected to oblige. Therefore, Naïve Bayesian Classification might be used to forecast and identify the most common crimes. Using certain conventional measurements, we can also calculate the algorithm's performance. When evaluating algorithms, the most important parameters to consider are accuracy, F1 score, recall, and average precision. By using machine learning methods, the

accuracy value might be significantly enhanced.

## 6. REFERENCE

- [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 Pimpuria, "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.