iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** XI **Month of publication:** November 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.47574

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue XI Nov 2022- Available at www.ijraset.com*

# Machine Learning Based Crime Rate Analysis Using Python

B. Thanuja[1], Konda Pranitha[2], Challagundla Vamshi[3], Pabbu Naveen[4], Puntikura Srinivas[5]
[1, 2, 3, 4, 5]*Dept.of ECE, CMR Technical Campus*

***Abstract:*** ***Crime is obvious that the rate of crimes were increasing day by day in all societies in world, but we personally do believe that there are a lot which can be done by both the governments and the individuals to reduce the crimes in communities. Crime analysis is a well-organized way of detecting and examining patterns and trends in crime. We should give utmost importance to study the reasons behind the crimes, so that we can prevent various crimes occurring and we can be able to find suitable solutions to prevent them. When people cannot find work, they have all the free time in the world. They think of crimes as a shortcut to obtaining and processing the riches of life, without any hardwork. To my mind, the overwhelming majority of people tend to participate in activities assisting the government to keep the society a safe place for their own families and the others and for all age groups. Our main aim of this project is to distinguish various crimes using clustering techniques based on the occurrences and regularity. In this project, the crime data is classified using the Support Vector Machine, Decision Tree, Random Forest Algorithm. This proposed system can indicate the areas which has more probability of occurring crimes so that we can easily identify the crimes based on the previous history and we can take measures to prevent the occurrences of crimes.***
***Keywords: Clustering Techniques, Criminal Intent, Data Mining***

## I. INTRODUCTION

Crime is increasing considerably day by day. Crime is among the main issues which is growing continuously in intensity and complexity. Crime patterns are changing constantly because of which it is difficult to explain behaviors in crime patterns. Crime is classified into various types like kidnapping, theft murder, rape etc. The law enforcement agencies collects the crime data information with the help of information technologies. But occurrence of any crime is naturally unpredictable and from previous searches it was found that various factors like poverty, employment affects the crime rate . It is neither uniform nor random. With rapid increase in crime number, analysis of crime is also required. Crime analysis basically consists of procedures and methods that aims at reducing crime risk. It is a practical approach to identify and analyze crime patterns. But, major challenge for law enforcement agencies is to analyze escalating number of crime data efficiently and accurately. So it becomes a difficult challenge for crime analysts to analyze such voluminous crime data without any computational support.The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster.

## II. LITERATURE SURVEY

Many researches have been done which address this problem of reducing crime and many crime-predictions algorithms has been proposed. Below are some researches:

*1)* Marchant, R.,Haan, S.Clancey, G.Cripps :Applying machine learning to criminology-semi parametric spatial demographic Bayesian regressions. A comparison of classification algorithms, Naive Bayes and decision tree was performed with an data mining software, WEKA. The datasets for this study was obtained from USCensus. The pattern of road accidents were studied after taking into consideration various factors like the driver, car, road conditions etc.Different classification algorithms used were K-Nearest Neighbour, Decision tree and Naive Bayes on a dataset containing around 18000 datapoints. The prediction accuracy for all three methods was between 79% to 81%.

*2)* Varshitha D N Vidyashree K P,Aishwarya P Janya T S,K R Dhananjay Gupta Sahana R,"Paper on Different Approaches for Crime Prediction system",International Journal of Engineering Research Technology (IJERT), ISSN: 2278-0181, 2017. Different approach for predicting like Data mining technique, Deep learning technique, Crime cast technique, Sentimental analysis technique were discussed and it was found that every method have some cons and pros. Every method gives better result for a particular instance. Clustering approaches were used for detection of crime and classification method were used for the prediction of crime.

©IJRASET: All Rights are Reserved 

*3)* Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav,"Crime pattern detection,analysis and prediction,International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2017. A thorough study of various crime prediction method like Support Vector Machine(SVM), Artificial neural networks(ANN) was done and concluded that there does not exist particular method which can solve different crime datasets problems.

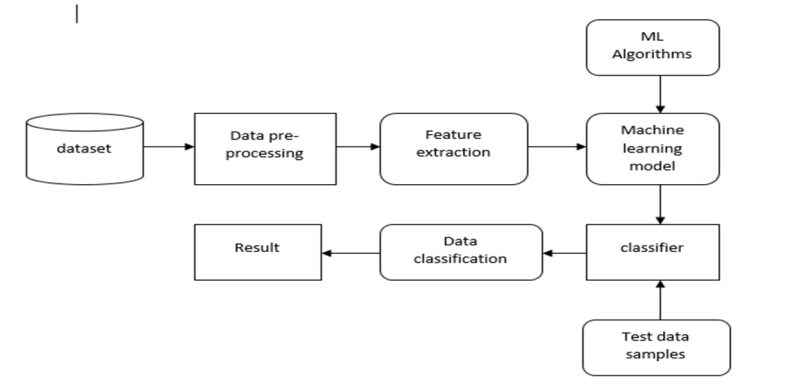
## III. PROPOSED SYSTEM

The proposed system is based on the machine learning algorithms that are available to users that can be implemented on datasets. However, there are two major types of learning algorithms: supervised learning and unsupervised learning algorithms. Here we use different machine learning algorithms such as Linear Regression, Random Forest , Support Vector Machine and clustering techniques. The dataset contains different types of crimes that are being committed in India according to the state and year respectively. The proposed model mainly has four phases: Data Collection, Data Preprocessing, Feature Extraction, Evaluation Model.

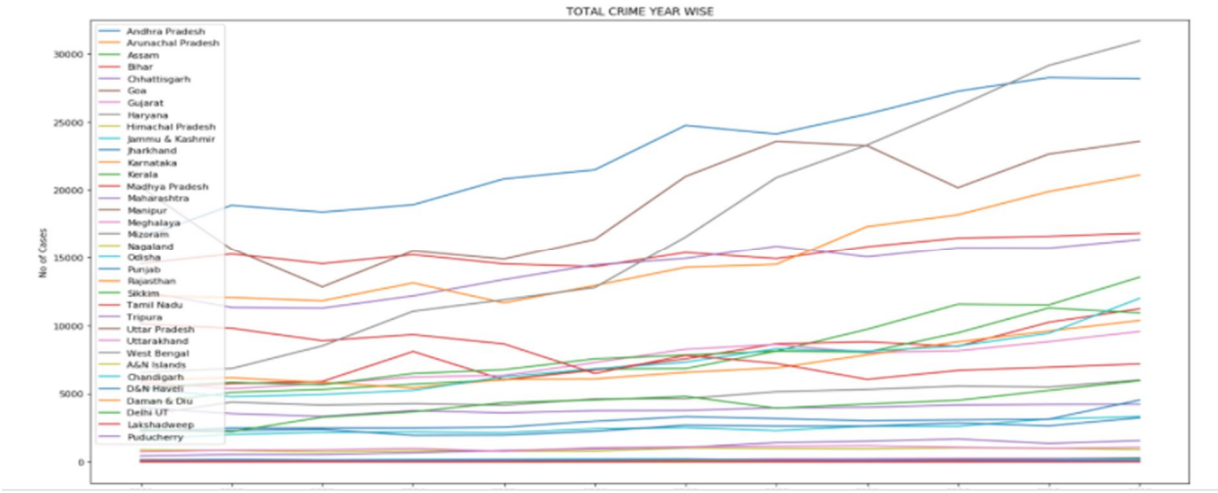
### *A. Steps Involved*

*1)* First, we take crime dataset.
*2)* Filter dataset according to requirements and create a new dataset which has attribute according to analysis to be done
*3)* Perform preprocessing and splitting on resultant dataset formed
*4)* From result plot data between crimes
*5)* Analysis can be done using Supervised classification algorithm on resultant dataset formed
*6)* Finally you will get results as accuracy metrics and the result with highest accuracy is considered as the required output.

System Architecture




## IV. RESULT



```
from sklearn.ensemble import RandomForestClassifier

random_forest = RandomForestClassifier(n_estimators=100)

random_forest.fit(X_train,y_train)

y_pred = random_forest.predict(X_test)
print(y_pred)

from sklearn.metrics import accuracy_score,classification_report
acc1=accuracy_score(y_pred,y_test)
print(acc1)

clf=classification_report(y_pred,y_test)
print(clf)
```

```
C:\Users\software\AppData\Local\Temp/ipykernel_15100/1328918803.py:7: DataCo
a 1d array was expected. Please change the shape of y to (n_samples,), for e
  random_forest.fit(X_train,y_train)
```

```
[8 0 0 0 0 0 0 0 7 0 0 0 7 0 2 0 0 0 0 8 7 0 0 0 0 0 5 0 0 0 0 0 0 0 7 0 0
 0 8 0 0 0 0 0 0 0 0 0 0 0 0 0 7 7 0 7 0 0 5 0 0 0 0 0 0]
0.9846153846153847
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        52
           2       0.00      0.00      0.00         1
           4       0.00      0.00      0.00         0
           5       1.00      1.00      1.00         2
           7       1.00      1.00      1.00         7
           8       1.00      1.00      1.00         3

    accuracy                           0.98        65
   macro avg       0.67      0.67      0.67        65
weighted avg       0.98      0.98      0.98        65
```

Fig.Accuracy result with Random Forest

```
from sklearn import svm

sv=svm.LinearSVC()

sv.fit(X_train,y_train)

predic3=sv.predict(X_test)

acc3=accuracy_score(predic3,y_test)
print(acc3)

clf3=classification_report(predic3,y_test)
print(clf3)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\valid
when a 1d array was expected. Please change the shape of y to
  return f(*args, **kwargs)
```

```
0.676923076923077
              precision    recall  f1-score   support

           0       0.85      0.90      0.87        49
           1       0.00      0.00      0.00        13
           2       0.00      0.00      0.00         1
           4       0.00      0.00      0.00         0
           5       0.00      0.00      0.00         0
           7       0.00      0.00      0.00         1
           8       0.00      0.00      0.00         1

    accuracy                           0.68        65
   macro avg       0.12      0.13      0.12        65
weighted avg       0.64      0.68      0.66        65
```

Fig.Accuracy result with SVM

```
from sklearn.linear_model import LogisticRegression

lg=LogisticRegression()

lg.fit(X_train,y_train)

predic2=lg.predict(X_test)

acc2=accuracy_score(predic2,y_test)
print(acc2)

clf2=classification_report(predic2,y_test)
print(clf2)
```

```
0.16923076923076924
              precision    recall  f1-score   support

           0       0.19      0.91      0.32        11
           1       0.00      0.00      0.00         1
           2       0.00      0.00      0.00         7
           3       0.00      0.00      0.00         9
           4       0.00      0.00      0.00         0
           5       0.00      0.00      0.00         8
           6       0.00      0.00      0.00         7
           7       0.14      0.08      0.11        12
           8       0.00      0.00      0.00        10

    accuracy                           0.17        65
   macro avg       0.04      0.11      0.05        65
weighted avg       0.06      0.17      0.07        65
```

Fig.Accuracy with Linear Regression

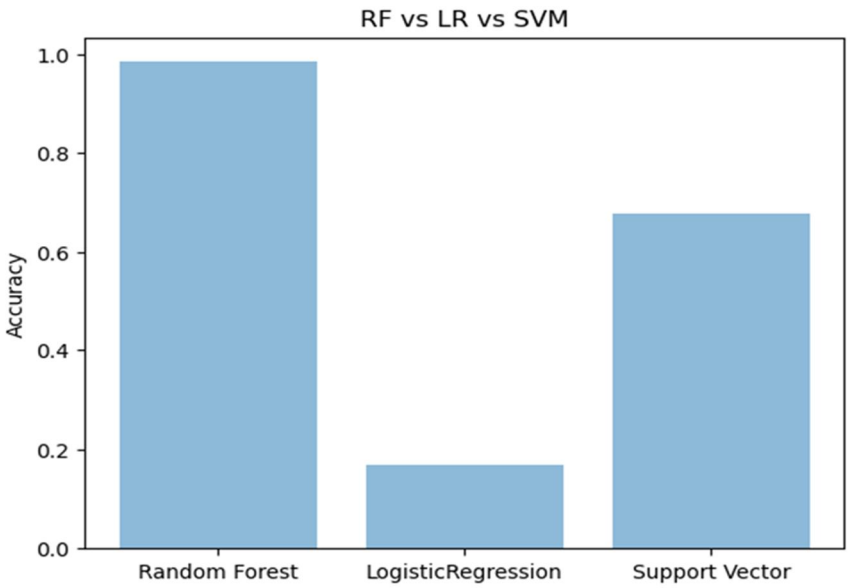



Fig.Comparative Model Accuracy Graph

## V. CONCLUSION

Crime prediction is one the current trends in the society. Crime prediction intends to reduce crime occurrences. It does this by predicting which type of crime may occur in future. Here, analysis of crime and prediction are performed with the help of various approaches. From the results obtained we saw that the training time of SVM is very high thus it should be avoided for this dataset. However which model will work best is totally dependent on the dataset that is being used.

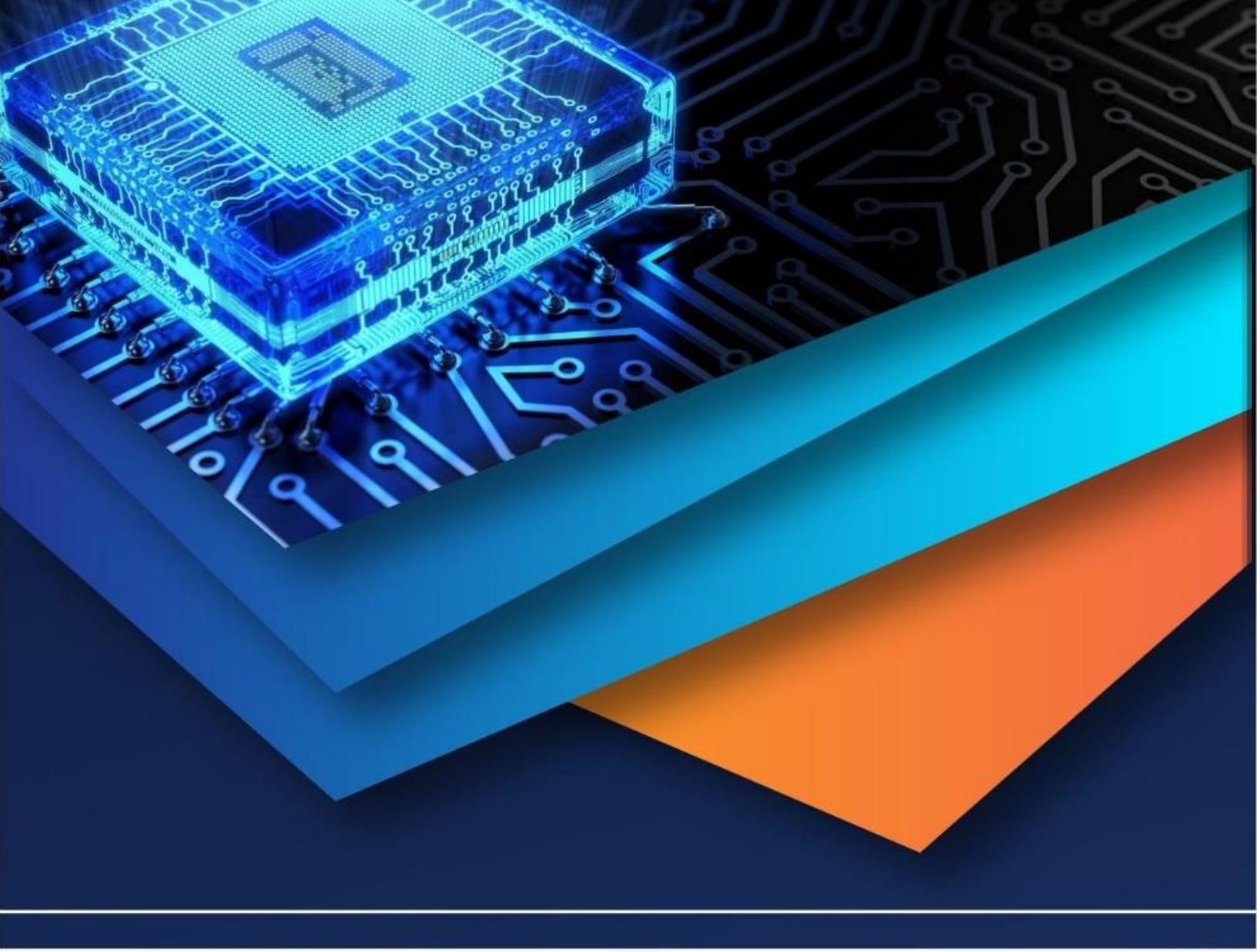
In this system, we get to classify and cluster to improve the accuracy of location and pattern-based crimes. This software predicts frequently occurring crimes , especially for particular state, and occurrences.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

## REFERENCES

[1] Marchant, R.,Haan, S.Clancey, G.Cripps :Applying machine learning to criminology-semi parametric spatial demographic Bayesian regression, 2018.
[2] Alkesh Bharati, Dr Sarvanaguru K "Crime Prediction and Analysis Using Machine Learning" in International Research Journal of Engineering and Technology ,Volume:05, 2018.
[3] Varshitha D N Vidyashree K P,Aishwarya P Janya T S,K R Dhananjay Gupta Sahana R,"Paper on Different Approaches for Crime Prediction system",International Journal of Engineering Research Technology (IJERT), ISSN: 2278-0181, 2017.
[4] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav,"Crime pattern detection,analysis and prediction,International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2017.
[5] Amanpreet Singh,Narina Thakur,Aakanksha Sharma,"A review of supervised machine learning algorithms",3rd International Conference on Computing for Sustainable Global Development,2016.
[6] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, Alex Pentland."Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data",in ACM International Conference on Multimodal Interaction (ICMI 2014).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)