

Crime Type and Occurrence Prediction Using Machine Learning

Mrs. B. Vijaya, MCA, M.Tech ^[1], B. Nagasaikiran ^[2]

^[1] Asst. Professor, Department of Computer Application

^[2] Student, Department of Computer Application

^{[1],[2]} Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRACT

In this era of recent times, crime has become an evident way of making people and society under trouble. An increasing crime factor leads to an imbalance in the constituency of a country. In order to analyse and have a response ahead this type of criminal activities, it is necessary to understand the crime patterns. This study imposes one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of most recently occurring crimes. The major aspect of this project is to estimate which type of crime contributes the most along with time period and location where it has happened. Some machine learning algorithms such as Naïve Bayes is implied in this work in order to classify among various crime patterns and the accuracy achieved was comparatively high when compared to pre-composed works.

Keywords: - Machine Learning, Kaggle, Crime Type, Crime patterns

I. INTRODUCTION

Crime has become a major thread imposed which is considered to grow relatively high in intensity. An action stated is said to be a crime, when it violates the rule, against the government laws and it is highly offensive. The crime pattern analysis requires a study in the different aspects of criminology and also in indicating patterns. The Government has to spend a lot of time and work to imply technology to govern some of these criminal activities. Hence, use of machine learning techniques and its records is required to predict the crime type and patterns. It imposes the uses of existing crime data and predicts the crime type and its occurrence bases on the location and time.

Researchers undergone many studies that helps in analysing the crime patterns along with their relations in a specific location. Some of the hotspots analysed has become easier way of classifying the crime patterns. This leads to assist the officials to resolve them faster. This approach uses a dataset obtained from Kaggle open source based on various factors along with the time and space where it occurs over a certain period of time. We implied a classification algorithm that helps in locating the type of crime and hotspots of the criminal actions that takes place on the certain time and day. In this proposed one to impose a machine learning algorithms to find the matching criminal patterns along with the assist of its category with the given temporal and spatial data.

II. RELATEDWORKS

Suhong Kim and Parminder Singh Kalsi proposed Crime Analysis Through Machine Learning. This Project investigates machine-learning-based crime prediction. In this work, Vancouver crime data for the last 15 years is analyzed using two different data-processing approaches. Machine-Learning predictive models, K-nearest-neighbour and boosted decision tree, are implemented and a crime prediction accuracy between 39% to 44% is obtained when predicting crime in Vancouver.

Shanjana A.S, Dr.R.Porkodi proposed crime analysis and prediction using datamining: a review. Crime analysis and prediction is a systematic approach for identifying the crime. This system can predict region which have high probability for crime occurrences and visualize crime prone area. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. The extraction of new information is predicted using the existing datasets. Crimes are treacherous and common social problem faced worldwide. Crimes affect the quality of life, economic growth and reputation of nation. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. propose a system which can analysis, detect, and predict various crime probability in given region. This project explains various types of criminal analysis and crime prediction using several data mining techniques.

Shruti S.Gosavi, Tilak Maharashtra Vidyapeeth and Shraddha S. Kavathekar proposed A Survey on Crime Occurrence Detection and prediction Techniques. Due to the increasing rate of crimes there is a need of system that will detect and predict crimes at dynamic time. The aim of this survey is to study Data Mining techniques that will help to detect and predict crimes using association rule mining, k-means clustering, decision trees & naive bayes and Machine learning techniques such as deep neural network and artificial neural network. Findings of this survey were that when the dataset instances have more number of missing values pre-processing becomes a vital task and crime does not occur uniformly across urban landscapes but concentrates in certain areas. So, predicting crime hotspots is a very important task also applying post-processing will help in lowering the rate of crimes. In pre-work, the dataset obtained from the open source are first pre-processed to remove the duplicated values and features. Decision tree has been used in the factor of finding crime patterns and also extracting the features from large amount of data is inclusive. It provides a primary structure for further classification process. The classified crime patterns are feature extracted using Deep Neural network. Based on the prediction, the performance is calculated for both trained and test values. The crime prediction helps in forecasting the future happening of any type of criminal activities and help the officials to resolve them at the earliest. The pre-existing works account for low accuracy since the classifier uses a categorical values which produces a biased outcome for the nominal attributes with greater value. The classification techniques does not suited for regions with inappropriate data and real valued attributes. The value of the classifier must be tuned and hence there is a need of assigning an optimal value.

III. PROPOSED SYSTEM ARCHITECTURE

The data obtained is first pre-processed using machine learning technique filter and wrapper in order to remove irrelevant and repeated data values. It also reduces the dimensionality thus the

data has been cleaned. The data is then further undergoes a splitting process. It is classified into test and trained data set. The model is trained by dataset both training and testing .It is then followed by mapping. The crime type, year, month, time, date, place are mapped to an integer for ensuring classification easier. The independent effect between the attributes are analysed initially by using Naïve Bayes. Bernouille Naïve Bayes is used for classifying the independent features extracted. The crime features are labelled that allows to analyse the occurrence of crime at a particular time and location. Finally, the crime which occur the most along with spatial and temporal information is gained. The performance of the prediction model is find out by calculating accuracy rate. The language used in designing the prediction model is python and run on the Colab – an online compiler for data analysis and machine learning models. The proposed algorithm is well suited for the crime pattern detection since most of the featured attributes depends on the time and location. It also overcomes the problem of analyzing independent effect of the attributes. The initialization of optimal value is not required since it accounts for real valued, nominal value and also concern the region with insufficient information. The accuracy has been relatively high when compared to other machine learning prediction model. In this proposed system there are two modules they are:

1. Service Provider
2. User

Service Provider module functionalities are Login, Train & Test Data Sets., View Trained & Tested Accuracy in Bar chart, View Trained &Tested Accuracy results, View Predicted crime type details, Find crime type ratio on datasets, Download trained datasets, View crime type ratio results, View all Remote users, Logout
User module functionalities are Register and login, Post crime datasets, Predict crime type, View your profile, Logout

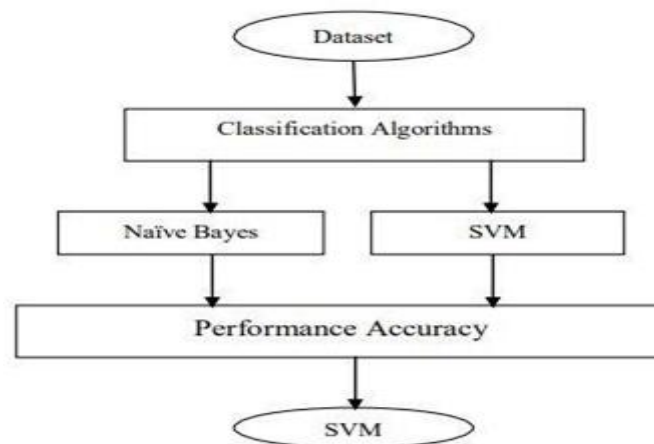


Fig.1 Proposed system

IV. RESULTS AND DISCUSSION

The output screens obtained after running and executing the system are shown from Fig.2 to Fig.7

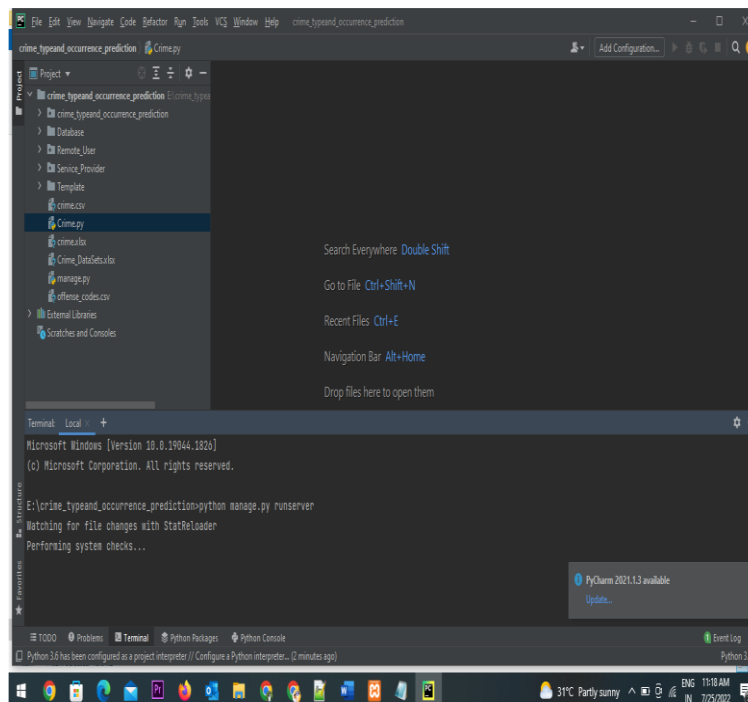


Fig 2. Running Program in Pycharm

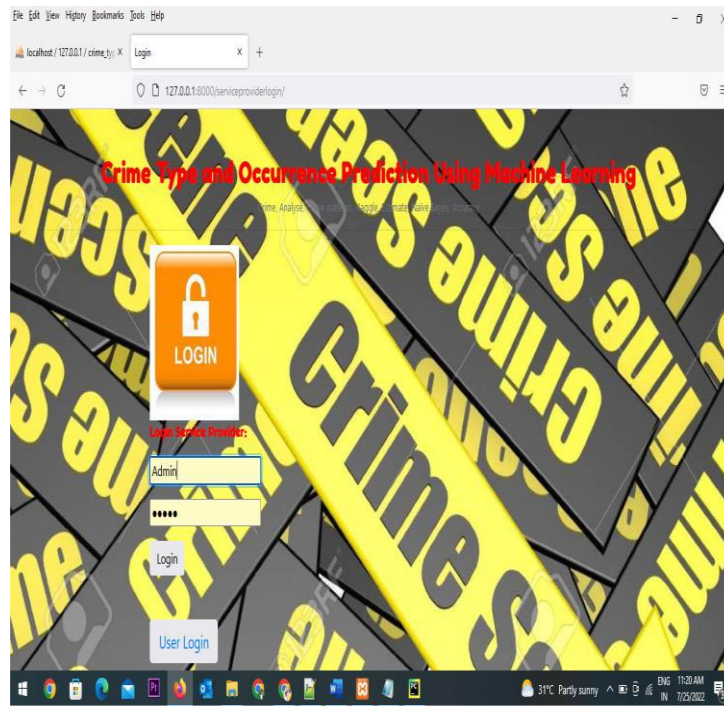


Fig 3.Admin Login

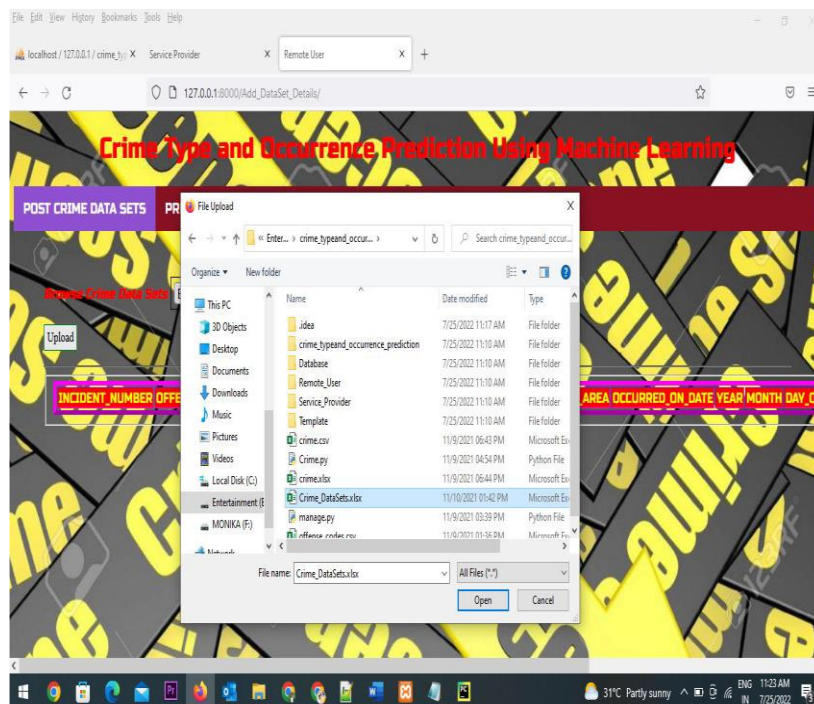


Fig.4 Browse Data set

INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	OCCURRED_ON_DATE	YEAR	MONTH
I182070945	619	Larceny	LARCENY ALL OTHERS	D14	808	2018-09-02 13:00:00	2018	9
I182070943	1402	Vandalism	VANDALISM	CT1	347	2018-08-21 00:00:00	2018	8
I182070941	3410	Towed	TOWED MOTOR VEHICLE	D4	151	2018-09-03 19:27:00	2018	9
I182070940	3114	Investigate Property	INVESTIGATE PROPERTY	D4	272	2018-09-03 21:16:00	2018	9
I182070938	3114	Investigate Property	INVESTIGATE PROPERTY	B3	421	2018-09-03 21:05:00	2018	9
I182070936	3820	Motor Vehicle Accident Response	M/V ACCIDENT INVOLVING PEDESTRIAN - INJURY	CT1	398	2018-09-03 21:09:00	2018	9
I182070933	724	Auto Theft	AUTO THEFT	B2	330	2018-09-03 21:25:00	2018	9
I182070932	3301	Verbal Disputes	VERBAL DISPUTE	B2	584	2018-09-03 20:39:00	2018	9
I182070931	301	Robbery	ROBBERY - STREET	C6	177	2018-09-03 20:48:00	2018	9
I182070929	3301	Verbal Disputes	VERBAL DISPUTE	CT1	364	2018-09-03 20:38:00	2018	9
I182070928	3301	Verbal Disputes	VERBAL DISPUTE	C6	913	2018-09-03 19:55:00	2018	9

Fig.5 data set Details

```

l-defined and being set to 0.0 in labels with no predicted samples. Use 'zero_division' parameter to control this behavior.
_warn_pnf(average, modifier, msg_start, len(result))
C:\Users\DELL\AppData\Local\Programs\Python\Python38\lib\site-packages\sklearn\metrics\classification.py:1248: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use 'zero_division' parameter to control this behavior.
_warn_pnf(average, modifier, msg_start, len(result))
precision  recall  f1-score  support

0         0.88    1.00    0.94    70566
1         0.00    0.00    0.00    9203

accuracy          0.88    79769
macro avg    0.44    0.50    0.47    79769
weighted avg 0.78    0.88    0.83    79769

CONFUSION MATRIX
[[ 70566  0]
 [  9203  0]]
SVM
    
```

Fig. 6 Training with ML Algorithms

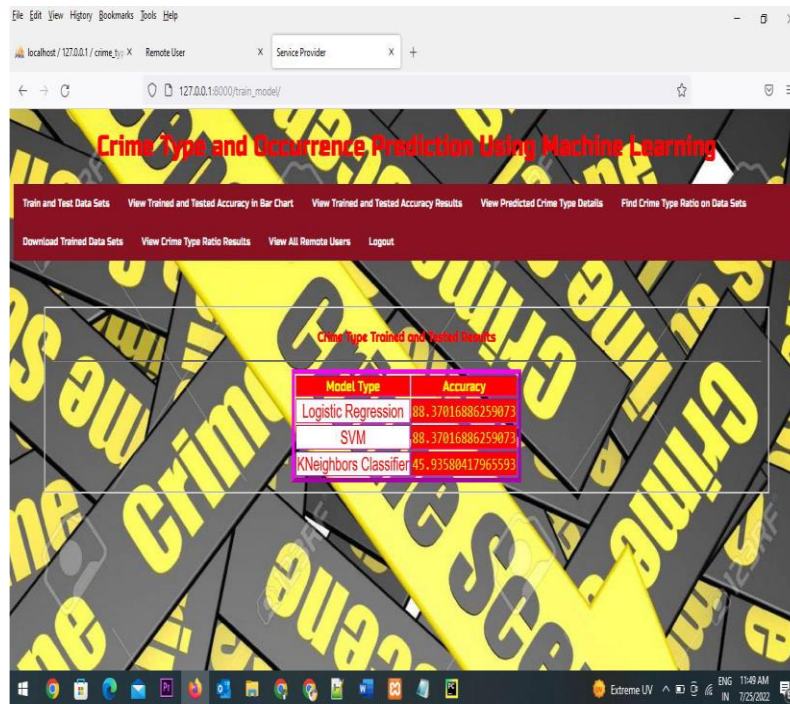


Fig.7 Generating Trained and Tested accuracy

V. FUTURE SCOPE AND CONCLUSION

In this project, the difficulty in dealing with the nominal distribution and real valued attributes is overcome by using two classifiers such as Multinomial NB and Gaussian NB. Much training time is not required and serves to be the best suited for real-time predictions. It also overcomes the problem of working with continuous target set of variables where the existing work refused to fit with. Thus the crime that occur the most could be predicted and spotted using Naïve Bayesian Classification. The performance of the algorithm is also calculated by using some standard metrics. In the future work, crime that occur the most could be predicted and spotted using Convolutional Neural network Classification.

REFERENCES

- [1] 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.
- [2] Chandy, Abraham, "Smart resource usage prediction using cloud computing for massive data processing systems" Journal of Information Technology 1, no. 02 (2019):108-118.
- [3] 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
- [4] 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.
- [5] 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).
- [6] Nikhil Dubey and Setu K. Chaturvedi, "A Survey Paper on Crime Prediction Technique Using Data Mining", Corpus ID: 7997627, Published on 2014.