TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 16235 - 16240

Crime Forecasting using Folium

Akshay Jawla, Chandigarh University, Mohali, Punjab, India
Manjot Singh, Student,Chandigarh University, Mohali, Punjab, India
Nishtha Hooda, Assistant Professor,Chandigarh University, Mohali, Punjab, India

Article Info
Volume 82
Page Number: 16235 – 16240
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 28 February 2020*

Abstract:
This research is related to machine learning, Aswe know that crime is an alarming aspect of the modern world. This research plays a vital in studying different factors and relationships among crimes and to discover the best possible approach to control crime. crime forecasting using folium is an approach for analyzing and predicting the pattern of crime in a specific area. This research can predict the area which has higher probabilities for crime occurrence and by using folium we can visualize the crime-prone areas, with the increasing emergence of the computerized system, this research can help to predict the crime. It will help in predicting the information about crime-prone specific areas in advance. In this, the concept of data mining is implementing, as extracting the previously known information. This research is using the concept of the K-means clustering algorithm to make a cluster of similar types of information. Once the clustering is done, we are able to predict a crime based on its historical data. With the help of this research, we can indicate regions which have higher crime rates

Keywords: *machine learning; crime forecasting; folium*

I. INTRODUCTION

1.1 In today's world, our society is suffering from various crimes. Crimes have become one of the serious issue of our society. Our government and Police have applied various methods to control the crime, but the result is not in their favor. Controlling crime doesn't mean that crime will end but it means the rate of crime will be reduced or we can prevent the occurrence of crime. Crimes are a vital threat to our society. There are several numbers of crimes that happen in our society in a regular period of time. Reasonably it is spreading at an enormous rate and this affects the whole of mankind. Crimes occur from big cities, towns to small towns. There are different types of crimes -kidnapping, rape, murder, robbery, assault, etc. Since the rate of crimes is increasing there is a need to provide a solution that helps in reducing the cases in a much faster way. if we are talking about some popular cities of India like Delhi, Chandigarh and many others. we notice that the rate of crime in these popular cities is very high. In Delhi, because of the high rate of crime i.e rape. Delhi is known as rape capital. as this research can work as a helping hand in preventing the crime because with the help of advanced technology the prediction of crime can be done, and prevention is always better than cure. For predicting the crime, we require the data. using this data, the crime prediction and identification of criminals can be done. but there is an immense problem of data size, so we require an efficient technology that provides a better way to overcome this problem.

1.2 The above-discussed problem made me think that how we can solve this issue and what are the various efficient approaches we can apply to use modern technology and research and provide a better solution. we have read some related work and do some research on related previous works we can say that technologies such as data science and machine learning can provide a solution for what we are looking for. with using these technologies, we can provide a better solution for this problem faster.

1.3 The main objective of this work and project is to make a prediction of crime using all the

January-February 2020
ISSN: 0193-4120 Page No. 16235 - 16240

necessary and individual features present in the dataset. This dataset is derived from government sites. By using features of the dataset and using the machine learning algorithm we can predict the crime in a specific area.

1.4 The goal of this research is to develop a model for efficient prediction of crime. we are using the dataset for training which is downloaded from the Delhi government sites. This dataset has real crime data of various districts of Delhi by using this dataset we can train our model. All the approaches and methods used in this research is being already used by previous researchers. so the accuracy of this work is assured. The algorithm we used in this work is the K-means Clustering algorithm for crime prediction.The k-means algorithm is an efficient algorithm for predicting the crime because in previous researcher work we found that K-means is better than algorithms such as KNN. Visualization is imposed to increase the flexibility of research as visualization on a map is relatively easy to understand. it will use by technical person as well as normal persons. The concerned authority can use this work and plan a strategy in advance to control the crime. It will help Delhi Police Department to take action against the criminals to reduce the crime. This will allow us to determine in advance which particular area is safe for women and children and where we can buy assets for us and as it will forecast to people where there are more property-related concerns and that place is secure from thieves. Later we can use powerful python library folium to map our obtained result with different colors to depict which place is more dangerous which place is safe for us to visit or buy the property or safe for children and women. Our work is about the crime prediction of Delhi. Here we are mapping for different crimes and enhance our decision-making in which crime is prominent in which area of the city. we are providing all the help we can provide to reduce the crime in Delhi as the crime is increased which means the plan of police to reduce crime is not serving as this problem required a modern solution.

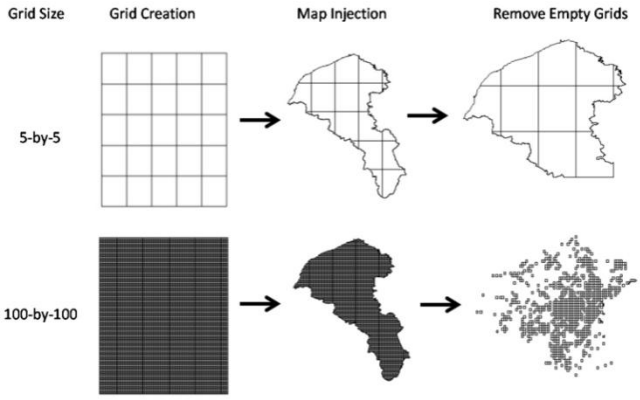


Fig.1

II. STATE-OF-THE-ART

This research investigation used algorithm machine-learning for crime prediction. the data of city Vancouver is analyzed and is uses two dataset processing techniques which is KNN and predictive models, with the accuracy of 39% to 44% are implemented[1].

This work is done on Real-time forecasting of crime. However, it is difficult to predict when and where crime will happen. No physical model contributes a sensible resemblance to such a compact system. crime data are inadequate in both time and space and there is a weak signal of interests. In this work, proper representation of crime data is presented and then we accommodate the spatial-temporal residual network on the represented data for the prediction of crime in LA at the scale of hours in neighborhood-sized parcels. these comparisons and experiments with other existing approaches for prediction show the sovereignty of the model in terms of accuracy [2].

The purpose of an adequate analysis of crime is profitable and valuable to guess the performance of the inclination and design of crimes. this work assists powers in implementing early crime

January-February 2020
ISSN: 0193-4120 Page No. 16235 - 16240

restriction measures. various crime forecasting models has been developed using this technique. It has been noted that in the forecasting of crime the researchers have started to turn their analysis credits from the statistical model to the AI model. This work is done to observe the abilities of artificial intelligence in enhancing crime forecasting. the main goal of this work is to carry a relative analysis of crime forecasting capabilities of four different artificial intelligence systems. i.e artificial neural network(ANN), random forest(RF),support vector regression(SVR),gradient tree boosting(GTB). the measurement of errors is used to evaluate the forecasting ability[3].

This work shows a strategy of deep learning for forecasting of crime related to urban areas. A structure of deep neural network is created so that by practicing geo-referenced data of criminal is trained to catch relative spatial models. The obtained result of this model advise that the model would be able to identify zones with criminal liveliness in sq. blocks of 500×500500×500 m22 on a weekly range [4].

III. MATERIAL AND METHOD.

To build a model that is capable of making predictions we use foretelling modeling predictive modeling. to make those predictions this method includes a machine learning algorithm. this algorithm learns certain characteristics from the training dataset. The predictive is further divided into two different areas.i.e regression and pattern classification. In regression models the analysis of the relationship between variables and trends to make predictions about continuous variables. In distinction to regression models, the work of pattern classification is to join critical data value as the

production of a prediction to class labels. The pattern of classification tasks can be further divided into two parts, unsupervised learning and supervised learning, the class descriptions in the dataset, which is applied in the creation of classification models. In a supervised learning issue, we would distinguish which training dataset is used to train and which dataset produces output so that the prediction can be made for unseen data.

Predictive Models

Decision tree: In this approach, a tree-shaped pattern of decisions including chances of event results, costs, and utility is used In machine learning, it is a way to display an algorithm, segment of the simplistic probabilistic classifier i.e naive Bayes classifier as it based on executing Bayes theorem making assumptions between the features individually. In this technique classifier model is developed which provides class labels to concerns, represented as a vector of values, we can draw these class labels using finite sets.

Linear regression: The study for evaluating the relationships between the variables. it is a proposal for modeling the relationship between explanatory variables X and scalar reliant variable Y. if it is the case of one illustrative variable then it is called simple linear regression otherwise it is called multivariate.

Logistic regression: In logistic regression design the value of the subordinate variable is binary or certain. to eliminate any null values or absolute values the system includes some methods. these absolute values may influence the correctness of the arrangement.

Proposed Work

We believe that finding the relationship between crime elements could help in forecasting or predicting the chances of occurring of crime in a specific area this work could help us in predicting the dangerous spot at a certain time in the future. This research aimed to focus on three elements of crime data we get of any specific area as in this research we are using data of Delhi. So, crime data of Delhi is being analyzed in this work. As we focused on three main elements of the crime data i.e type of crime, the occurrence of crime and location of the crime. In this work, we extract all interesting

January-February 2020
ISSN: 0193-4120 Page No. 16235 - 16240

patterns based on criminal elements and variables. After that, we applied some classification method and with the help of these classification methods, we create some classes of the group of required data and on this required data we can apply our clustering algorithm which provides us results which are sufficient to predict crime types in a specific location within a specified range of time.

Functional Diagram of proposed work.

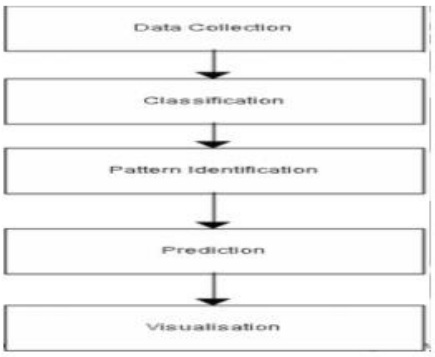
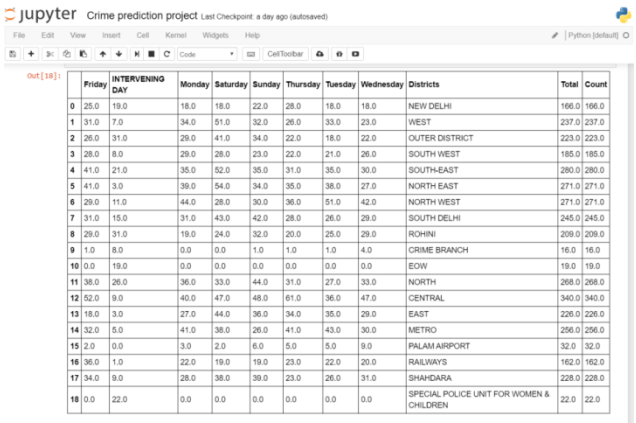


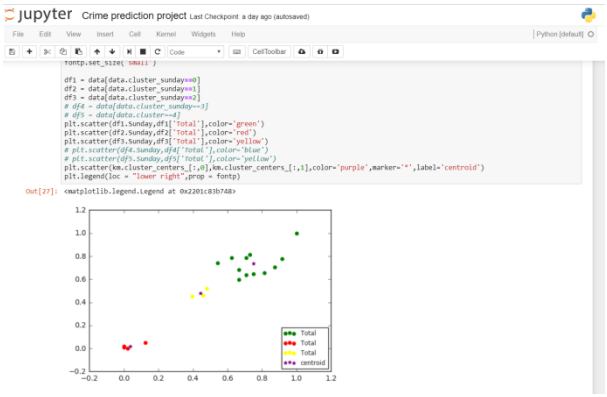
Figure 1- Architecture

Data pre-processing is a data mining technique that is used for the conversion of raw data into anmeaningfulformat.Data is often incompatible and lacking in trends and contains many errors and faults. that's why we need to clean data and make suitable for our clustering algorithm.

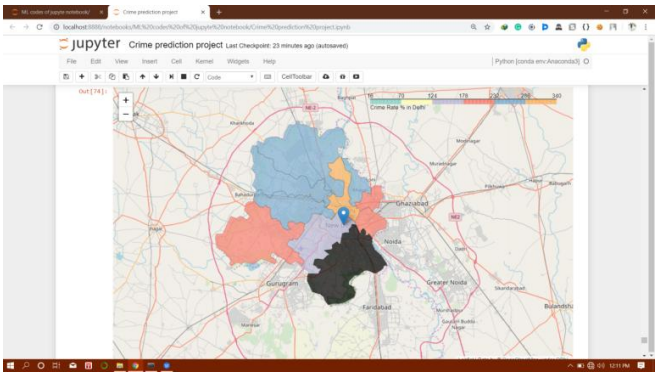


	Friday	INTERVENING DAY	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday	Districts	Total	Count
0	25.0	19.0	18.0	18.0	22.0	28.0	18.0	18.0	NEW DELHI	166.0	166.0
1	31.0	7.0	34.0	51.0	32.0	26.0	33.0	23.0	WEST	237.0	237.0
2	26.0	31.0	29.0	41.0	34.0	22.0	18.0	22.0	OUTER DISTRICT	223.0	223.0
3	28.0	8.0	29.0	28.0	23.0	22.0	21.0	26.0	SOUTH WEST	185.0	185.0
4	41.0	21.0	35.0	52.0	35.0	31.0	35.0	30.0	SOUTH-EAST	280.0	280.0
5	41.0	3.0	39.0	54.0	34.0	35.0	38.0	27.0	NORTH EAST	271.0	271.0
6	29.0	11.0	44.0	28.0	30.0	36.0	51.0	42.0	NORTH WEST	271.0	271.0
7	31.0	15.0	31.0	43.0	42.0	28.0	26.0	29.0	SOUTH DELHI	245.0	245.0
8	29.0	31.0	19.0	24.0	32.0	20.0	25.0	29.0	ROHINI	209.0	209.0
9	1.0	8.0	0.0	0.0	1.0	1.0	1.0	4.0	CRIME BRANCH	16.0	16.0
10	0.0	19.0	0.0	0.0	0.0	0.0	0.0	0.0	EOW	19.0	19.0
11	38.0	26.0	36.0	33.0	44.0	31.0	27.0	33.0	NORTH	268.0	268.0
12	52.0	9.0	40.0	47.0	48.0	61.0	36.0	47.0	CENTRAL	340.0	340.0
13	18.0	3.0	27.0	44.0	36.0	34.0	35.0	29.0	EAST	226.0	226.0
14	32.0	5.0	41.0	38.0	26.0	41.0	43.0	30.0	METRO	256.0	256.0
15	2.0	0.0	3.0	2.0	6.0	5.0	5.0	9.0	PALAM AIRPORT	32.0	32.0
16	36.0	1.0	22.0	19.0	19.0	23.0	22.0	20.0	RAILWAYS	162.0	162.0
17	34.0	9.0	28.0	38.0	39.0	23.0	26.0	31.0	SHAHDARA	228.0	228.0
18	0.0	22.0	0.0	0.0	0.0	0.0	0.0	0.0	SPECIAL POLICE UNIT FOR WOMEN & CHILDREN	22.0	22.0

and classification of data is done to classify the data into classes to get target data from the collected data as our algorithm requires some data to make clusters that are required for prediction.



We identify the pattern for specifying the location and corresponding to this location we take attributes and these attributes predict some pattern to that cluster. With the help of historical information, the prediction is done as the information is able to classify the clusters and after the cluster forms the prediction is also possible. We have used folium to represent our obtained result on the Delhi map.



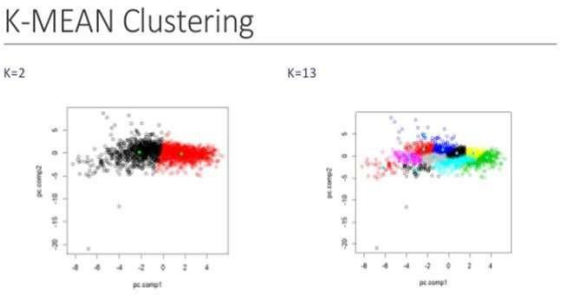
IV.RESULTSANDDISCUSSION

We have used k-means clustering algorithm in our work.it is one of the methods of cluster analysis in this partition and observation into k number of clusters in this each iteration belongs to the cluster with the closet mean.

Firstly, we choose the set of K instances as centers of the clusters. Then the algorithm considers each instance and allocates that instance to the nearest cluster. Recalculation of the cluster is done even after the whole cycle of reassignment is done for each assignment.

The analysis of K-means clustering involves tracking the crime occurs in a particular area on a specified day. Here we consider homicide crime and plot it against the districts of the Delhi and day on which the crime is reported.

January-February 2020
ISSN: 0193-4120 Page No. 16235 - 16240

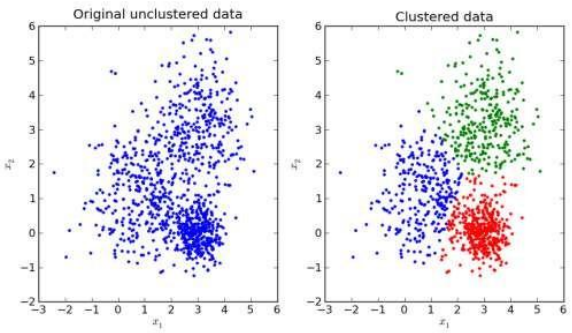


(a)Non-classified clusters (b)Classified clusters

The result of this work will be to predict and correlate the crimes from crime datasets available. Results will be in the form of a correlation between the location of crime and crime. The prediction of crime will be displayed using folium and tables for tabular representation.

As we know that the aspect which must be given higher priority by all political and government organizations is Security.

These organization aiming to lower the incidences of the crime. As data mining is the correct field to apply or implement on more massive datasets, the output which we are provided by this approach will be highly useful and support the law to reduce the crime. So, in this research, the crime analysis is done by performing the algorithm which is named K-means clustering algorithm on real-life crime datasets.



Clustered Data and Unclustered Data

Future Scope

As of now in this work, we provide the dataset for a particular area like in our work the crime dataset of Delhi is being used what if we make a centralized system by which we can check the crimes rate for the specific location and connect this system to all the police station countrywide. This will help the police force to control the crime. And also, the reporting should become digital i.e FIR reporting should become digital so that by using this reporting data we should able to predict the crimes in a particular location.

Conclusion

In this work, we apply the K-means algorithm to form clusters which improve the accuracy of location and pattern-based crimes in a particular location. And visualize the data using folium.we have not adopted a traditional technique. traditionally researchers used to clean data then start working on the dataset, but we instead identified the data elements that we can proceed accordingly. we met our goal of accurate and user-friendly visualization of our work and for visualization, we have used Folium. in the end, we met our objective as now our model can efficiently predict and visualize crime. we can provide support to the Delhi police and this will help in stabilizing the condition of crime in Delhi.

Acknowledgment

We owe a very sincere thanks to our university CHANDIGARH UNIVERSITY for providing us platform to prepare a project on such important topic CRIME FORECASTING USING FOLIUM and we would like to thank our project Mentor Dr. NishthaHooda for instigating within the requirements of the research and providing us the opportunities and time to conduct the research. We are grateful to our Mentor as it would have seemed difficult without her motivation, constant support and suggestion.

References

[1] Kim, S., Joshi, P., Kalsi, P. S., &Taheri, P. (2018, November). Crime Analysis Through Machine Learning. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 415-420). IEEE.

January-February 2020
ISSN: 0193-4120 Page No. 16235 - 16240

[2] Wang, Bao, et al. "Deep learning for real-time crime forecasting and its ternarization." *arXiv preprint arXiv:1711.08833* (2017).

[3] Khairuddin, AlifRidzuan, RazanaAlwee, and HabibollahHarun. "Comparative Study on Artificial Intelligence Techniques in Crime Forecasting." Applied Mechanics and Materials. Vol. 892. Trans Tech Publications Ltd, 2019.