

An Efficient Crime Forecasting Using Machine Learning and Folium

Abhay Singh¹, Dr. Ruchika Gupta², Dr. Ruchika Gupta³, Navpreet Kaur⁴, Manish Kumar⁵

^{1,2} Student, ^{3,4,5} Project Mentor

^{1,2,3,4} Department of Computer Science, Chandigarh University Mohali, Punjab

Article Info

Volume 82

Page Number: 16743 - 16749

Publication Issue:

January - February 2020

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 29 February 2020

Abstract

In our country India crime is one of the major issues which we have to tackle and overcome. Our government have taken various steps but to solve this problem we have to act smart and make use of the technology to overcome such major issue. This research is done keeping this issue in mind that we have to think global and act local. So, this research plays a vital in studying different factors and analyzing relationships among crimes and to discover the best possible approach to tackle and control crime. Smart and Efficient Crime forecasting using folium is an approach for analyzing and predicting the pattern of crime in a specific area.

I. INTRODUCTION

Crime is one major issue of concern in our society.

Various techniques are adopted to overcome this issue but no major change has been observed. Controlling crime rate can re-enforce peace and harmony in our society. We cannot completely finish or control crime but we can prevent it because it is often said prevention is better than cure.

1.1 Crimes are the significant challenge and threat to mankind. There are several crimes that occur at the regular interval of time again and again. Perhaps it is increasing and spreading at a faster and rate than probably what we think. Crimes are happening everywhere be it small town be it village or any metro city including Delhi. There are various types of crimes that are taking place such as- rape, murder, robbery, honor killing etc, Since crimes are increasing there is a need of preventing these crimes to happen we are here trying to prevent crime rather than curing it .As we have all observed that the criminal activities have increased at a faster rate and we think or believe that it is the only accountability of police department to monitor as well as control

the criminal activities . crime reduction process. These researches can act as catalyst. I think through such studies and researches we can enhance or improve in a more efficient way the crime prediction process. As the data is too much or too large in size crime prediction is the real issue faced. There is a need of smarter, more reliable and at the same time efficient technology so that police can tackle such huge problem better.

1.2 This issue made me to think how we can be smarter in our approach and come up with better solution through the use of proper and efficient use of modern technology. After proper research and reading several previous works done on the same issue, we realized technologies such as machine learning and data science can solve this problem.[1]

1.3 The main aim of this research and project is to predict crime using all the unique features present in the dataset. The dataset is obtained from the official and government sites.

1.4 The main and top priority is to train the model for as accurate crime prediction as possible. The

Training will be performed on the data set downloaded from the official government website of Delhi Government, so training is done on real data not on any imaginary data. The algorithm chosen purely on the basis of accuracy according to the that was observed by previous researchers, nothing is hit and trial everything is tried and tested. The K-Means Clustering algorithm will be used for efficient crime prediction as statically in the past it has shown better results than other algorithms such as KNN . Visualization is the key and is done so the even non technical person for example – various law enforcement agency in this case so that they can understand and predict on which day which part of city is more susceptible to crime or is more crime prone and plan their crime control strategies accordingly and can understand in advance which part of city is safe for women, children and where we can buy property for us and different people as it will forecast to people where there are more property related issues and that place is safe from thieves and mafia's.

Later we can use folium to map with different colors to depict which place is more dangerous which place is safe for us to visit or buy property or safe for children and women. Here we are mapping for different crimes and enhance our decision-making which crime are prominent in which part of city. [2]

Here in our research we are concerned about crime forecasting of capital of our country New Delhi, because all the necessary steps required to control crime Delhi Police is falling short, it is needed to come up with a smart solution to an age-old problem of the city.

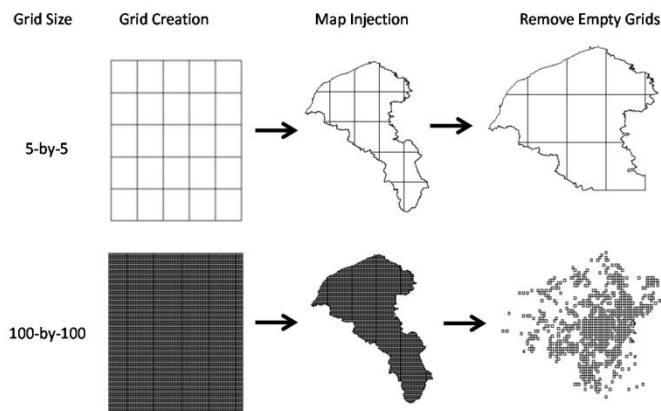


Fig.1 SAMPLE DIADRAM

II. RELATED WORK

In past several researchers have worked on this issue, so there is lot of related work done on predicting crime in rural as well as urban areas, few of them are listed below-

2.1 PREDICTING NEIGHBOURHOOD RISK OF CRIME

This study was done by Adele V Haroll, and it tried to describe the relationship between crime and urban decay as well as focused on predicting crime using decision tree.

2.2 CRIME FORECASTING USING DATA MINING TECHNIQUES

This study was done by Max W. Ward along with his three team members. This used various data science techniques to predict crime and it was done in collaboration with Police Department of United States city situated in Northeast.

2.3 AN EXPERIMENTAL STUDY OF IMPLEMENTATION OF VARIOUS CRIME PREDICTION ALGORITHMS

This study was conducted by R Iqbal along with other team members. This study was rather focused on accuracy of different algorithms rather than applying one particular algorithm so that we can identify which algorithm is more reliable and gives best results for real time data.

III. MATERIAL AND METHOD

3.Forecasting Modeling Predictive modeling is the method of structuring a model that is more capable of coming up with accurate predictions. The technique involves machine learning algorithm that learns various properties and attributes from a training dataset in order to make correct and efficient predictions.

Prediction model that we select depends on our data set. If we are choosing dataset or target data that is not random rather it is continuous then we will rather choose Regression Model, but if we are choosing target dataset as discrete then we will choose Classification Model. [3]

Types of Predictive Models

Algorithms Classification and Decision Trees-It is one of those very few or if any such model exists which offers this property of customizability. It in turn act as the basic foundation of more optimized models including- Random Forest also known as gradient boosting etc.

Linear Regression – It is quite easy to identify when to use Linear Regression. When the data varies continuously i.e. data is continuous say from all real numbers as the case of temperature, we can use linear regression.

There are certain methods that we will follow such as-

Data Preprocessing- This procedure includes methods to remove or eliminate any irrelevant values which may not require or is of no use to us or null values or infinite values which may affect the accuracy and efficiency of the system. The main steps include

- Formatting,
- cleaning,
- sampling.

Cleaning procedure is used for elimination or fixing of some missing values in data such as few values in a row or column is missing or there may be a case that data is incomplete.

Sampling is the procedure in which we choose the appropriate data in accordance or keeping in mind that we have to reduce the running time so as to obtain the real time results. We are using python in order to complete our sampling process.

2.2 Functional Diagram of Methodology-

It can be divided into 5 parts:

1. Data Collection- It may include the process of downloading the data set, it may be from caggle or any other reliable government website, prediction should be done on reliable data , thus we have downloaded this data from trusted and reliable government website

2. Classification

acts	address	datefrom	dateto	day	district
0	JHUGGI NO. A-603, SANJAY CAMP, CHANAKYA PURI, ...	2017-01-01	2017-01-01	Sunday	NEW DELHI
1	PANCHSHEEL MARG, ASHOKA HOTEL PARKING, CHANAKYA...	2017-01-05	2017-01-05	Thursday	NEW DELHI
2	H.NO. 773, 7TH FLOOR, BLOCK-7, BAPU DHAM, CHAN...	2014-04-21	2016-04-11	INTERVENING DAY	NEW DELHI
3	RIDGE ROAD TOWARDS KAROL BAGH, NEAR SIMON BOLI...	2017-01-01	2017-01-01	Sunday	NEW DELHI
4	MTC ROAD NEAR 11 MURTI	2017-01-13	2017-01-13	Friday	NEW DELHI

Fig.2 Classification

3. Pattern Identification- which help us finding pattern to apply k-means algorithm on data for clustering we required some specific data and also for visualization as we take specified columns i.e. day and districts and then plot them on bar graph to get conclusion.

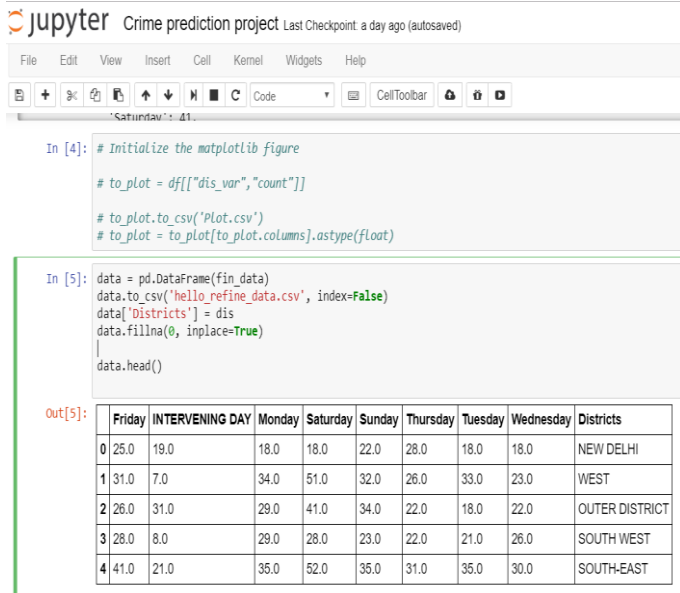


Fig.3-Pattern Identification

4. Prediction

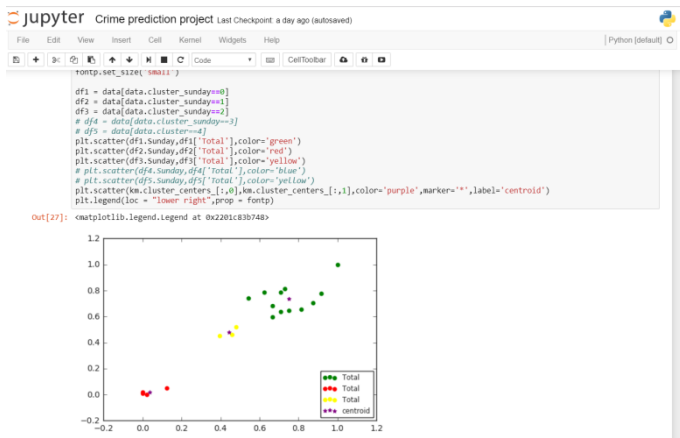


Fig.4 Prediction

5. Visualization (Novelty)- Here we will use folium to map the various regions in Delhi which lie in red or danger zone try to avoid those areas on that particular day of the week.

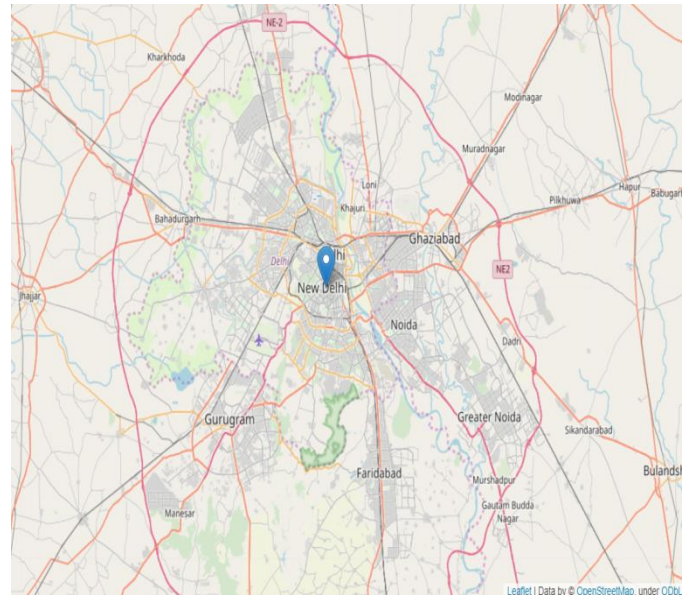
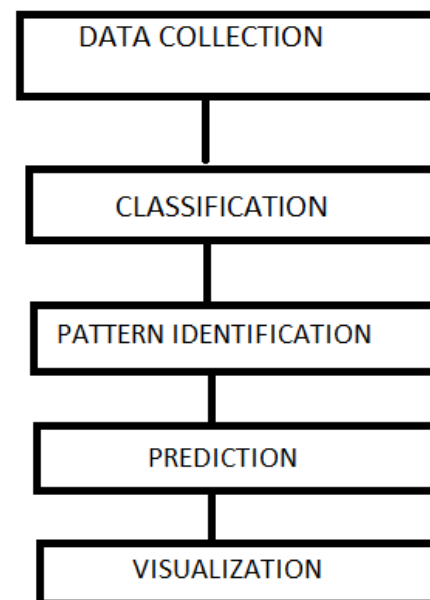


Fig.5-Visualization



IV. PROPOSED WORK

4.1 Here our main objective is to find or try to find some kind of relationship among different elements in our dataset and then use our algorithm to come up with the solution.

4.2 We could have used the traditional technique to clean the data first then work on the cleaned data, rather we have avoided this and used only those elements of dataset which are useful to us for ex-

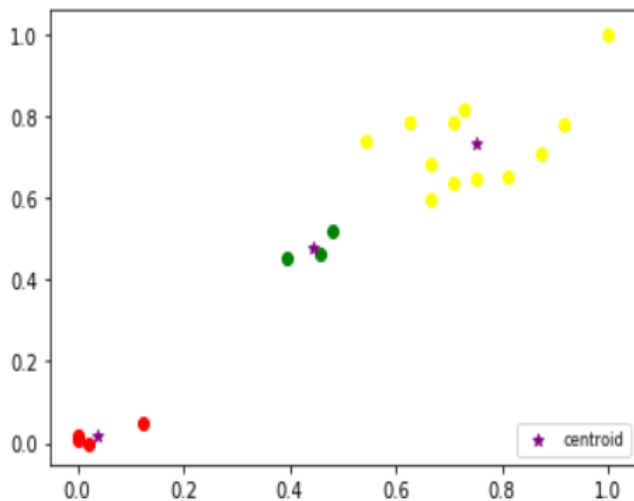
first we calculated total no of crimes that took place in a particular region on a particular day.[5]

4.3 Then we plotted all the data day wise and worked on that data in order to meet our objective to predict which region is safe on which day and avoid that region.

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

Fig.7-Day-wise crime data collection

After doing this, we have applied K-Means Clustering algorithm on our data, this will form clusters and this is tried and tested algorithm for prediction, hence we will obtain centroid as well as final cluster of data.



```
lt = {}
lt['Sunday'] = df2.Districts
```

Fig.8 K-Means Clustering

We can also use elbow method to calculate the centroid, which is one of the techniques used to find the centroid-

```
plt.xlabel('K')
plt.title('The Elbow Method')
plt.ylabel('Sum of squared error')
plt.plot(k_rng,sse)
```

[<matplotlib.lines.Line2D at 0x1e978d09b38>]

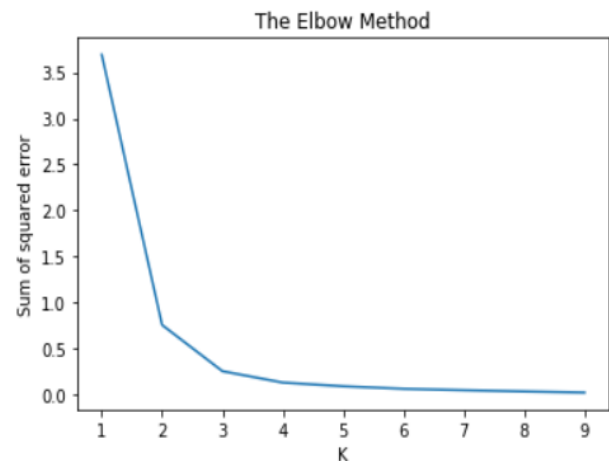


Fig.9 Elbow Method

The final cluster obtained will look something like this-

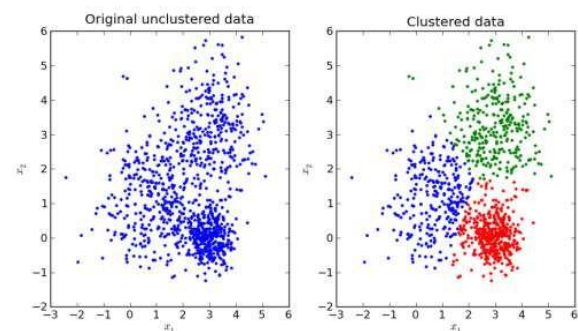


Fig-10 Clustering Sample Diagram

4.4 INNOVATION

Now we have to visualize this result, for this we will use folium which has never been used in other crime prediction researches. This is the novelty in our project. Folium will enhance decision making or we can say it will make easier for person to decide which areas to avoid in Delhi on a particular day.

Mapping different regions of Delhi with colors brings novelty in our project as we will map each region for different crimes and depending on the basis of color shown on our map we can decide or choose which area in Delhi is safe and maintain law and order. This will also enable government to cope with crime better.

```
#generate map
sf_map.choropleth(
    geo_data=gjson,
    data=table,
    columns=['Districts', 'Count'],
    key_on = 'feature.properties.Dist_Name',
    fill_color = 'Set3',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Crime Rate % in Delhi'
)
# folium.LayerControl().add_to(sf_map)
# Save to html
sf_map.save('Result.html')
```

Fig.11 Folium Implementation Code

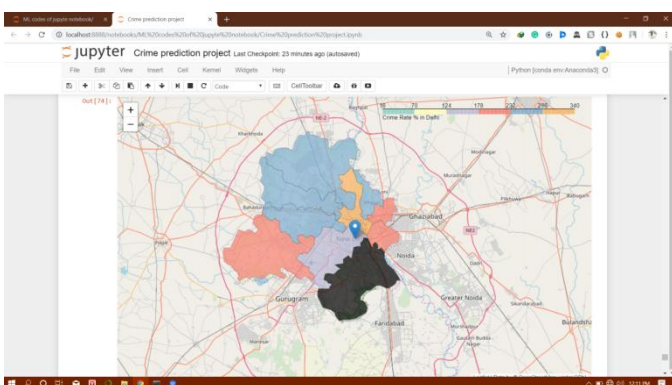


Fig.12 Folium Implementation

V. CONCLUSION

In this study our focus was to introduce more smarter way to avoid and anticipate crime. This study is really need of an honour our aim was to come up with accurate prediction model which is user friendly as well, someone from non technicalbackground can also understand our user interface made using folium. We made a project for

every citizen including male,female, old as well as young.

We adopted technique which is not traditional. Traditionally researchers used tpo clean data then work on it instead we identified the data elements that we can use and proceeded accordingly.

We met our objective of accurate and user friendly visualiation of our project and we achieved this through folium.

In the end we can conclude that our research and study can efficiently predict and visualize crime and we met our objective and our model can revelutionize the condition of crime in Delhi.

VI.ACKNOWLEDGEMENT-

We owe a very sincere thanks to our university CHANDIGARH UNIVERSITY for providing us platform to prepare a project on such important topic AN EFFICIENT CRIME FORECASTING USING FOLIUM and we would like to thank our project Mentor Assistant Professor Miss NavpreetKaur 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.

I would also like to thank Dr. Ruchika Gupta for guiding us and providing us with her valuable feedback which enabled to improve various aspects of our project. Her weekly assessment of the project led us to build reliable and accurate predicting model.

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.

- [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.
- [4] Piraján, Freddy, AndreyFajardo, and Miguel Melgarejo. "Towards a Deep Learning Approach for Urban Crime Forecasting." Workshop on Engineering Applications. Springer, Cham,2019.
- [5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. MagnetismJapan, p.301,1982].