



PROJECT REPORT

Machine Learning V1

SAMAN MANSOOR - F2019332020

ALI MUNAWAR - F2019332007

**MUHAMMAD TAIMOOR -
F2019332012**

**ZEESHAN MUSTAFA-
F2019332005**

Sameed _ f2019332008

Submitted to: Waqar Hussain

Project Report

Topic: Crime Prediction

Abstract:

Analysis of crime is a methodological approach to the identification and assessment of criminal patterns and trends. In a number of respects cost our community profoundly. We have to go many places regularly for our daily purposes, and many times in our everyday lives we face numerous safety problems such as hijack, kidnapping, and harassment. In general, we see that when we need to go anywhere at first, we are searching for Google Maps; Google Maps show one, two, or more ways to get to the destination, but we always choose the shortcut route, but we do not understand the path situation correctly. Is it really secure or not that's why we face many unpleasant circumstances; in this job, we use different clustering approaches of Machine Learning to analyze the crime rate of world and we also any algorithm like K-mean, K-nearest neighbor (KNN) algorithm to train our dataset? The biggest challenge will probably be "proving" to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative. Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology. We also attempt to make our classification task more meaningful by merging multiple classes into larger classes. Finally, we report and reflect on our results with different classifiers, and dwell on avenues for future work.

Introduction

Crime, in a way, influences organizations and institutions when occurred frequently in a society. Thus, it is necessary to study the factors and relations between different crimes and to find a way to accurately predict and avoid these crimes. Recently law enforcement agencies have been moving towards a more empirical, data driven approach to predictive policing. However, even with new data-driven approaches to predict crime, the fundamental job of crime analysts still remains difficult and often manual; specific patterns of crime are not very easy to find by way of automated tools, whereas larger-scale density-based trends comprised mainly of background crime levels are much easier for data-driven approaches and software to estimate. With the advent of the Big Data era and the availability of fast, efficient algorithms for data analysis, understanding patterns in crime from data is an active and growing field of research.

Literature Review

In this report we have studied the relationship between crime and different features in the criminology literature. Reduce the crime and detect the techniques of crime and stop the crime before the author uses different techniques. Use Z-Crime Tools and Advanced algorithm with Machine Learning technology to predict criminal activity. K-Nearest neighbors is used for classification to single and multi-class variable. The Neural Network is used for the prediction's precision. The precision rate by using Neural Network the model accuracy is 60, 93 and 97%. The Author presents a geographical analysis-based and self-regressive approach to automatically identify large danger urban crime areas and to represents crime patterns in each region reliably. Show the result of the algorithm of crime prediction system consisting of a collection of thick crime areas and a set of related crime forecaster. This operates primarily for the large region where the large amounts of individuals live and demonstrates that the suggested strategy achieves excellent precision over rolling time horizons in spatial and temporal crime forecasting. This paper's working process collects raw data the hotspot uses after splitting the data to create the new hotspot model and finally shows the predictive crime rate.

Through many documentations and cases, it came out that machine learning and data science can make the work easier and faster. The inputs to our algorithms are time (hour, day, month, and year), place (latitude and longitude), and class of crime:

- **Act 379 - Robbery**
- **Act 13 - Gambling**
- **Act 279 - Accident**
- **Act 323 - Violence**
- **Act 302 - Murder**
- **Act 363 - Kidnapping**

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests. We also perform multiple classification tasks – we first try to predict which of 6 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

Statement of the problem:

The main problem is that day to day the population is going to be increased and by that the crimes are also going to be Increased in different areas by this the crime rate cannot be accurately predicted by the officials. The officials as they focus on many issues may not predict the crimes to be happened in the future. The officials/police officers although they try to reduce the crime rate, they may not reduce in full-fledged manner. The crime rate prediction in future may be difficult for them. There has been countless of work done related to crimes. Large datasets have been reviewed, and information such as location and the type of crimes have been extracted to help people follow law enforcements. Existing methods have used these databases to identify crime hotspots based on locations. Even though crime locations have been identified, there is no information available that includes

the crime occurrence date and time along with techniques that can accurately predict what crimes will occur in the future.

Goal

Much of the current work is focused in two major directions:

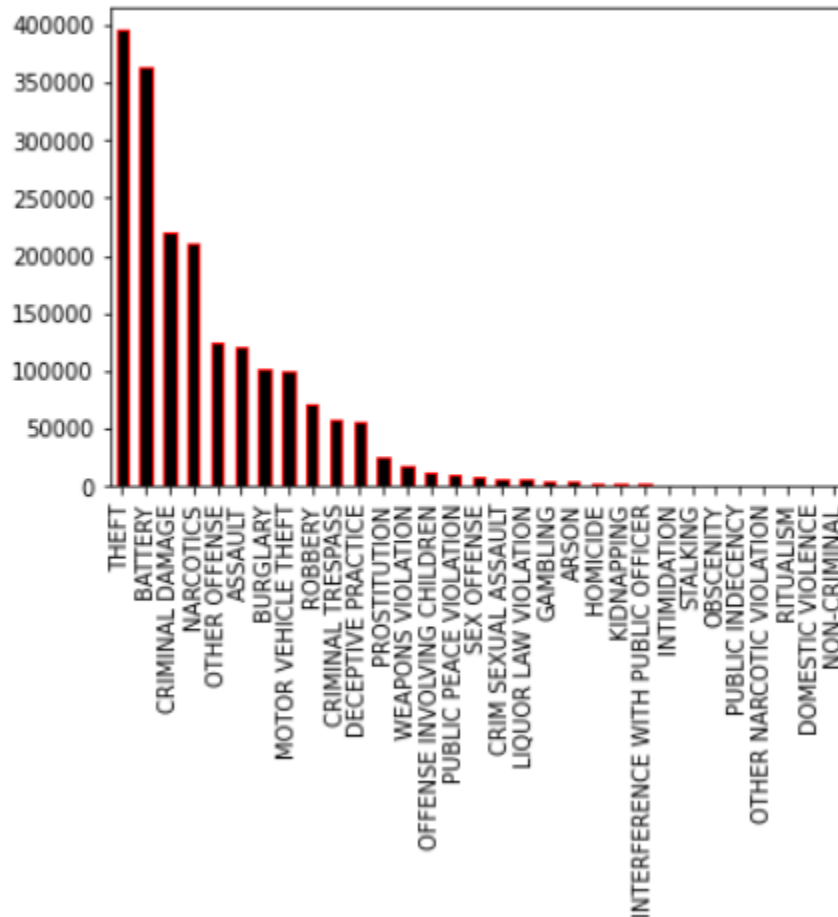
- Predicting surges and crime.
- Understanding patterns of criminal behavior that could help in solving criminal investigations.

Objective

The objective of our work is to:

- Predicting crime before it takes place.
- Understanding crime pattern.
- Classify crime based on location.
- Analysis of crime in Indore.

Trends and Patterns:



Methodology.

Dataset

The crime dataset is extracted from primary data collection based on field work. This dataset consists of about 1923515 and columns 23. The key features such as primary type, location area, month, date, domestic, longitude, and latitude.

Preprocessing

We decided to remove unknown values from the information set like Dates and times was recorded as a times window in MM/DD/YY HH/MM format from the documents as direct matching of these dates are complicated by

the classification system to match date and time with each other. Classification of the date was considered to classify a particular date into three groups as weekends, weekdays and unaware. This classification is based on the features of the date time windows. Explains the system's work flow. The workflow begins by extracting data from data collection, which is a dataset repository on different roles. The primary data will be preprocessed and transformed into a criminal data. Four target variables are predicted.

Algorithm

The domain contains many clustering algorithms. There is widespread use and acceptance of the K-means partitioning method. Apart from the K-means strategy, the Linear regression algorithm is the one we used because it enables consumers to determine the number of clusters based on those values Naïve Bayes is also pretend good result but above two are provides the best accuracy.

K-Nearest Neighbors

K-nearest neighbors is used when the target variable must be classified in more than two classes. In this dataset, there are three classes of target variable perpetrator sex: male, female, and unknown. Similarly, three categories of young, old and kid are defined in age. To classify these target variables, K-nearest neighbors Classifier is used.

$$D(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

Pseudo Code:

- At first KNN Classifier (Data Entry)
- Assign the Number of Cluster in K
- Choose a set of K instances to be cluster centers
- Data points for each output
- Calculate the Euclidean range
- Assign next to the data point the cluster

- Perpetually Calculate centroids and reassign the variables of the cluster
- Repeat until you reach an appropriate cluster Result.
- Give back the clusters and their values

Advantages

- Helps to prevent crime in society
- System will keep historical record of crime.
- System is user friendly
- Saves time

Disadvantages

- Users who don't have internet connection can't access the system.
- Admin must enter correct records otherwise system will provide wrong information

Implementation

```
In [51]: from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier()
classify(model, X, y)
```

Accuracy is 93.57697050609404

C:\Users\Taimoor\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\model_selection_split.py:680: UserWarning:
The least populated class in y has only 1 members, which is less than n_splits=5.
warnings.warn(

Cross validation is 93.53293319781754

```
In [57]: y_pred=model.predict(X)
```

```
In [58]: y_pred
```

```
Out[58]: array([29, 21, 19, ..., 29,  3, 29])
```



```
print("Training Score:\n",model.score(x_train,y_train)*100)
print("Mean Squared Error:\n",mean_squared_error(y_test,y_pred))
print("R2 score is:\n",r2_score(y_test,y_pred))
```

```
Training Score:
92.20683526544464
Mean Squared Error:
1.8555748951399416
R2 score is:
0.9841753250860553
```

Result and Discussion

In this part summarizes the paper and make aware about the future crime based on algorithms and crime data set we find out the crime rate in various section like age based, male vs female, area based and monthly crime rates. The data sources and methods used to guide forecasting include various type crime statistics, survey of the general people data, literature reviews and statistical models that extrapolate crime trends into the future. Algorithms models that describe the behavior of observe past values can be used to forecast future crime trends by projecting a time series analysis of crime trends into the future.

Reference.

- Feng, A., and Allan, J. 2007. Finding and linking incidents in news. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, 821–830. ACM.
- Gerber, M. S. 2014. Predicting crime using twitter and kernel density estimation. Decision Support Systems 61:115– 125.
- Anon: [online] Available at https://www.researchgate.net/publication/322000460_A_review_

Crime_analysis_using_data_mining_techniques_and_algorithms.
Accessed 30 Aug. 2019 (2019)

- Kerr, J.: Vancouver police go high tech to predict and prevent crime before it happens. Vancouver Courier, July 23, 2017. [Online] Available <https://www.vancourier.com/news/vancouver-police-go-high-tech-to-predict-and-prevent-crime-before-it-happens-1.21295288>. Accessed 09 Aug 2018
- Marchant, R., Haan, S., Clancey, G., Cripps, S.: Applying machine learning to criminology: semi parametric spatial demographic Bayesian regression. Security Inform. (2018)