ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

CRIME DATA ANALYSIS USING MACHINE LEARNING

M. kalyan sagar
Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad-500090, Telangana State, India
Email: kalyansagarmarka@gmail.com

P Gopala Krishna
Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad-500090, Telangana State, India
Email: gopalakrishna@griet.ac.in

ABSTRACT:
Crime is one of the biggest and dominating problem in our society and its prevention is an important. task. Daily there are huge numbers of crimes committed frequently. This require keeping track of all the crimes and maintaining a database for same which may be used for future reference. The current problem faced are maintaining of proper dataset of crime and analyzing this data to help in predicting and solving crimes in future. The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. In this project, we will be using the technique of machine learning and data science for crime prediction of Chicago crime data set. The crime data is extracted from the official portal of Chicago police. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training of the model data preprocessing will be done following this feature selection and scaling will be done so that accuracy obtain will be high. The K-Nearest Neighbor (KNN) classification and various other algorithms will be tested for crime prediction and one with better accuracy will be used for training. Visualization of dataset will be done in terms of graphical representation of many cases for example at which time the criminal rates are high or at which month the criminal activities are high. The soul purpose of this project is to give a jest idea of how machine learning can be used by the law enforcement agencies to detect, predict and solve crimes at a much faster rate and thus reduces the crime rate. It not restricted to Chicago, this can be used in other states or countries depending upon the availability of the dataset.

Keywords –K-Nearest Neighbor Support, Vector Machine Autoregressive moving average, recurrent neural network

1. INTRODUCTION

Crimes are the significant threat to the humankind. There are many crimes that happens regular interval of time. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type – robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. 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.

ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

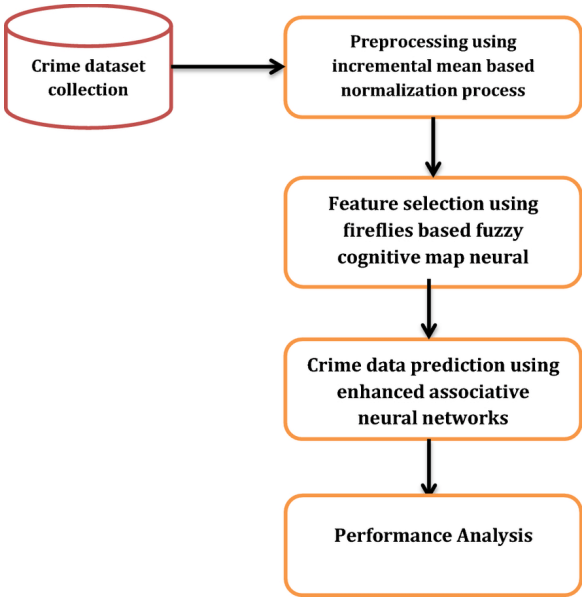

Fig.1: Example figure

The above problem made me to go for a research about how can solving a crime case made easier. Through many documentation and cases, it came out that machine learning and data science can make the work easier and faster.

The aim of this project is to make crime prediction using the features present in the dataset. The dataset is extracted from the official sites. With the help of machine learning algorithm, using python as core we can predict the type of crime which will occur in a particular area.

The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. Building the model will be done using better algorithm depending upon the accuracy. The K-Nearest Neighbor (KNN) classification and other algorithm will be used for crime prediction. Visualization of dataset is done to analyze the crimes which may have occurred in the country. This work helps the law enforcement agencies to predict and detect crimes in Chicag.

2. LITERATURE REVIEW

Crime pattern detection, analysis & prediction:

Criminal analysis is a methodical approach for identifying and analyzing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, system can analyze previously unknown, useful information from an unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, system will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. In this paper, the main focus is on the review of algorithms and techniques used for identify the criminals.

An overview on crime prediction methods:

In the recent past, crime analyses are required to reveal the complexities in the crime dataset. This process will help the parties that involve in law enforcement in arresting offenders and directing the crime prevention strategies. The ability to predict the future crimes based on the location, pattern and time can

ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

serve as a valuable source of knowledge for them either from strategic or tactical perspectives. Nevertheless, to predict future crime accurately with a better performance, it is a challenging task because of the increasing numbers of crime in present days. Therefore, crime prediction method is important to identify the future crime and reduces the numbers of crime. Currently, some researchers have been conducted a study to predict crime based on particular inputs. The performance of prediction models can be evaluated using a variety of different prediction methods such as support vector machine, multivariate time series and artificial neural network. However, there are still some limitations on their findings to provide an accurate prediction for the location of crimes. A large number of research papers on this topic have already been published previously. Thus, in this paper, we thoroughly review each of them and summarized the outcomes. Our objective is to identify current implementations of crime prediction method and the possibility to enhance it for future needs.

Crime prediction and forecasting in Tamilnadu using clustering approaches:

Crime is one of the most predominant and alarming aspects in our society and its prevention is a vital task. Crime analysis is a systematic way of detecting and investigating patterns and trends in crime. In this work, we use various clustering approaches of data mining to analyse the crime data of Tamilnadu. The crime data is extracted from National Crime Records Bureau (NCRB) of India. It consists of crime information about six cities namely Chennai, Coimbatore, Salem, Madurai, Thirunelvelli and Thiruchirapalli from the year 2000-2014 with 1760 instances and 9 attributes to represent the instances. K-Means clustering, Agglomerative clustering and Density Based Spatial Clustering with Noise (DBSCAN) algorithms are used to cluster crime activities based on some predefined cases and the results of these clustering are compared to find the best suitable clustering algorithm for crime detection. The result of K-Means clustering algorithm is visualized using Google Map for interactive and easy understanding. The K-Nearest Neighbor (KNN) classification is used for crime prediction. The performance of each clustering algorithms are evaluated using the metrics such as precision, recall and F-measure, and the results are compared. This work helps the law enforcement agencies to predict and detect crimes in Tamilnadu with improved accuracy and thus reduces the crime rate.

Crime analysis and prediction using data mining:

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Instead of focusing on causes of crime occurrence like criminal background of offender, political enmity etc we are focusing mainly on crime factors of each day.

Crime pattern detection using data mining:

Data mining can be used to model crime detection problems. Crimes are a social nuisance and cost our society dearly in several ways. Any research that can help in solving crimes faster will pay for itself. About 10% of the criminals commit about 50% of the crimes. Here we look at use of clustering algorithm for a data mining approach to help detect the crimes patterns and speed up the process of solving crime. We will look at k-means clustering with some enhancements to aid in the process of identification of crime patterns. We applied these techniques to real crime data from a sheriff's office and validated our results. We also use semi-supervised learning technique here for knowledge discovery from the crime records and to help increase the predictive accuracy. We also developed a weighting scheme for attributes here to deal with limitations of various out of the box clustering tools and techniques. This easy to implement data mining framework works with the geospatial plot of crime and helps to improve the productivity of the detectives and other law enforcement officers. It can also be applied for counter terrorism for homeland security.

ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

3. METHODOLOGY

In existing studies crime prediction is done on Chicago data set in which various machine learning models are used. Comparison of models like KNN, Naïve Bayes, SVM is done this paper. It is seen that prediction varies depending upon the dataset and features that have been selected. The prediction accuracy found in is 78% for KNN, 64% for GaussianNB, 31% for SVC. Auto regressive integrated Moving average models were used in to make machine learning algorithms to forecast crime trends in urban areas. One of the major problems in crimes is detecting and analyzing the pattern of crimes. Understanding datasets is also an important concept in this case. We surely want to accurately predict so that we don't waste our resources due to false signals. Also proposed a method for classifying the crime rate as high, medium or low. None of them has classified the type of crime that can happen and its probability of happening. Analysis and prediction of crime is an important activity that can be optimized using various techniques and processes. Lot of research work is done by various researchers in this domain.

Disadvantages:

1. The existing work is limited to use the datasets to identify locations of crime.

In this project, we will use data science and machine learning to predict crimes using a set of crime data from Chicago. The Chicago police's official portal is where the crime statistics are taken from. It includes details on the offence, including the time, date, place, and kind of crime. Data preprocessing will be performed prior to training the model, and this will be followed by feature selection and scaling to ensure high accuracy. Several other algorithms, including the K-Nearest Neighbor (KNN) classification, will be examined for their ability to forecast crimes, and the algorithm that performs the best will be used to train others.

Advantages:

1. The main goal of this project is to provide a basic understanding of how machine learning may be utilized by law enforcement organizations to identify, anticipate, and solve crimes considerably more quickly, which lowers the crime rate.

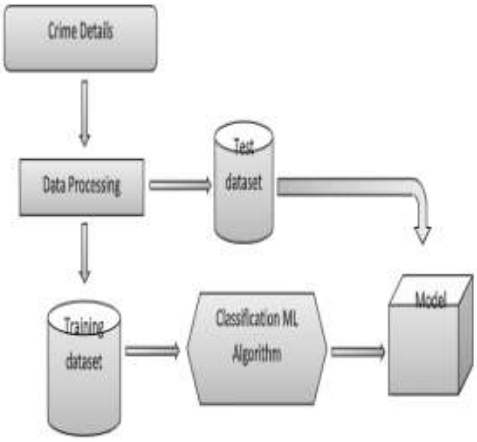

Fig.2: System architecture

Exploratory Data Analysis of crime rate Prediction Machine learning is a computer system's method of learning by way of examples. There are many machine learning algorithms available to users that can be implemented on datasets. However, there are two major types of learning algorithms: supervised learning and unsupervised learning algorithms. Supervised learning algorithms work by inferring information or "the right answer" from labeled training data. The algorithms are given a particular attribute or set of attributes to predict. Data preprocessing process includes methods to remove any null values or infinite values which may affect the accuracy of the system. The main steps include Formatting, cleaning and

ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

sampling. Cleaning process is used for removal or fixing of some missing data there may be data that are incomplete. Crimes Prediction ways:

- To utilize the resources identify the hotspots of crimes and allocate vigilante resources such as policeman, police cars, weapons etc. reschedule patrols according to the vulnerability of a place.
- Through that avoid crimes Ensure better civilization through avoiding happening crimes such as murder, rapes, thefts, drug, smugglings etc.

4. IMPLEMENTATION

The implementation of this project is divided into following steps:

Data collection:
Crime dataset from kaggle is used in CSV format.

Data Preprocessing:
10k entries are present in the dataset. The null values are removed using df = df.dropna() where df is the data frame. The categorical attributes (Location, Block, Crime Type, Community Area) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like month and hour which can be used as feature for the model.

Feature selection:
Features selection is done which can be used to build the model. The attributes used for feature selection are Block, Location, District, Community area, X coordinate, Y coordinate, Latitude, Longitude, Hour and month,

Building and Traning Model:
After feature selection location and month attribute are used for training. The dataset is divided into pair of xtrain ,ytrain and xtest, y test. The algorithms model is imported form skleran. Building model is done using model. Fit (xtrain, ytrain).

Prediction:
After the model is build using the above process, prediction is done using model. predict(xtest). The accuracy is calculated using accuracy_score imported from metrics - metrics.accuracy_score (ytest, predicted).

Visualization:
Using mathpoltlib library from sklearn. Analysis of the crime dataset is done by plotting various graphs.

5. EXPERIMENTAL RESULTS

Fig.3: webpage address

ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

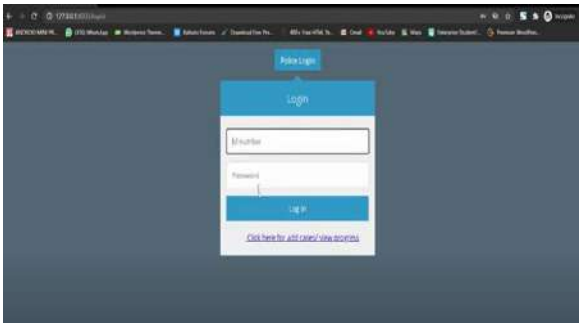

Fig.4: User signin

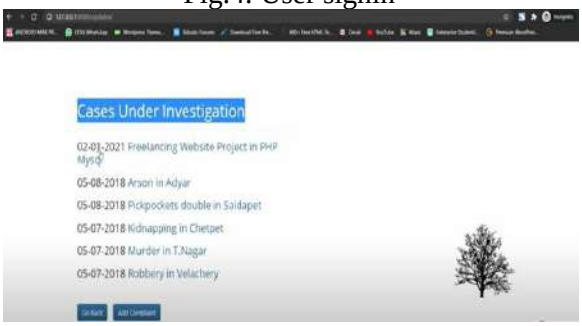

Fig.5: Cases under investigation

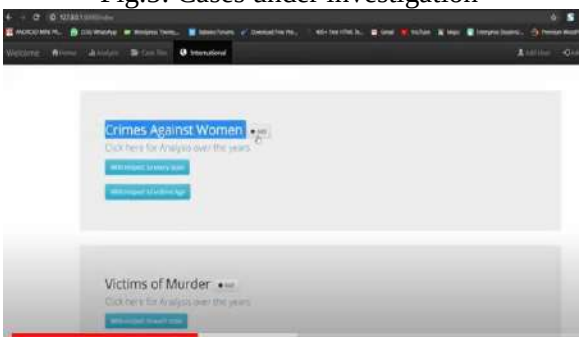

Fig.6: Crime details

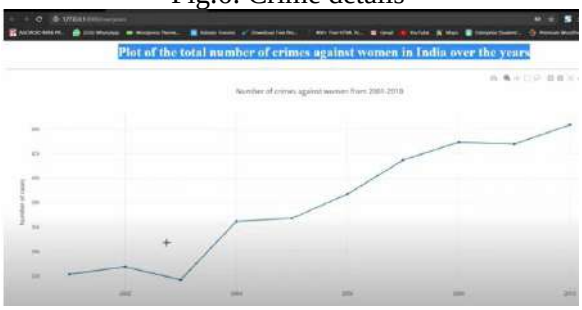

Fig.7: Total number of crimes graph

ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

Fig.8: State wise graph

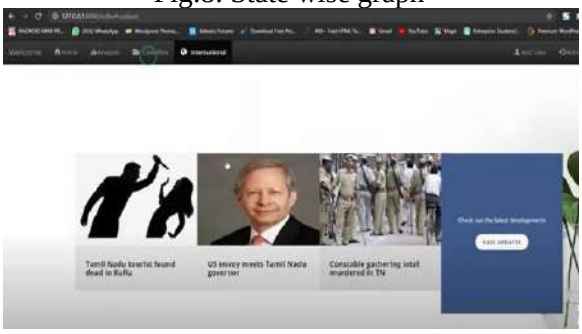

Fig.9: case updating

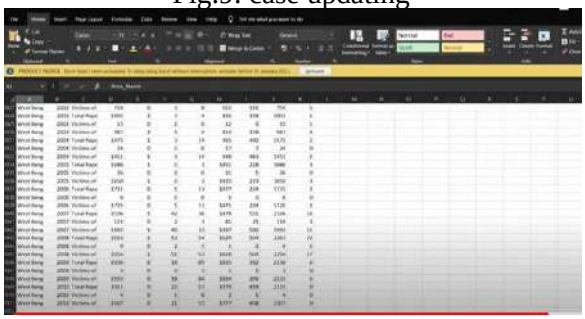

Fig10: Dataset

6. CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns among various data. The work in this project mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning we have built a model using training data set that have undergone model predicts the type of crime with accuracy of 0.789. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding Chicago crimes datasets that can help in capturing the factors that can help in keeping society safe.

REFERENCES

[1] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and AerospaceTechnology (ICECA), 2017 International conference of (Vol. 1, pp. 225- 230). IEEE.

[2] Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.

[3] Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016, October). Crime prediction and forecasting in Tamilnadu using clustering approaches. In Emerging Technological Trends (ICETT) International Conference on (pp. 1-6). IEEE.

[4] Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp. 406-412). IEEE.

[5] Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on (pp. 41-44). IEEE.

ISSN: 0972-2750
Vol-12 Issue-12 No.05: 2022

[6] Zhao, X., & Tang, J. (2017, November). Exploring Transfer Learning for Crime Prediction. In Data Mining Workshops (ICDMW), 2017 IEEE International Conference on (pp. 1158-1159). IEEE.

[7] Al Boni, M., & Gerber, M. S. (2016, December). AreaSpecific Crime Prediction Models. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 671-676). IEEE.

[8] Tayebi, M. A., Gla, U., & Brantingham, P. L. (2015, May). Learning where to inspect: location learning for crime prediction. In Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on (pp. 25-30). IEEE.

[9] E. W. T. Ngai, L. Xiu, and D. C. K. Chau. "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification," Expert Systems with Applications, pp. 2592–2602, 2008.

[10] J. McCarthy, "Arthur Samuel: Pioneer in Machine Learning," AI Magazine, vol. 11, no. 3, pp. 10-11, 1990.