

CRIME RATE PREDICTION

Ch. Mahendra¹, G. Nani Babu², G. Balu Nitin Chandra³, A. Avinash⁴, Y. Aditya⁵

¹ Department of Computer Science & Engineering, Gudlavalleru Engineering College, India

² Department of Computer Science & Engineering, Gudlavalleru Engineering College, India

³ Department of Computer Science & Engineering, Gudlavalleru Engineering College, India

⁴ Department of Computer Science & Engineering, Gudlavalleru Engineering College, India

⁵ Assistant Professor, Department of Computer Science & Engineering, Gudlavalleru Engineering College, India

mahendrachilakalapudi@gmail.com, ghantasalanani@gmail.com, garlapatibalu11677@gmail.com,

avinashadapa98@gmail.com adityaalu@gmail.com

Abstract - Crime is one of the biggest and dominating problem in our society. Daily there are huge number of crimes committed frequently. Here the dataset consists of the date and the crime rate that has taken place in the corresponding years. In this project the crime rate is only based on the robbery. We use linear regression algorithm to predict the percentage of the crime rate in the future by using the previous data information. The date is given as an input to the algorithm and the output is the percentage of the crime rate in that particular year.

Keywords: Crime rate, number of crimes, regression algorithm, Machine learning.

1. INTRODUCTION

Crimes are the significant threat to the humankind. There are many crimes that happen in regular intervals 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.

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 with crime percapita.

The objective would be to train a model for prediction. The training would be done using Training data set which will be validated using the test dataset. The Multi Linear Regression (MLR) will be used for crime prediction. Visualization of dataset is done to analyze the crimes which may have occurred in a particular year and based on population and number of crimes. This work helps the law enforcement agencies to predict and detect the crime_percapita in an area and thus reduces the crime rate.

MACHINE LEARNING:

Machine Learning is a sub-area of artificial intelligence, whereby the term refers to the ability of IT systems to independently find solutions to problems by recognizing patterns in databases. In other words: Machine

Learning enables IT systems to recognize patterns on the basis of existing algorithms and data sets and to develop adequate solution concepts. Therefore, in Machine Learning, artificial knowledge is generated on the basis of experience.

In order to enable the software to independently generate solutions, the prior action of people is necessary. For example, the required algorithms and data must be fed into the systems in advance and the respective analysis rules for the recognition of patterns in the data stock must be defined. Once these two steps have been completed, the system can perform the following tasks by Machine Learning:

2. OBJECTIVES OF THE PROJECT

→The main objective of the project is to predict the crime rate and analyze the crime rate to be happened in future. Based on this Information the officials can take charge and try to reduce the crime rate.

→The concept of Multi Linear Regression is used for predicting the graph between the Types of Crimes (Independent Variable) and the Year (Dependent Variable)

→The system will look at how to convert crime information into a regression problem, so that it will help detectives in solving crimes faster.

→Crime analysis based on available information to extract crime patterns. Using various multi linear regression techniques, frequency of occurring crime can be predicted based on territorial distribution of existing data and Crime recognition.

3. PROBLEM STATEMENT

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 can't 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 tries 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. There are several maps applications that show the exact crime location along with the crime type for any given city (see Figure 1). 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.

4. LITERATURE REVIEW

Crime rate prediction is different in various applications, some of the studies are given below:

C.P. Chaithanya, N. Manohar, Ajay Bazil Issac, describes Text detection is the method of locating areas in a picture wherever, text is present. Text detection and classification in natural pictures is very important for several computer vision applications like optical character recognition, distinguish between human and machine inputs and spam removal. Currently the challenge in text identifying is to detect the text in natural pictures due to many factors like, low- quality image, unclear words, typical font, image having a lot of color stroke than the background color, blurred pictures due to some natural problems like rain, sunny, snow, etc. The main aim of this work is to identify and classify the text in natural pictures. Here system detects the text and finds the connected regions, chain them together in their relative position. Uses a text classification engine to filter chains with low classification confidence scores [1].

Shiju Sathyadevan, Devan M.S, proposed that Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime sensitive areas. So for building such a powerful crime analytics tool we have to collect crime records and evaluate it .It is only within the last few decades that the technology made spatial data mining a practical solution for wide audiences of Law enforcement officials which is affordable and available. Since the availability of criminal data or records is limited we are collecting crime data from various sources like web sites, news sites, blogs, social media, RSS feeds etc. This huge data is used as a record for creating a crime record database. So the main challenge in front of us is developing a better, efficient crime pattern detection tool to identify crime patterns effectively [2].

PAC Duijn, V Kashirin, proposed, policymakers and law enforcement agencies across the globe struggle to find effective strategies to control criminal networks. The effectiveness of disruption strategies is known to depend on both network topology and network resilience. However, as these criminal networks operate in secrecy, data-driven knowledge concerning the effectiveness of different criminal network disruption strategies is very limited. By combining computational modeling and social network analysis with unique criminal network intelligence data from [3].

Nishat Shama proposed that Criminal activities are present in every region of the

world affecting quality of life and socio-economical development. As such, it is a major concern of many governments who are using different advanced technology to tackle such issues. Crime Analysis, a sub branch of criminology, studies the behavioral pattern of criminal activities and tries to identify the indicators of such events. Machine learning agents work with data and employ different techniques to find patterns in data making it very useful for predictive analysis. Law enforcement agencies use different patrolling strategies based on the information they get to keep an area secure. A machine learning agent can learn and analyze the pattern of occurrence of a crime based on the reports of previous criminal activities and can find hotspots based on time, type or any other factor. This technique is known as classification and it allows to predict nominal class labels. Classification has been used on many different domains such as financial market, business intelligence, healthcare, weather forecasting etc. In this research, a dataset from San-Francisco Open Data[8] is used which contains the reported criminal activities in the neighbour hoods of the city San Francisco for a duration of 12 years. I used different classification techniques like Decision Tree, Naive Bayesian, Logistic Regression, k-Nearest Neighbour, Ensemble Methods to find hotspots of criminal activities based on the time of day. Results of different algorithms have been compared and most the effective approach has also been documented [4].

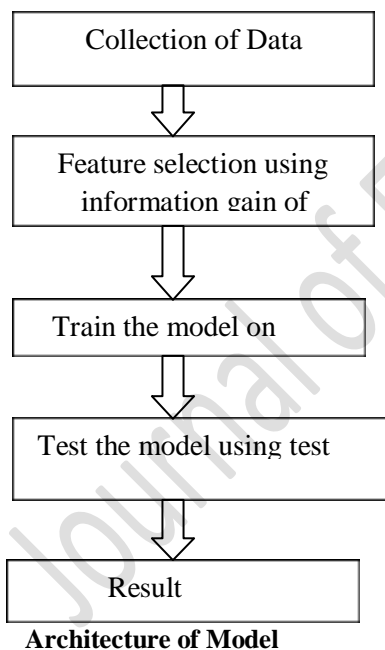
Riya Rahul Shah proposed that Vancouver is most populated city in Canada. It is most ethnically diverse cities in Canada. Crime is one of the biggest and dominating problem in our society and its prevention is an important task [1]. Even though Vancouver known to be the safest city it is observed that vehiclebreak-ins and many more thefts is still a problem. There has been tremendous increase in machine learning algorithms that have made crime prediction feasible based on past data. The aim of this project is to perform analysis and prediction of crimes in states using

machine learning models. It focuses on creating a model that can help to detect the number of crimes by its type in a particular state. In this project various machine learning models like K-NN, boosted decision trees will be used to predict crimes. Area Wise geographical analysis can be done to understand the pattern of crimes. Various visualization techniques and plots are used which can help law enforcement agencies to detect and predict crimes with higher accuracy. This will indirectly help reduce the rates of crimes and can help to improve securities in such required areas [5].

PROPOSED METHODOLOGY

METHODOLOGY:

This project has undergone the following process:



Initially, the data set is prepared manually. After identifying the relationships and visualizing the data, we create a regression model for forecasting the percapita. For this model, we have used Multi Linear regression model. Other models such as the Linear

Regression and Logistic Regression models were also tested, but the Multi Linear regression produced the minimal error while training the model. This regression model predicts the percapita of Crime rate that is going to be happen in future by taking different parameters.

Logistic Regression: In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist.

In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression). Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail which is represented by an indicator variable, where the two values are labeled "0" and "1".

KNN: In pattern recognition, the k -nearest neighbors algorithm (k -NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k -NN is used for classification or regression. In k -NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor. In k -NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors.

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship

between the explanatory (independent) variables and response (dependent) variable. So far, we have seen the concept of simple linear regression where a single predictor variable X was used to model the response variable Y. In many applications, there is more than one factor that influences the response. Multiple regression models thus describe how a single response variable Y depends linearly on a number of predictor variables. In essence, multiple regression is the extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable.

RESULTS

The Support vector Regression model results in classifying the test data. It predicts the stock crime rate value that is going to be happening in future. It uses train set to train and fit the model. The test data is transformed and predicts the accurate result

By Using Django and Aws we prepared the UserInterface.

The below figure describes that home page takes the year, murder, attempts, cheating etc., values as input in text boxes.

The below figure describes that new page which displays the predicted crime rate of the inputs produced by our algorithms.

Crime Rate Prediction

Crime rate is :[3930.8491711]

CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. The work in this project mainly revolves around predicting the type of crime and crime percapita which may happen in future. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation using Multi Linear Regression Algorithm. The model predicts the type of crime and Data visualization helps in analysis of data set and prediction of crimes. The graphs include bar,

line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding different crime datasets that can help in capturing the factors that can help in keeping society safe.

REFERENCES

- [1] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
- [2] Machine Learning in Production: Developing and Optimizing Data Science Workflows and Applications (Addison-Wesley Data & Analytics)
- [3] Python : The Complete Reference - Martin C. Brown
- [4] <https://docs.anaconda.com/anaconda/install/windows/>
- [5] A.U.S.S Pradeep, Soren Goyal, J. A. Bloom, I. J. Cox, and M. Miller, —Detection of statistical arbitrage using machine learning techniques in Indian Stock market, IIT Kanpur, April 15, 2013.
- [6] Prashant S. Chavan, Prof. Dr. Shrishail. T. Patil —Parameters for Stock Market Prediction, Prashant S Chavan et al, Int.J.Computer Technology & Applications, Vol 4 (2), 337-340.
- [7] Neelima Budhani, Dr. C. K. Jha, Sandeep K. Budhani —Prediction of Stock Market Using Artificial Neural Network, International Conference on Soft Computing Techniques for Engineering and Technology (ICSCTET)- 2014.
- [8] Sharvil Katariya, Saurabh Jain —Stock Price Trend Forecasting using Supervised Learning Methods.
- [9] Dhiraj Mundada, Gaurav Chhaparwal, Sachin Chaudhari, and Trupti Bhamare — Stock Value Prediction System, International Journal on Recent and Innovation Trends in Computing and Communication, April 2015
- [10] Hochreiter, Sepp, and Jürgen Schmidhuber — Long short-term memory, Neural computation 9.8 (1997): 1735-1780