Empirical Analysis for Crime Prediction and Forecasting using Machine Learning and Deep Learning Techniques

Dr. Suma T[1], Megha C[2], Mittal Savan Kumar[3], Mahesh Jadhav[4]

Assistant Professor, Department of Computer Science[1]

Sri Venkateshwara College of Engineering, Bangalore, India

suma.t_cse@svcengg.edu.in

Abstract: *Crime Forecasting refers to the basic process of predicting crimes before they occur. Crimes are a common social problem affecting the quality of life and the economic growth of a society. A crime is a deliberate act that can cause physical or psychological harm, as well as property damage or loss, and can lead to punishment by a state or other authority according to the severity of the crime. For our daily purposes we have to go many places every day and many times in our daily lives we face numerous security issues such as hijacking, kidnapping, harassment, etc. Daily there are huge numbers of crimes occurring frequently. These 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 main objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime which might occur in future depending upon various conditions. We will be using the technique of machine learning and data science for crime prediction of Chicago and Los Angeles crime data set. 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. The main purpose of this project is to give a brief 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 reduce the crime rate. It is not restricted to Chicago and Los Angeles, this can be used in other states or countries depending upon the availability of the dataset.*

Keywords: Crime Forecasting

I. INTRODUCTION

Crime is defined as an illegal act for which a person can be punished by the government. There are many types of crimes that keep occurring in a regular interval of time. Perhaps it is increasing and spreading at a fast rate. Crimes are of different types – robbery, murder, rape, kidnapping and so on. . The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the criminal activities. The major problem faced by the Police department is Crime prediction because there are enormous amount of crime data that exist. There is a need of technology through which the case solving couldbe faster.

II. CONCEPTS OF THE PROPOSED SYSTEM

2.1 Predictive Modeling

Predictive modeling is a mathematical process used topredict future events or outcomes by analysing patterns in a given set of input data. Predictive modelling can be divided further into two areas: Regression and pattern classification. Pattern classification tasks can be divided into two parts, Supervised and unsupervised learning.

III. IMPLEMENTATION

The dataset used in this project is taken from Kaggle.com.

The implementation of this project can be done in the following steps:

- **Data Collection:** Crime dataset from kaggle is used in CSV format.

- **Data Pre-processing:** This is the first step in Machine Learning. We need to clean the data and remove the unnecessary noises, null data and missing values.
- **Feature Selection:** The attributes used for feature selection are Block, Location, District, Community area, X coordinate, YCoordinate, Latitude, Longitude, Hour and month.
- **Building and Training Model:** The dataset is divided into pair of xtrain, ytrain and xtest, ytest.
- **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:** To analyse the Crime dataset in the Graphical Representation such as pie chart and bar graph. More than the written results the pictorial view would be easier to understand and analyse.

IV. RESULTS

The graph below shows the arrest ratio made in the city. 67.2 % of crimes committed by the criminals are not arrested and rest 32.8 % are arrested.
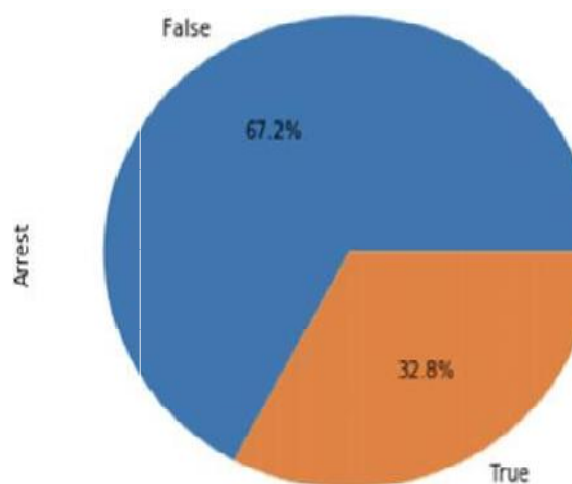

Figure 1: Percentage of arrest made

4.1 Forecasting with an ARIMA MODEL

An Autoregressive Integrated Moving Average OR **ARIMA MODEL** is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.

The algorithm first splits the data set into training and testing sets (70% and 30%, of the original data, respectively). A variety of models are currently available in the literature, and ARIMA models are considered a standard method for time series forecasting.

Figure 2 - shows the annual trends for Chicago, showing a significant decrease in the crime rate, while Los Angeles shows an increasing trend in recent years.
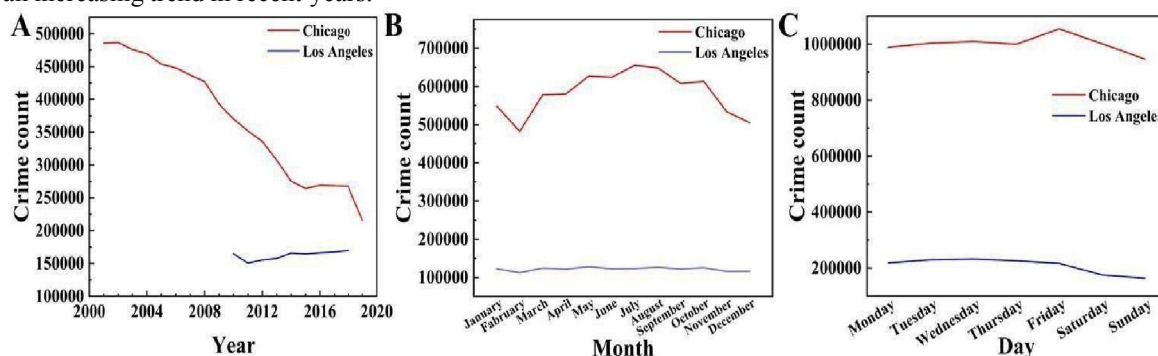

Figure 2-Yearly, monthly and daily crime rate trend analysis.

Figure 3 - Graph shows which crimes have occurred most in the city. The x coordinate denotes the Types of crimes committed and y coordinate denotes the number of crimes committed.
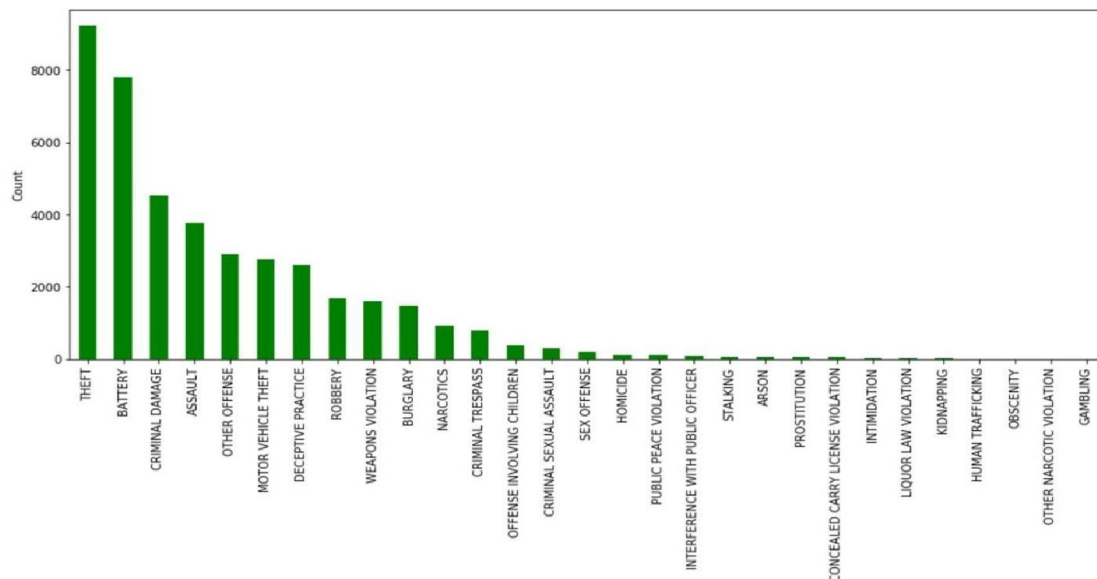

Figure 3- Types of crimes versus Number of crimes occurred

V. CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. Predicting crimes before they happen is simple to understand, but it takes a lot more than understanding the concept to make it a reality.

Using the concept of machine learning we have built a model using training data set that have undergone data cleaning. 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 and Los Angeles crimes datasets that can help in capturing the factors that might 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 Aerospace Technology (ICECA),2017 International conference of (Vol. 1, pp. 225-230). IEEE.

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

[3]. A. J. Hussain, P. Liatsis, M. Khalaf, H. Tawfik, and H. Al-Asker, ''A dynamic neural network architecture with immunology inspired optimization for weather data forecasting,'' Big Data Res., vol. 14, pp. 81–92, Dec. 2018.

[4]. 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.

[5]. M. Alsharif, M. Younes, and J. Kim, ''Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea,'' Symmetry, vol. 11, no. 2, p. 240, Feb. 2019

[6]. Chicago Data Portal. Accessed: Nov. 2, 2019.[Online].Available:https://data.cityofchicago.org/Public-Safety/ Crimes-2001-topresent- Dashboard/5cd6ry5g.

[7]. Los Angeles County GIS Data Portal. Accessed: Nov. 2,2019.[Online]. Available:http://egis3.lacounty. gov/data portal/?s=crime