

CRIME DATA ANALYSIS**KOMBATHUAL D.N. M. KUMAR*, N SRINIVASA RAO******PG SCHOLAR*, ASSISTANT PROFESSOR******E-Mail Id: meenakumar792@gmail.com*, naagaasrinu@gmail.com******SKBR PG COLLEGE, AMALAPURAM, E.G.DIST, ANDHRA PRADESH – 533201****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.

Key Words: K-Nearest Neighbor Support, Vector Machine Autoregressive moving average, recurrent neural network, Recursive Feature Elimination, National Crime Records Bureau

1. INTRODUCTION

1.1 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.

1.2 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.

1.3 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. 1.4 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 Chicago with improved accuracy and thus reduces the crime rate.

2. CONCEPTS OF THE PROPOSED SYSTEM

2.1 Predictive Modeling Predictive modeling is the way of building a model that is capable of making predictions. The process includes a machine learning algorithm that learns certain properties from a training dataset in order to make those predictions. Predictive modeling can be divided further into two areas: Regression and pattern classification. Regression models are based on the analysis of relationships between variables and trends in order to make predictions about continuous variables. In contrast to regression models, the task of pattern classification is to assign discrete class labels to particular data value as output of a prediction. Example of a classification model is - A pattern classification task in weather forecasting could be the prediction of a sunny, rainy, or snowy day. Pattern classification tasks can be divided into two

parts, Supervised and unsupervised learning. In supervised learning, the class labels in the dataset, which is used to build the classification model, are known. In a supervised learning problem, we would know which training dataset has the particular output which will be used to train so that prediction can be made for unseen data.

Types of Predictive Models Algorithms

Classification and Decision Trees □ A decision tree is an algorithm that uses a tree shaped graph or model of decisions including chance event outcomes, costs, and utility. It is one way to display an algorithm.

Naive Bayes -In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem with independence assumptions between the features. The technique constructs classifier models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set.

Linear Regression – The analysis is a statistical process for estimating the relationships among variables. Linear regression is an approach for modelling the relationship between a scalar dependent variable Y and one or more explanatory variables denoted X . The case of one explanatory variable is called simple linear regression. More than one variable is called multivariate. Logistic

Regression - In statistics, logistic regression, is a regression model where the dependent variable is categorical or binary.

Data Preprocessing This 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 sampling. □ Cleaning process is used for removal or fixing of some missing data there may be data that are incomplete. Sampling is the process where appropriate data are used which may reduce the running time for the algorithm. Using python, the preprocessing is done.

2.2 Functional Diagram of Proposed Work

It can be divided into 4 parts:

1. Descriptive analysis on the Data
2. Data treatment (Missing value and outlier fixing) □
3. Data Modelling □
4. Estimation of performance □

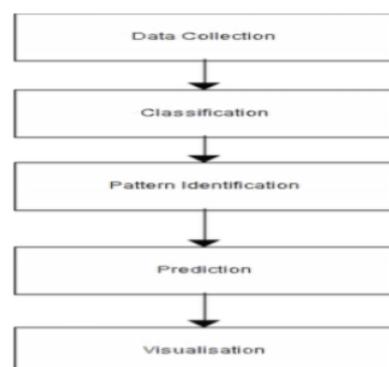


Figure 1- Architecture

3.IMPLEMENTATION

3.1. Data collection

Crime dataset from kaggle is used in CSV format.

3.2. 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.

3.3 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,

3.4. Building and Training 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 from sklearn. Building model is done using `model = model.fit(xtrain, ytrain)`.

3.5. 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)`.

3.6. Visualization Using `matplotlib` library from sklearn. Analysis of the crime dataset is done by plotting various graphs.

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 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 data cleaning and data transformation. The 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 ieeewic/acm international conference on (pp. 41-44). IEEE.

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.