

Crime Prediction and Analysis Using Machine Learning

Anuja. R¹, Remya. R. S²

Anuja. R¹, Department of CSE, Sivaji College of Engineering and Technology.

Remya. R. S², Assistant Professor, Department of CSE, Sivaji College of Engineering and Technology.

Abstract— Crime analysis and prediction is a systematic approach for identifying the crime. This system can predict region which have high probability for crime occurrences and visualize crime prone area. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. The extraction of new information is predicted using the existing datasets. Crimes are treacherous and common social problem faced worldwide. Crimes affect the quality of life, economic growth and reputation of nation. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. We propose a system which can analysis, detect, and predict various crime probability in given region. This paper explains various types of criminal analysis and crime prediction using several data mining techniques.

I. INTRODUCTION

1.1. Overview

Day by day crime data rate is increasing because the modern technologies and hitech methods are helps the criminals to achieving the illegal activities. According to Crime Record Bureau crimes like burglary, arson etc have been increased while crimes like murder, sex, abuse, gang rap etc have been increased. Crime data will be collected from various blogs, news and websites. The huge data is used as a record for creating a crime report database. The knowledge which is acquired from the data mining techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime. Data mining helps in solving the crimes faster and this technique gives good results when applied on crime dataset, the information obtained from the data mining techniques can help the police department. A particular approach has been found to be useful by the police, which is the identification of crime 'hot spots' which indicates areas with a high concentration of crime. Use of data mining techniques can produce important results from crime report datasets. The very step in study of crime is crime analysis. Crime analysis is exploring, inter relating and detecting relationship between the various crimes and

characteristics of the crime. This analysis helps in preparing statistics, queries and maps on demand. It also helps to see if a crime in a certain known pattern or a new pattern necessary. Crimes can be predicted as the criminal are active and operate in their comfort zones. Once successful they try to replicate the crime under similar circumstances. The occurrences of crime depended on several factors such as intelligence of criminals, security of a location, etc. The work has followed the steps that used in data analysis, in which the important phases are Data collection, data classification, pattern identification, prediction and visualization. The proposed framework uses different visualization techniques to show the trends of crimes and various ways that can predicts the crime using machine learning algorithm. 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

1.2. SCOPE OF THE PROJECT

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

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

1.3. OBJECTIVE OF PROJECT The objective of our work is to:

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

1.4.1. MACHINE LEARNING

Machine Learning combines computer science, mathematics, and statistics. Statistics is essential for drawing inferences from the data. Mathematics is useful for developing machine learning models and finally, computer

science is used for implementing algorithms. However, simply building models is not enough. You must also optimize and tune the model appropriately so that it provides you with accurate results. Optimization techniques involve tuning the hyper parameters to reach an optimum result.

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.

SUPERVISED LEARNING

In the majority of supervised learning applications, the ultimate goal is to develop a finely tuned predictor function $h(x)$ (sometimes called the “hypothesis”). “Learning” consists of using sophisticated mathematical algorithms to optimize this function so that, given input data x about a certain domain (say, square footage of a house), it will accurately predict some interesting value $h(x)$ (say, market price for said house).

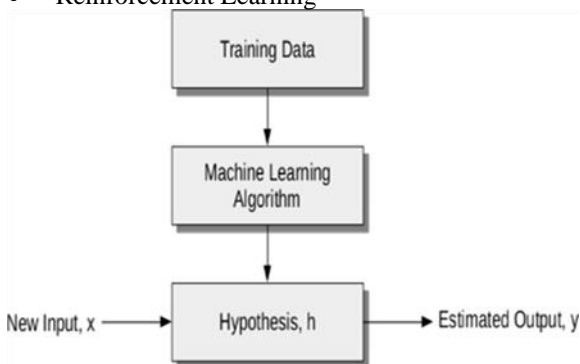
$$h(x_1, x_2, x_3, x_4) = \theta_0 + \theta_1 x_1 + \theta_2 x_3^2$$

This function takes input in four dimensions and has a variety of polynomial terms. Machine Learning model learns to adapt to new examples and produce better results.

Types of Machine Learning

Machine Learning Algorithms can be classified into 3 types as follows –

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning



Deriving a normal equation for this function is a significant challenge. Many modern machine learning problems take thousands or even millions of dimensions of data to build predictions using hundreds of coefficients. Predicting how an organism’s genome will be expressed, or what the climate will be like in

fifty years, are examples of such complex problems.

Under supervised ML, two major subcategories are:

- Regression machine learning systems: Systems where the value being predicted falls somewhere on a continuous spectrum.
- Classification machine learning systems: Systems where we seek a yes-or-no prediction.

In practice, x almost always represents multiple datapoints. So, for example, a housing price predictor might take not only square-footage (x_1) but also number of bedrooms (x_2), number of bathrooms (x_3), number of floors (x_4), year built (x_5), zip code (x_6), and so forth. Determining which inputs to use is an important part of ML design.

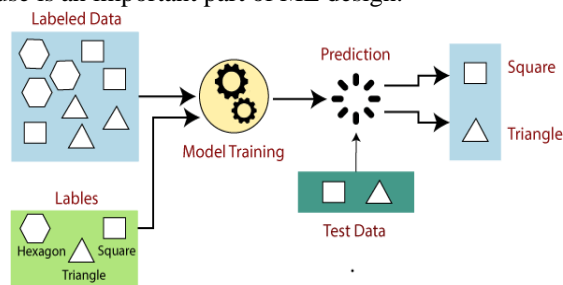


Fig – 1.2 Block Diagram

1.4.2. PROPOSED ALGORITHMS Decision Tree Classification Algorithm:

- Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
- A decision tree simply asks a question, and based on the answer (Yes/No), it further splits the tree into sub trees.

- Below diagram explains the general structure of a decision tree.

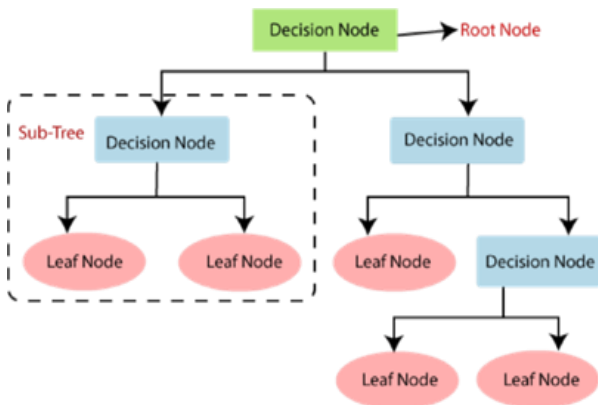


Fig-1.3 Structure of decision tree

There are various algorithms in Machinelearning, so choosing the best algorithm for the given dataset and problem is the main point to remember while creating a machine learning model. Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree- like structure.

2. SYSTEM ANALYSIS

2.1 PROPOSED SYSTEM

In this project, we will be using the technique of machine learning and data science for crime prediction of crime data set. The crime data is extracted from the official portal of 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 Logistic Regression classification and various other algorithms (Decision Tree and Random Forest) 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 whole purpose of this project is to give a just 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. This can be used in other states or countries depending upon the availability of the dataset.

2.3 SYSTEM ARCHITECTURE

There are many kinds of architecture diagrams, like a software architecture diagram, system architecture diagram, application architecture diagram, security architecture diagram, etc.

For system developers, they need system architecture diagrams to understand, clarify, and communicate ideas about the system structure and the user requirements that the system must support.

It describes the overall features of the software is concerned with defining the requirements and establishing the high level of the system. During architectural design, the various web pages and their interconnections are identified and designed. The major software components are identified and decomposed into processing modules and conceptual data structures and the interconnections among the modules are identified.

The following modules are identified in the proposed system.

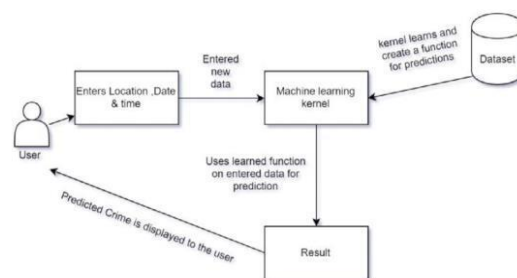


Fig- 3.1 Architecture diagram

The system architectural design is the design process for identifying the subsystems making up the system and framework for subsystem control and communication. The goal of the architectural design is to establish the overall structure of software system.

3.4 DATA FLOW DIAGRAM:

- The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
- The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
- DFD shows how the information moves through the system and how it is modified by a series of

transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

3.5 SYSTEM REQUIREMENTS HARDWARE REQUIREMENTS

- System : intel Core i5.
- Hard Disk : 1 TB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 8GB

SOFTWARE REQUIREMENTS

- Operating system: Windows 10.
- Coding Language: Python

3.5 SOFTWARE DESCRIPTION

Python is a free, open-source programming language. Therefore, all you have to do is install Python once, and you can start working with it. Not to mention that you can contribute own code to the community. Python is also a cross-platform compatible language. So, what does this mean? Well, you can install and run Python on several operating systems. Whether you have a Windows, Mac or Linux, you can rest assure that Python will work on all these operating systems. Python is also a great visualization tool. It provides libraries such as Matplotlib, seaborn and bokeh to create stunning visualizations.

In addition, Python is the most popular language for machine learning and deep learning. As a matter of fact, today, all top organizations are investing in Python to implement machine learning in the back-end.

APPENDIX

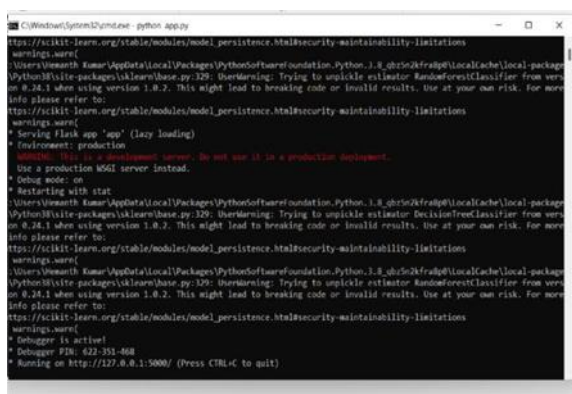


Figure-1 shows the page of running our application

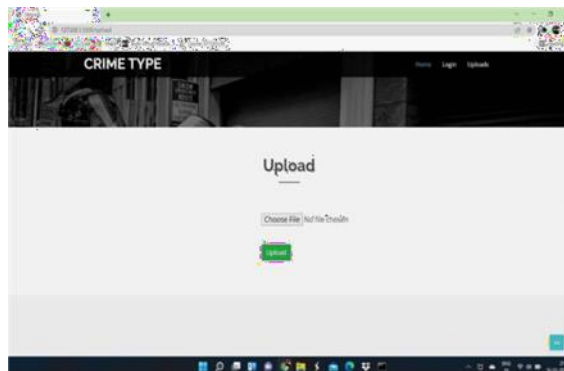


Figure-2 shows the page to upload the dataset of crime prediction

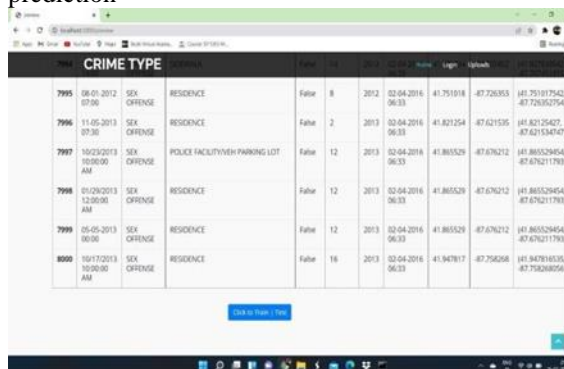


Figure-3 shows the dataset to train and test using machine learning algorithms.



Figure-4 shows the prediction of type of crime.

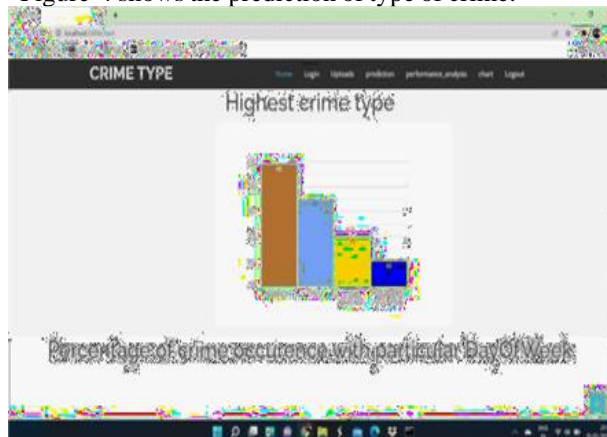


Figure-5 shows the performance analysis for the type of crime

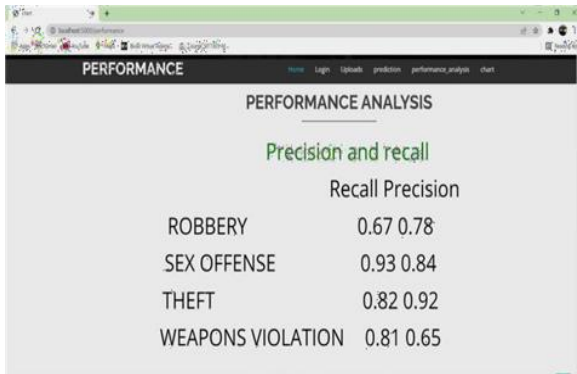


Figure-6 shows the confusion matrix for the analysis of crime.



Figure-7 shows the performance of crime occurrences with particular day of week.



Figure-8 shows the graph for percentage of crime occurrences with particular day of week.

5 CONCLUSION AND FUTURE WORK

5.4 Conclusion

In this paper focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police official in taking necessary steps to reduce crime. The project helps the crime analysis to analyze these crime networks by means of various interactive visualization. Future enhancement of this research work on training

bots to predict the crime prone areas by using machine learning techniques. Since, machine learning is similar to data mining advanced concept of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction.

5.5 Future Enhancement

Crime analysis takes past crime data to predict future crime locations and time. Crime prediction for future crime is a process that finds out crime rate change from one year to the next and projects those changes into the future. Crime predictions can be made through both qualitative and quantitative methods.

REFERENCES

- [1] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making, 2017.
- [2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.
- [3] Khushabu A. Bokde, Tisksha P. Kakade, Dnyaneshwari S. Tumasare, Chetan G. Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K- Means, International Journal of Engineering Research & technology (IJERT), 2018.
- [4] H. Benjamin Fredrick David and A. Suruliandi, Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.
- [5] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCS), 2015.
- [6] RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research, 2019.
- [7] Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science, 2015.
- [8] Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics, 2018.