International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Systametic Approach Towards the Description and Classification of Crime

Mrs. D Kanaka Satya[1], Yanamadala Renuka Suprathi[2], Thethala Ratna Reddy[3], Peruri Kodanda Sriram[4], Vasamsetti Chandu Priya Mahalakshmi[5], Pampana Bharat Raja[6]

[1]Assistant Professor, CSE Department, Vsm College Of Engineering, Ramachandrapuram, Dr.B.R.Ambhedkar Konaseema Dt, Andhra Pradesh - 533255
[2,3,4,5,6]Student, CSE Department, Vsm College Of Engineering, Ramachandrapuram, Dr.B.R.Ambhedkar Konaseema Dt, Andhra Pradesh - 533255

ABSTRACT
Understanding and addressing crime is a complex challenge that requires a structured and analytical framework. This project aims to develop a systematic approach to the classification and description of crime, facilitating more effective law enforcement, policy-making, and academic research. By categorizing crimes based on their severity, nature, legal framework, and societal impact, this study provides a comprehensive model that enhances clarity and consistency in crime analysis. The project explores various dimensions of crime, including violent, property, white-collar, cyber, environmental, and organized crime, and establishes criteria for their classification. Through this structured methodology, the project seeks to improve the understanding of criminal behavior patterns and contribute to the development of more targeted crime prevention strategies. The findings may also serve as a foundation for legal reforms and improved data collection practices within the criminal justice system.

INTRODUCTION

Day by day crime data rate is increasing because the modern technologies and hi- tech 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.

EXISTING SYSTEM

Data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression. De Bruin et. Al. introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. Manish Gupta et. Al. highlights the existing systems used by Indian police as e- governance initiatives and also proposes an interactive query based interface as crime analysis tool to assist police in their activities. He proposed interface which is used to extract useful information from the vast crime database maintained by National Crime Record Bureau (NCRB) and find crime hot spots using crime data mining techniques such as clustering etc. The effectiveness of the proposed interface has been illustrated on Indian crime records. Sutapat Thiprungsri examines the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. The purpose of his study is to examine the use of clustering technology to automate fraud filtering during an audit. He used cluster analysis to help auditors focus their efforts when evaluating group life insurance claims.

PROBLEM STATEMENT

With the rapid increase in crime rates and the growing complexity of criminal activities, law enforcement agencies face significant challenges in identifying patterns, predicting crime occurrences, and allocating resources effectively. Traditional methods of crime analysis are often manual, time-consuming, and lack the predictive power needed for proactive decision-making. This project aims to develop a machine learning-based crime classification and prediction system that can analyze historical crime data, identify patterns based on time and location, and accurately predict the type of crime likely to occur. By automating crime analysis and visualization, the system provides a data-driven solution to support faster, more informed responses to criminal activity.

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.

IMPLEMENTED ALGORITHMS

a) LOGISTIC REGRESSION

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

RANDOMFOREST ALGORITM

Random forest algorithm can use both for classification and the regression kind of problems. In this you are going to learn, how the random forest algorithm works in machine learning for the classification task. Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

c) DECISION TREE ALGORITM

When it comes to sequence prediction problems, recurrent neural networks (RNNs) of the Long Short-Term Memory (LSTM) type can identify order dependence. This is most frequently used in challenging problems like machine translation, speech recognition, and others. This problem was noted when training traditional RNNs because little to no training can take place if the gradients are very small or zero as we proceed through the neural network, which leads to poor prediction performance. The classification, processing, and prediction of time series data are ideally suited for LSTM networks since there may be lags of unknown length between important events in a time series. Because LSTM circumvents the RNN's short-term memory limitations, it is far more effective and superior to the standard RNN. The LSTM can ignore irrelevant information while processing inputs and relevant information.

WORKING OF MACHINE LEARNING ALGORITHMS

Machine Learning is the most popular technique of predicting the future or classifying information to help people in making necessary decisions. Machine Learning algorithms are trained over instances or examples through which they learn from past experiences and also analyze the historical data. Therefore, as it trains over the examples, again and again, it is able to identify patterns in order to make predictions about the future. Data is the core backbone of machine learning algorithms. With the help of the historical data, we are able to create more data by training these machine learning algorithms. For example, Generative Adversarial Networks are an advanced concept of Machine Learning that learns from the historical images through which they are capable of generating more images. This is also applied towards

speech and text synthesis. Therefore, Machine Learning has opened up a vast potential for data science applications.

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. 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

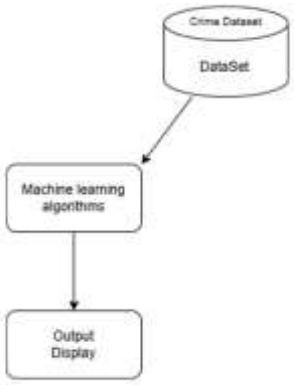


Fig:2 System Architecture Diagram

1. Algorithm Implementation:

The algorithm implementation in the crime classification system involves applying supervised machine learning techniques to historical crime data. After preprocessing the dataset—removing null values and converting categorical data into numerical form using Label Encoding—relevant features such as latitude, longitude, hour, and crime type are selected. The dataset is then split into training and testing sets, and models like Random Forest, Decision Tree, and Logistic Regression are trained using the training data. The Random Forest Classifier, known for its accuracy and ability to handle overfitting, is commonly selected due to its ensemble approach of combining multiple decision trees. The trained model is then used to predict crime types on unseen data, and its performance is evaluated using accuracy score, confusion matrix, and classification reports.

2. Feature Extraction:

Feature extraction in the crime classification system involves identifying and transforming key attributes from the raw dataset into meaningful inputs for the machine learning model. Important features such as date, time, location, and crime type are extracted and processed. The date is split into components like month and day of the week, while the time is converted to hour to understand crime patterns based on time

of occurrence. Geographical coordinates like latitude and longitude are used to identify location-based trends. Categorical features such as crime type, victim's gender, and premise description are encoded into numeric values using Label Encoding. These extracted features help the model detect patterns and improve the accuracy of crime prediction**.**

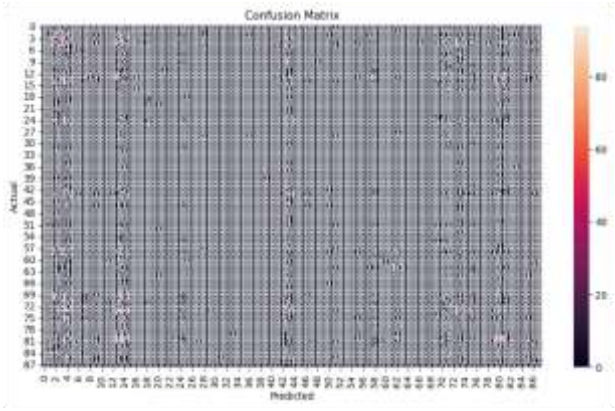
3. Caption Generation:

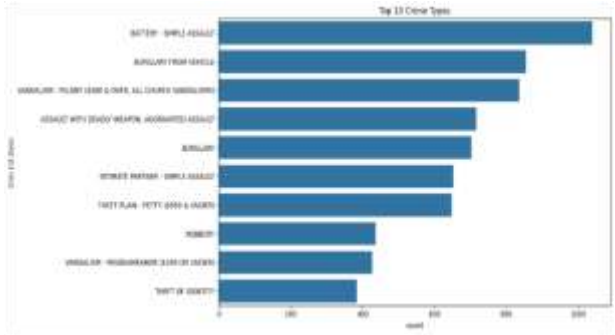
Caption generation in the crime classification system involves creating short, descriptive texts based on the input data and predicted outcomes from the machine learning model. After processing features like location (latitude and longitude), time (hour, day, month), and crime type, the system generates natural language captions summarizing the prediction. For example, if the model predicts a robbery at a specific time and place, the system might generate a caption such as "Robbery likely to occur near downtown at 9 PM on Friday." These captions help in quickly communicating insights, especially in visualizations, alerts, and dashboards, making the system's output more user-friendly and actionable.

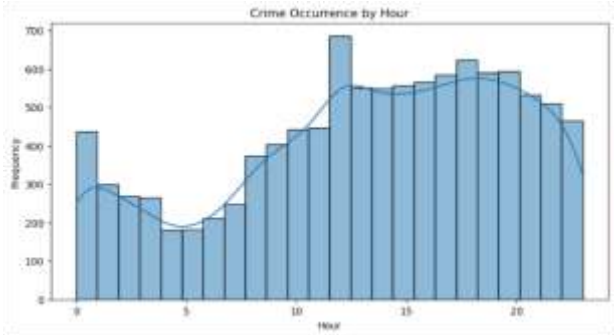
RESULT ANALYSIS

Classification Report:

	precision	recall	f1-score	support
ARSON	0.00	0.00	0.00	7
ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER	0.12	0.17	0.14	6
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	0.19	0.25	0.21	138
ATTEMPTED ROBBERY	0.00	0.00	0.00	15
BATTERY - SIMPLE ASSAULT	0.16	0.31	0.21	197
BATTERY ON A FIREFIGHTER	0.00	0.00	0.00	0
BATTERY POLICE (SIMPLE)	0.33	0.12	0.18	8
BATTERY WITH SEXUAL CONTACT	0.00	0.00	0.00	8
BIKE - STOLEN	0.22	0.13	0.17	30
BOMB SCARE	0.00	0.00	0.00	1
BRANDISH WEAPON	0.00	0.00	0.00	36
BUNCO, ATTEMPT	0.00	0.00	0.00	1
BUNCO, GRAND THEFT	0.00	0.00	0.00	18
BUNCO, PETTY THEFT	0.00	0.00	0.00	6
BURGLARY	0.23	0.32	0.27	139
BURGLARY FROM VEHICLE	0.34	0.49	0.40	186
BURGLARY FROM VEHICLE, ATTEMPTED	0.00	0.00	0.00	2
BURGLARY, ATTEMPTED	0.20	0.09	0.12	11
CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT	0.00	0.00	0.00	0
CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	0.41	0.75	0.53	12
CHILD ANNOYING (17YRS & UNDER)	0.00	0.00	0.00	4
CHILD NEGLECT (SEE 300 W.I.C.)	0.25	0.17	0.20	6
CHILD PORNOGRAPHY	0.00	0.00	0.00	0
CHILD STEALING	0.00	0.00	0.00	1
CONTEMPT OF COURT	0.00	0.00	0.00	4
COUNTERFEIT	0.00	0.00	0.00	0







Crime Hotspots

CONCLUSION

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 analysis 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.

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.

4. Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.
5. RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research,2019.
6. Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science , 2015.
7. Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.