

CRIME DATA ANALYSIS AND PREDICTION USING MACHINE LEARNING

K. Santoshi Rupa¹, N. Niha², M. Jey Keerthana³, N. Manasa Sammeta⁴, M. Santhi⁵

Department of Computer Science & Engineering, Vignan's Institute of Engineering
For Women, Duvvada, Visakhapatnam, Andhra Pradesh, India.

ABSTRACT

The crime data analysis and prediction project aims to explore patterns and trends in crime data and develop a model to predict the type of crime based on certain factors. The project involves data collection, cleaning, and preparation, followed by exploratory data analysis to identify patterns in the data. Machine learning algorithms such as K-Nearest Neighbours, Gaussian Naive Bayes, and Random Forest are utilized to develop a predictive model for crime type based on factors such as location, timing, and other relevant variables. The KNN algorithm, out of the three, predicts outcomes with an accuracy of 87%. The project also involves the use of data visualization techniques to present insights and findings from the analysis. The objective of this project is that both the user, who is unfamiliar with the city, and the police will benefit from its assistance in assessing crime in the area.

Keywords: Machine learning, prediction, crime data analysis, KNN, Gaussian Naive Bayes, and Random Forest

INTRODUCTION

Crime is a concept that has existed for a very long time and has evolved as society has developed. Nowadays, nobody is secure. As more crimes are committed, there are an increasing number of criminal cases waiting in India. An extensive internal investigation and analysis should be done before making a decision about how to handle a case based on specific evidence. It is challenging for the authorities to analyse and decide on these criminal cases given the volume of crime data that is currently available in this country. By following the recognition of this significant problem, the research is concentrated on developing a remedy for the choices that cause crime. Science's field of machine learning focuses on automated The data pre-processing involves data cleaning, feature selection, dropping null values, etc. After data pre-processing, the data is free of null values, which may alter the accuracy of the model significantly, and feature selection is used to select only the required features that won't affect the accuracy of the model. The fact that crimes are rising at an alarming rate makes it difficult to foresee them effectively. So, it is important to determine potential crimes now in order to prevent them in the future. In recent years, numerous academics have tested various machine learning techniques and specific inputs to predict crimes. As science and technology grow, criminals are using scientific methods to commit crimes, and as society develops, authorities have been shocked by the increased knowledge of criminals. Family disputes over property distribution are the most frequent motive for many crimes committed with the intent of attaining wealth in money or property. The prediction and analysis of the crime can be enhanced with the use of machine learning. It is asserted to be a pervasive social problem that affects both the economic development and quality of life in a society. It is emphasized as a significant problem that directly affects a person's quality of life as well as having indirect societal repercussions.

LITERATURE SURVEY

Arunmozhi and Vijayalakshmi's "Crime Type Using Machine Learning Techniques": In this study, decision trees and k-nearest neighbour algorithms are used to discuss the usage of the machine learning approaches for crime type and prediction. Using crime datatype from various cities, this research analyses the performance of a different machine learning techniques, such as decision trees, random forests, and support vector machines, for crime prediction. In order to anticipate crime in a specific location, this research

suggests a hybrid strategy that blends machine learning methods with geospatial data. To identify criminal hotspots, the authors employ the K-means clustering technique and decision trees.

By Swathi J. and Shilpa. B., "Crime Prediction and Analysis Using Machine Learning": In order to assess crime data and forecast the likelihood of future crimes, this study provides a crime prediction system that makes use of machine learning algorithms like decision trees, random forests, and Naive Bayes. Predictive policing uses machine learning algorithms, which are discussed in this paper, to predict and prevent crime. The authors discussed about the many kinds of data used for crime prediction, including data from social media, data from sensors, and data from crime reports. In order to predict criminal recidivism, this paper conducts a thorough literature assessment on the topic. The writers go over the many categories of data utilized for forecasting, including demographic information, criminal records, and psychiatric evaluations.

Hamed Zarei and Mehran Abolhasan: "Predicting Crime Type in Smart Cities Using Machine Learning," Using the machine learning methods like decision trees and support vector machines, this article suggests a framework for smart cities that can anticipate criminal type. Social media, security cameras, and crime records are just a few of the sources the authors draw their information from. This is survey of the machine learning methods for predicting crime type. This study introduces a crime prediction system that analyses crime data type from social media and forecasts future crimes using the machine learning algorithms and natural language processing methods. The authors demonstrate that their approach predicts crime types with a high degree of accuracy.

Salwa Ismail, Ruaa Shabbar, and Mohamed Elmogy's "Crime Prediction Using Machine Learning Methods" This work provides a crime type and prediction system that analyses crime data from several sources, such as social media, security cameras, and crime reports, using machine learning approaches, such as decision trees and random forests. This study suggests a crime prediction system that analyses crime data and forecasts future crimes using time series analysis and machine learning techniques like decision trees and support vector machines. The authors demonstrate that their approach has a high level of crime prediction accuracy. A method for predicting crimes that uses crime data and makes predictions about upcoming crimes using machine learning algorithms. The authors demonstrate how their approach offers secure and open crime type prediction.

Seyed Aliakbar Kiaei, Seyed Vahid Mirisae, and Amir Hassan Monadjemi's paper "Crime Analysis and Prediction Using Machine Learning and Social Media Data": The crime prediction system described in this study analyses crime trends and forecasts future crimes using the machine learning algorithms and social media data. The authors demonstrate how their approach works better than established crime prediction models. A review of the decision trees, logistic regression, and artificial neural networks utilized in the machine learning for crime type prediction. For appropriate crime type prediction, the authors also cover the significance of data preprocessing, feature selection, and model validation. This study suggests a system for predicting crimes that analyses several types of crime data and does so by using machine learning. In order to examine crime data and forecast future crimes, this study proposes a crime prediction system that makes use of social network analysis and machine learning. This study suggests a machine learning method for classifying different sorts of crimes by analysing criminal descriptions.

PROPOSED SYSTEM

The proposed system goal would be to develop a prediction model. The training data set would be used for training, and the test dataset would be used for validation. The system processes a given crime dataset and provides insights into the dataset by performing the following operations:

Data Pre-processing: The system pre-processes the given data set to extract the crime date time and crime day features. Data Summary: The system provides a summary of the given dataset, which includes data set information, the first few rows of the dataset, the number of null values in each column, and descriptive statistics for the dataset.

Data Day-wise Plotting: The system plots the crime rate of each type of crime on the day of the week. It also generates a dictionary with the crime rate of each crime on each day of the week.

Classifier Model: The system builds a classifier model based on the given dataset to predict whether a given

crime is of a particular type or not. The system uses KNN, Naive Bayes, and Random Forest classifiers and reports their accuracy using k-fold cross-validation .Overall, the proposed system takes in a crime dataset and provides insights into the dataset to help in understanding the trends and patterns in crime. It also provides a classifier model that can be used to predict the type of crime based on given features.

METHODOLOGY

The crime data analysis and prediction project employs the following machine learning algorithms, namely K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Random Forest, to build a predictive model for crime type prediction.

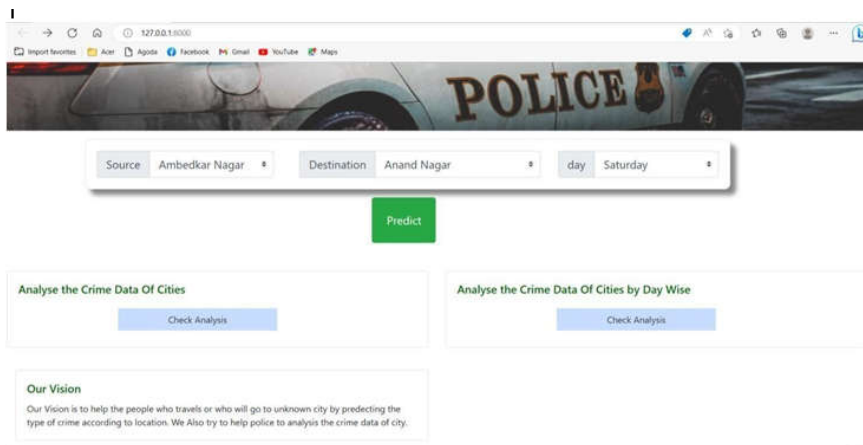
The KNN algorithm is a simple and effective algorithm that classifies an unknown data point based on its closest k neighbors. In this project, KNN is used to classify crime types based on geographical location, time, and other relevant variables. KNN works by calculating the distance between the test data point and all the training data points, and then selecting the k nearest neighbors. The majority class of the k nearest neighbours then determines the class of the unknown data point. The Gaussian Naive Bayes method is a probabilistic algorithm that selects the class with the highest probability after calculating the probability of each class given the input data. In this study, GNB is used to categorize the different types of crimes according to their characteristics. According to GNB, each feature is thought to be independent and to follow a Gaussian distribution. To determine the probability of the class given the input features, it first calculates the conditional probability of each feature given the class and multiplies those results together.

Random Forest algorithm builds many decision trees and aggregates their output to provide a final prediction. It is an ensemble learning algorithm. Random Forest is employed in this project to categorize crime types based on all of the available features.

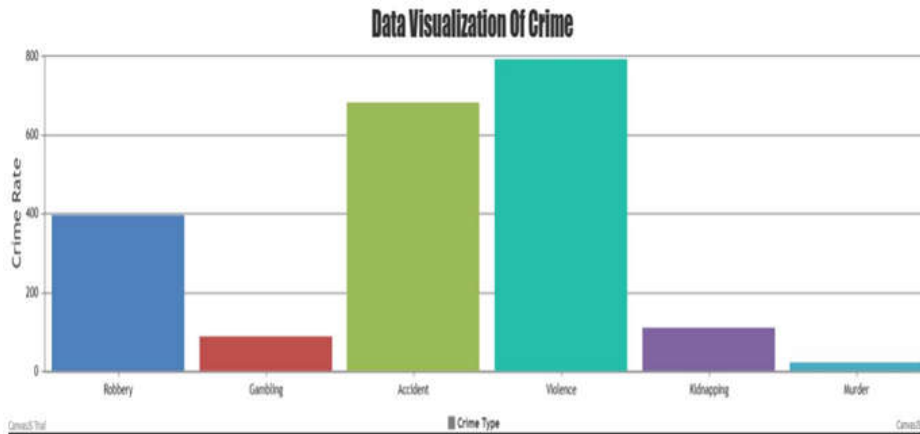
RESULTS



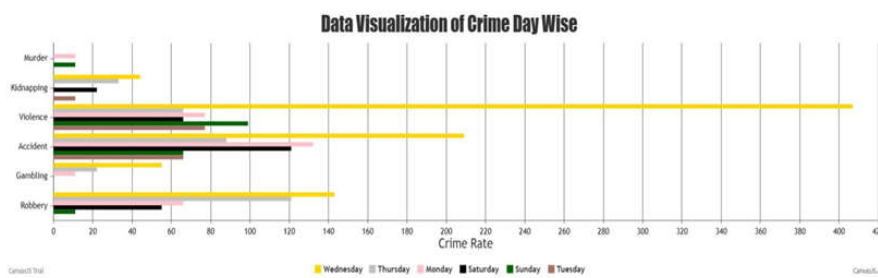
HOME PAGE WITH INPUT



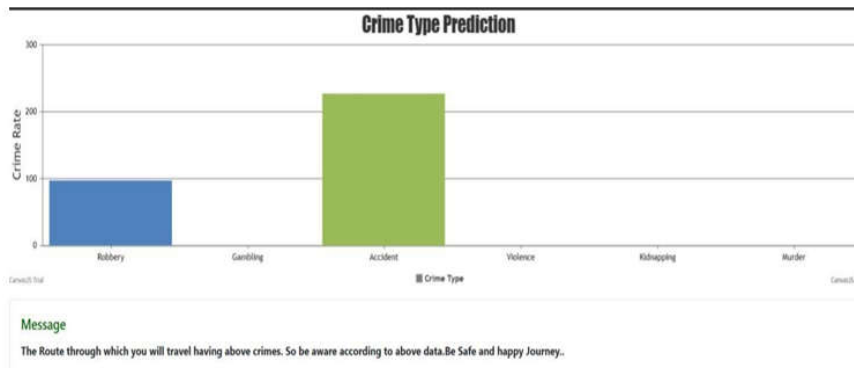
CRIME DATA VISUALIZATION



DAY WISE DATA VISUALIZATION



CRIME PREDICTION GRAPH



CONCLUSION

In conclusion, the crime data analysis and prediction using machine learning project provides insights into patterns and trends in crime data and develops a model to predict the type of crime based on certain factors. The project employs various machine learning algorithms such as K-Nearest Neighbors, Gaussian Naïve Bayes, and Random Forest to build a predictive model that can assist law enforcement and criminal justice officials in predicting the type of crime based on location, time, and other relevant variables. The project's findings reveal that certain factors such as location, time of day and a day of the week have a significant impact on the type of crime that occurs. Moreover, the predictive model developed through this project offers a promising tool for crime prevention and intervention, by helping law enforcement and criminal justice officials allocate resources effectively and take proactive measures to prevent crime.

FUTURE SCOPE

The system is designed to work with historical crime data. However, integrating it with realtime data could help law enforcement agencies to respond more quickly to crime incidents. The current system uses a variety

of machine learning algorithms, but there are many more advanced algorithms that could be used to improve accuracy and performance. The user interface could be enhanced to make it more intuitive and user-friendly. This could include better visualizations and more interactive features. The system could be integrated with other law enforcement systems, such as dispatch systems and crime reporting systems. This would allow for more seamless data sharing and communication between different agencies. Overall, these enhancements could improve the accuracy, speed, and usability of the system, helping law enforcement agencies to prevent and solve crimes more effectively and make people feel much safe.

REFERENCES

1. Arun mozhi and Vijaya lakshmi's" Crime Type Using Machine Learning Techniques"
2. BySwathiJ.andShilpa.B.,"CrimePredictionandAnalysisUsingMachineLearning"
3. HamedZareiandMehranAbolhasan:"PredictingCrimeTypeinSmartCitiesUsingMachineLearning,"
4. SalwaIsmail,RuaaShabbar,andMohamedElmoggy's"CrimePredictionUsingMachineLearningMethods"
5. SeyedAliakbarKiaei,SeyedVahidMirisae, andAmirHassanMonadjemi'spaper"CrimeAnalysisandPredictionUsingMachineLearningandSocialMediaData"
6. OltaLaha,“CrimeAnalysisandPredictionUsingMachineLearning”,IEEE2020
7. Suhong Kim, Param Joshi, Parminder Singh Kalsi,Pooya Taheri, “Crime Analysis is Through Machine Learning”
8. “Crime Prediction and Analysis” by Kumari, Pratibha & Gahalot, Akanksha & Uprant, & Dhiman, Suraina & Chouhan, Lokesh. (2020).
9. Akash Kumar, Aniket Verma, Gandhali Shinde, Yash Sukhdeve, NidhiLal,“Crime Prediction Using K-Nearest Neighboring Algorithm”.