



CRIME RATE PREDICTION USING MACHINE LEARNING

1*Mrs.Bindu K P, 2*Mohammed Yaseen, 3*Nagendra B N, 4*Nitesh A, 5*Prajwal R

*1Assistant Professor, Department Of CSE, K S School of Engineering and Management, Bengaluru , Karnataka, India.

*2345 Department Of CSE, K S School of Engineering and Management, Bengaluru ,Karnataka, India.

Abstract : The real world of crime analyzing and the investigative process, especially in developing nations like India, the endeavor is often time-intensive. This research endeavors to explore the methodology that can expedite the anticipation of crime types in certain area. Crime, a pervasive social issue impacting individuals worldwide, has experienced a significant surge in years, with a projected in the future. This study aims to furnish law officers and the public with effective crime forecasting tools to optimize resource allocation to prospective crime hotspots. To this end, the previous research work proposes the utilization of a machine algorithm for the detection of crimes cases on collected data, revealing the extent of criminal activities in specific areas. Given the existing gaps in prior studies concerning the crime prediction through learning models, this study has various algorithms, including K-nearest neighbors , decision, support machine and random forest. The algorithm demonstrating the highest accuracy is selected for evaluating the results.

IndexTerms - Crime rate, Random Forest, Crime prediction, Machine Learning, Analysis and Forecast.

I. INTRODUCTION

The Crime analyzing is a way to identify crime. This method can predict areas or cities with more crime rates and crime areas. Crime often affects a country's quality of life and economic growth. To protect people from crime, advanced techniques and new methods are needed to improve crime detection in order to protect society. This system to detect and predict various crimes in a particular area. Crime, as a social threat and illegal behavior, offers many opportunities to people. The crimes limited to assault, murder, theft, robbery and drug crimes. The crimes of the first category, crimes include violation of tourist rights, damage to living spaces and citizens, and other crimes.

Crime analyzing is an aspect of crime that requires analyzing of crime patterns to understand the relationship between crime and the situation in an area. These issues have inspired many studies in recent years on predicting future occurrence crime to help police allocate resources. Machine learning technique's is used to get the information from this large database and discover new, never-before-seen connections. The information is then used to note and investigate crime, which can help the crime analyst investigate these crimes using a variety of interactive methods, thus helping prevent crime. Various kinds of learning techs were applied to predict crime based on data from the last 15 years in Vancouver, and K nearest neighbor algorithm and P-based augmented decision algorithm achieved 39% and 44% accuracy respectively. Kumar et al. Predicting the average probability of many types of crime at different places and times in the city using additive classification trees, K-Neighbour, Support machine, decision trees classification and k-nearest neighbor algorithm.

The data gives the security situation in the country, according to data given by the Crime Bureau of India, which shows that crime is increasing every year. Crime today can take in different types , including crimes such as theft and vandalism, emerging problems such as digital crime. These figures are a reminder of strengthening counter-terrorism strategies, law enforcement and investigate in urine technology to prevent a rise in crime in India. By this problem does not the guarantee the safety and health of the public. The speed at which new solutions are implemented to keep up with the changing crime scene is also important.

According to data released by the Indian National Crime Records Bureau, the information presents an image of the country security situation indicating a rise, in crime rates every year. This statistical pattern highlights the increasing difficulty faced by law enforcement agencies. Emphasizes the need for measures and strategic interventions to tackle the surge in criminal activities. Crimes today come in forms, including offenses like theft and assault as well as emerging challenges such as digital crimes, which necessitate a comprehensive approach. The data serves as a reminder of the importance of enhancing crime prevention strategies strengthening law enforcement capabilities and investing in technologies to effectively combat the growing number of criminal incidents, in India. Addressing this issue not ensures the safety and well being of citizens. Also underscores the urgency of adopting innovative solutions to stay ahead of evolving criminal methods.

II. LITERATURE SURVEY

A. "The Crime Detection Technique by Using data Mining and K-Means"

A high rate of social crime can have an effective impact on companies and organizations. This study on the variety of crime groups according to their occurrence within a year. We drew on theoretical models of real crime data collected by the officers of England country and Wales of the years 1990 to 2011, using exploratory mining techniques for analysis, exploration and pattern discovery. We focused on working to increase the quality of our advertisements and emphasized their importance. Eliminates products of low importance. Using the Rapid Miner tool, a genetic algorithm was used to utilize the operator detection.

This model includes a limited number of data breaches and cannot be used for multiple data breaches. This project focuses on crime analysis using crime analysis techniques on crime data using speed mining tools, here we do crime analysis by taking into account murders and organizing them according to years, the result is that murders saw a decline from 1991 to 2011. Reports says that the output of the joint effort, crimes committed over the years are easily detected and used to develop preventive measures in the future.

B. "Crime Rate Using K-Nearest Neighboring Algorithm"

The model is identification of crimes by using classification techniques and crime can be done using data set of six cities of Tamil Nadu by using KNN classification, K Means clustering, Agglomeration hierarchical clustering, whose main aim was to use a datasets where the data positions were divided into the different classes to get clarity of a new sample positions. Using features like Day, Date, Year of the using KNN - algorithm it is found to be 40% accurate. Their model used techniques like Logistic-regression, Decision trees, Bayesian Methods and Support Vector Machine. There were many trails of different methods were used on the training datasets by splitting it into two sets; training and testing, both validation and cross-testing were conducted, the method with the lowest loss was applied to get the results for the test data. This research work offers a way to foresee and predict crimes and frauds within a city.

It focuses on having a crime prediction tool that can be helpful to law enforcement. This paper is aimed at increasing the prediction accuracy as much as possible. As compared to the others work, this work was successful in achieving the highest accuracy in prediction. The KNN system helps law implementing agencies for improved and exact crime analysis.

By traversing through the crime datasets, we have to find out different reason that lead to crime. Since this paper is bearing in mind only some limited factors, full accuracy cannot be accomplished. This system has trained using certain attributes but we can take account of more factors to improve accuracy. In the future, this work can be stretched to have developed classification algorithms to detect criminals more efficiently. The crime rates that are increasing non-stop may go down in the future due to such prediction techniques.

C. "A Study on Crime Examination and Forecasting"

This research intends to predict criminal behavior by making use of the myriad of characteristics included in the datasets. It is possible to apply "machine learning algorithms", the primary language of which is Python, to make predictions on the type of criminal activity that are take place in a specific location. The goal would be to train a model such that it could make predictions. The training dataset will be checked with test dataset. Depending up on the level of precision required, a more suitable method will be used to generate the model. The prediction of criminal activity will have the k nearest-neighbor (KNN) classification, other techniques. Visual and graphical representations of the datasets are provided for the aim of conducting an investigation into possible offenses that were committed within the nation.

Crime prediction system using various techniques as machine learning, deep learning technologies, specifically employing "the k-Nearest Neighbour" classifier and support machine algorithms. The models analysed show promising accuracy and effectiveness, aiding law enforcement in resource allocation and decision-making. By analysing crime datasets, visualizing data through graphs and charts, and comparing various algorithms to optimize accuracy, the project demonstrates the potential for the machine learning in the field of crime prevention. Importantly, the system are adaptable to various regions and countries, provided that relevant datasets are available, highlighting its potential for widespread implementation and significant impact on global crime reduction efforts. However, addressing ethical, privacy, and fairness concerns is crucial to avoid biases and discrimination.

D. "Analysis for Crime Prediction and Forecasting Using ML and DL Techniques"

Overcoming critical challenges arising from crimes and criminal offenses before justice requires effective management. Improving urban safety with computational tools on accurate of the crimes and prediction models. There are still gaps in the use of best predictive techniques that can guide well about the law enforcement in tracking crime. This research aims to fill this gap using different machine algorithms, including support vector, naive Bases, k-nearest, tree decision trees, and multilayer perceptron (MLP), random forest. and increasing cloud slope. Also perform time analysis using Long and Short Term Memory and Overaggressive Integrated Moving Average

models to better fit crime data. This comprehensive program uses machine learning and real-time analysis to prediction models. More importantly, the entire system is not tied to a specific time when a crime occurs, making it more effective and less time-consuming.

III. DATA AND PREPROCESSING

The data utilized in this study originates from the National Crime Records Bureau (NCRB) encompassing records from 19 metropolitan areas across India. These datasets are crucial for achieving accurate crime prediction models. Accessible via the NCRB's open data portals, they are downloadable. Given the diversity and scale of the dataset, it offers comprehensive insights into crime patterns across these metropolitan areas. Through careful preprocessing and partitioning, the dataset was prepared for training and testing machine learning models aimed at crime prediction and analysis. The dataset obtained from the National Crime Records Bureau (NCRB) of 19 metropolitan areas in India presents a wealth of information crucial for understanding and addressing criminal activities across urban regions. Comprising data on 10 different types of crimes, this dataset offers a comprehensive view of crime trends, patterns, and dynamics. However, before this data can be effectively utilized for modeling and analysis, it must undergo thorough preprocessing to ensure its quality, consistency, and usability.

Data preprocessing involves several steps aimed at cleaning, transforming, and organizing the raw dataset into a format suitable for analysis. One of the primary tasks in preprocessing is handling missing values, outliers, and inconsistencies that may exist within the data. This often requires careful examination and imputation of missing values using appropriate techniques to maintain the integrity of the dataset. Additionally, outliers, which can skew the analysis and model performance, need to be identified and either corrected or removed. Furthermore, data preprocessing includes feature engineering, where new features may be created or existing features may be transformed to enhance the predictive power of the models. This step may involve scaling, normalization, or encoding categorical variables to ensure that the data is appropriately represented for modeling purposes. Moreover, data partitioning is essential to separate the dataset into training and testing subsets, with a typical split of 80% for training and 20% for testing. This ensures that the model is trained on a sufficient amount of data while also allowing for unbiased evaluation of its performance on unseen data.

Overall, data preprocessing plays a critical role in ensuring the quality and reliability of the dataset, laying the foundation for the development of accurate and effective predictive models. By carefully cleaning, transforming, and partitioning the data, researchers can extract meaningful insights and patterns that contribute to a deeper understanding of crime dynamics and support informed decision-making by law enforcement agencies and policymakers.

IV. PROPOSED SYSTEM

The proposed system for crime rate prediction aims to leverage the dataset obtained from the National Crime Records Bureau (NCRB) of 19 metropolitan areas in India, which encompasses records of 10 different types of crimes. This dataset will be divided into 20% for testing the model and 80% for training purposes, ensuring an unbiased evaluation of the predictive algorithms. The system will employ various machine learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Random Forest, to develop predictive models capable of forecasting crime rates across different regions and crime types.

Support Vector Machine (SVM) is a powerful supervised learning algorithm that is effective for classification and regression tasks. It works by finding the hyperplane that best separates the data into different classes or categories, maximizing the margin between the classes. In the context of crime rate prediction, SVM can be trained on historical crime data to identify patterns and trends, allowing it to make accurate predictions.

K-Nearest Neighbors (KNN) is a simple yet effective algorithm used for both classification and regression tasks. It operates by finding the K nearest data points to a given point and making predictions based on the majority class or average value of the neighboring points. In the context of crime rate prediction, KNN can be trained on historical crime data to identify similar neighborhoods or regions with comparable crime patterns, enabling it to predict crime rates for new locations.

Decision Tree is a versatile algorithm that is widely used for classification and regression tasks. It works by recursively partitioning the data into subsets based on the values of different features, creating a tree-like structure where each internal node represents a decision based on a feature, and each leaf node represents the predicted outcome. In the context of crime rate prediction, Decision Tree can be trained on historical crime data to identify important features and their interactions, facilitating accurate predictions of crime rates.

Random Forest is an ensemble learning algorithm that combines multiple Decision Trees to improve predictive performance. It works by training a collection of Decision Trees on random subsets of the data and aggregating their predictions to make a final prediction. In the context of crime rate prediction, Random Forest can effectively capture complex relationships and interactions between different features, leading to more accurate and robust predictions of crime rates.

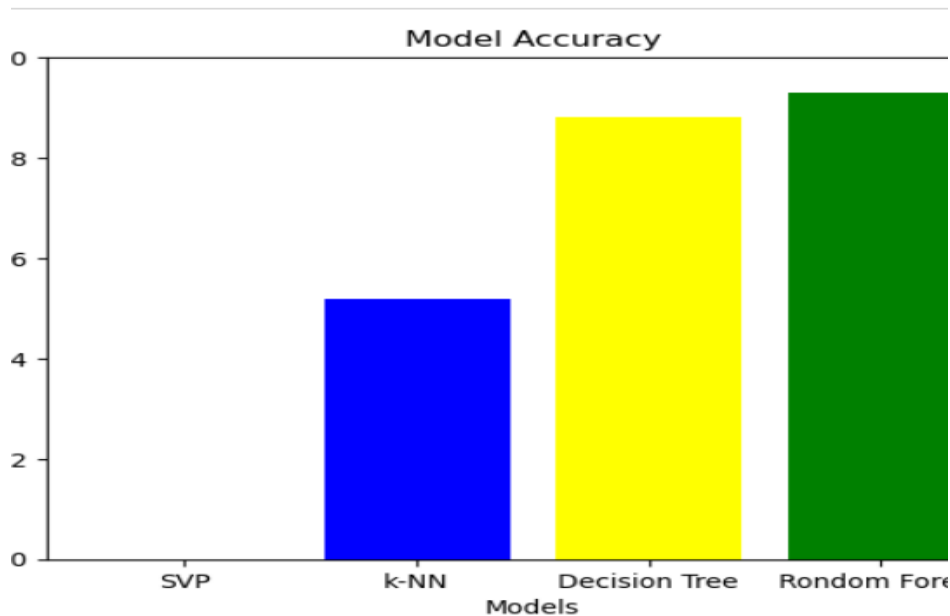


Fig - Models Accuracy Comparison.

Support Vector Machine (SVM), known for its effectiveness in handling high-dimensional data, is utilized as one of the predictive models. K-Nearest Neighbors (KNN), which relies on the similarity of data points in feature space, is also employed to provide an alternative approach to crime rate prediction. Additionally, Decision Tree, a simple yet powerful algorithm for classification tasks, is incorporated into the system. Finally, Random Forest, an ensemble learning technique that aggregates multiple decision trees, is employed to further enhance prediction accuracy.

Upon training the models on the designated training dataset, their performance is evaluated using the testing subset. The accuracy scores obtained from the evaluation demonstrate the effectiveness of each algorithm in predicting crime rates. Specifically, the accuracy scores achieved by each algorithm are as follows: SVM with an accuracy score of 0.17, KNN with 0.55, Decision Tree with 0.88, and Random Forest with 0.93.

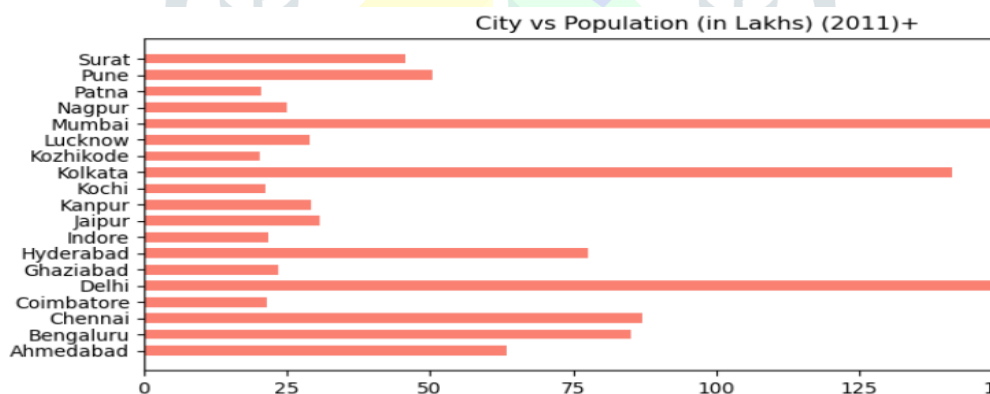


Fig - City vs Population

To calculate the crime rate, we divide the total number of crimes by the total population of the city, expressed in lakhs (100,000). The formula for calculating the crime rate is:

$$\text{Crime Rate} = \frac{\text{Total Population of the City (in lakhs)}}{\text{Number of Crimes}}$$

For example, if a city has a total population of 5 lakhs and there were 20 reported crimes, the crime rate would be calculated as follows:

$$\text{Crime Rate} = \frac{2000}{5} = 4$$

This means that there are 40 crimes reported per lakh (100,000) of the city's population. It provides a standardized measure to assess the prevalence of crime in relation to the population size of the city, allowing for meaningful comparisons between different cities or regions.

By utilizing these machine learning algorithms and the rich dataset provided by the NCRB, the proposed system for crime rate prediction aims to provide law enforcement agencies and policymakers with valuable insights and forecasts to support proactive crime prevention strategies and resource allocation efforts, ultimately contributing to the enhancement of public safety and security in metropolitan areas across India.

The proposed crime rate prediction system incorporates a user-friendly web interface to facilitate seamless interaction between users and the predictive models. This web interface serves as a central platform where users, such as law enforcement agencies, policymakers, and researchers, can access and utilize the predictive capabilities of the system.

One of the primary features of the web interface is the ability for users to input parameters and customize their crime rate predictions. Users can specify the geographic region of interest, select the types of crimes to be predicted, and adjust other relevant parameters to tailor the predictions to their specific needs. This interactive functionality empowers users to explore different scenarios and assess the potential impact of various interventions and policy changes.

Fig - Prediction Page

Overall, the proposed web interface serves as a powerful tool for users to interact with the crime rate prediction system, enabling them to harness the predictive capabilities of machine learning algorithms and make data-driven decisions to enhance public safety and security in metropolitan areas across India.

V. CONCLUSION

The primary objective of this study is to estimate crime rates in various locations, with a specific focus on factors such as population density, country, crime rate, and centrality. The chosen destination for this investigation is India, and a Random Forest model has been employed to forecast crime rates. The study presents a graph following the implementation of the Random Forest model. Recognizing that crime is a significant global issue impacting communities and societies worldwide, the research emphasizes the need to address and manage this challenge effectively. The study acknowledges the complexity of predicting crime and extracting meaningful insights from extensive crime datasets. The potential reduction of crime through advanced forecasting is highlighted, emphasizing the ongoing efforts to enhance prediction systems through data gathering and mining technologies. The survey aims to improve crime predictions by identifying trends and patterns, anticipating the types of violations likely to occur in specific districts during specific time periods and seasons. The practical application of crime prediction is underscored, emphasizing its role in helping individuals make informed decisions about their living and visiting choices based on the anticipated crime scenarios in different neighborhoods at various times.

VI. ACKNOWLEDGMENT

We would like to express my sincere gratitude to our project guide, **Mrs. Bindu K P, Assistant Professor** for their invaluable guidance, unwavering support, and insightful feedback throughout the course of this research, which have significantly contributed to the successful completion of this research project.

REFERENCES

- [1] "Crime Detection Techniques Using Data Mining and K-Means" Khushabu A. Bokde, Tisksha P. Kakade, Dnyanes hwari S. Tumsare, Chetan G. Wadhai (2018)
- [2] "Crime Prediction using KNN and Data Mining" Vrushi Pednekar, Trupti Mahale, Pratiksha Gadhave and Arti Gore (IJRITCC- 2018)
- [3] "Crime Rate Prediction Using K-Nearest Neighboring Algorithm" Akash Kumar, Aniket Verma, Gandhali Shinde and Yash Sukhdeve (IEEE-2020)
- [4] "Study on Crime Examination and Forecasting using Machine Learning" Sakshi mench, Saloni Kuge and Hafsa Bagwan (Kalahari-2021)
- [5] "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques" Wajiha Safat, Sohail Asghar, Saira Andleeb Gillani (IEEE-2021)