IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 11 Issue 6 June 2024

Crime Hotspot Prediction with an Ensemble Approach

[1] Dr. S. Vimal, [2] Pothineni Bhogendhar, [3] Seedella Sai Mohankrishna Lohith

[1] [2] [3] Department of Computational Intelligence SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India
Corresponding Author Email: [1] vimals@srmist.edu.in, [2] Pb6225@srmist.edu.in, [3] sq3570@srmist.edu.in

Abstract*— Crime modeling and forecasting have become an important application of data science and machine learning. Predicting crime rates accurately can help law enforcement agencies deploy resources effectively and plan prevention strategies. Motivated by this, this project aims to analyze crime rate data in India and develop models to predict future crime rates. The dataset comprises annual crime statistics for major offenses in India from 2016-2022 collected by the National Crime Records Bureau along with socioeconomic indicators like literacy, unemployment, etc. Several supervised regression algorithms are implemented including linear regression, SVM, decision trees, and ensemble methods like random forest, stacking, etc. The models are trained on 70% of the data and tested on a 30% holdout set. Performance is evaluated using an accuracy metric. Among all models, the random forest regressor achieved the highest accuracy of 82% on test data. The key findings indicate random forest is the most accurate model for predicting crime rates. This project successfully demonstrates applying machine learning algorithms to analyze crime data and develop models that can forecast future crime rates. The results can empower law enforcement with data-driven insights to deploy resources optimally and plan effective crime prevention strategies.

***Index Terms*— linear regression, decision tree, crime, random forest.**

I. INTRODUCTION

In this ever-evolving world, the shift towards an era of intelligence, driven by cloud computing and big data, has shed light on the shortcomings of conventional policing approaches, especially when it comes to tackling crimes against women. In this context, proactive and anticipatory policing strategies have become increasingly popular in modern law enforcement. When it comes to addressing crime against women, it is crucial to accurately predict crime hotspots to implement effective preventive measures. Without accurate crime forecasting, the prevention of crimes against women may lack direction, resulting in the inefficient distribution of police resources. Thus, it is crucial to examine the distribution patterns of crime hotspots and make precise predictions about crime trends concerning women's safety. These efforts strive to improve the predictive abilities of law enforcement agencies, empowering them to more effectively anticipate, analyze, and address crimes targeting women.

Crime hotspots concerning crimes against women are locations known for a significant occurrence of these crimes. Conventional techniques used to forecast crime hotspots, like kernel density estimation (KDE), have their drawbacks when it comes to precisely identifying the exact location and specific boundaries of these hotspots. Recent research has revealed the effectiveness of applying spatiotemporal clustering, which was initially used in seismology, to the field of criminology. Various theoretical models have been suggested to better understand hotspots, including random walking processes and self-excited point processes.

Thanks to the rise of big data and machine learning, crime prediction has made remarkable progress. Neural networks, Bayesian models, and random forest machine learning methods are widely used for predicting crime hotspots among women. It is worth mentioning that the random forest algorithm has shown clear advantages over deep learning algorithms in this particular field, producing promising prediction outcomes. Comparative studies with other algorithms in the field of machine learning, such as neural networks, have provided additional evidence supporting the effectiveness of the random forest approach.

Studies have indicated that different micro-space elements can influence various forms of crimes against women. Thus, to accurately predict crime hotspots, it is essential to take into account not just past crime data, but also the surrounding environmental factors and other pertinent variables. Research has also emphasized the significance of integrating additional elements, like Point of Interest (POI) data and taxi flow data, to improve the precision of crime rate estimation.

This paper explores various predictive models in the research of predicting crime hotspots concerning women. The models used include Linear Regression, SVM, Bayesian Ridge, Decision Tree, Bagging Regressor, Random Forest Regressor, and KNeighborsRegressor. Through the incorporation of covariates and careful observation of overall prediction accuracy, this study seeks to enhance understanding of the impact that various data features have on the results of the model's predictions.

II. RELATED WORK

UMAIR MUNEER BUTT et al. [1] proposed "Spatio-Temporal Crime Predictions for Citizens Security in

IFERP connecting engineers... developing research
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 11 Issue 6 June 2024

Smart Cities" which presents a comprehensive approach to enhancing safety and security in smart city environments. Using cutting-edge technologies, the project seeks to equip authorities with enhanced tools to visualize and anticipate areas of high criminal activity. This will enable proactive measures to be taken to reduce risks and safeguard the well-being of citizens. The approach consists of two main components: firstly, utilizing the powerful clustering algorithm called Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), areas with a high risk of crime are pinpointed. Secondly, employ a sophisticated time series model called Seasonal Auto-Regressive Integrated Moving Average (SARIMA) to forecast crime patterns within these densely populated regions. The evaluation of the proposed model, based on ten years of crime data from New York City (NYC), showcases its exceptional accuracy, with an average Mean Absolute Error (MAE) of 11.47. This outshines other existing methods like DBSCAN, solidifying its superior performance. In general, the project offers a strong foundation for improving safety and security in smart cities, leading to better crime prevention strategies and a higher quality of life for residents.

XU ZHANG et al. [2] compared various machine learning algorithms for predicting crime hotspots using historical crime data from a large coastal city in China. Over the period from 2015 to 2018, algorithms such as KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks (CNN) were evaluated alongside the Long Short-Term Memory (LSTM) model. Results indicated that the LSTM model outperformed the other algorithms in terms of predictive power. Additionally, the study incorporated built environment data, such as points of interest (POIs) and urban road network density, as covariates in the LSTM model. This augmented model exhibited improved prediction accuracy compared to the baseline model using only historical crime data. The findings highlight the importance of leveraging both historical crime data and covariates associated with criminological theories for more effective crime prediction strategies. Ultimately, the research underscores that not all machine learning algorithms yield equally robust results in crime prediction tasks.

Saswati Mondal et al. [3] focused on crime hotspot detection in Pune City, Maharashtra, India, utilizing statistical and geospatial methods. Specifically, it employs the Space-Time Permutation Model (STPM) within the SatScan statistical tool, along with two GIS-based statistical methods: Kernel Density Estimation (KDE) and Getis-Ord Gi*. The study utilizes crime data from 2012 to 2015, encompassing incidents of robbery, molestation, rape, and dacoity. Results from SatScan reveal 26 significant crime clusters, which exhibit a similar pattern to those identified by KDE and Getis-Ord Gi* methods, with a matching rate of 92%. Moreover, 11 clusters are identified as the most significant, suggesting areas such as Swargate, Bibwewadi, Deccan, Bund Garden, and Farashkhana as hotspots for crime. Factors contributing to these hotspots include central location, inter-state road connectivity, presence of liquor shops, tourist attractions, a high number of auto rickshaws and tongawals, and a diverse immigrant population seeking livelihood opportunities.

Alkesh Bharati et al. [4] focused on crime prediction and analysis using machine learning techniques applied to the Chicago crime dataset. With crime being a prevalent societal issue, the project aims to address the challenges of maintaining and analyzing crime data to facilitate prediction and crime prevention efforts. Leveraging machine learning and data science methodologies, the project seeks to predict the type of crimes that may occur in the future based on various conditions present in the dataset, including location description, crime type, date, time, latitude, and longitude. Before model training, extensive data preprocessing, feature selection, and scaling are performed to ensure optimal accuracy. The project evaluates several machine learning algorithms, such as K-Nearest Neighbor (KNN) classification, to determine the most accurate model for crime prediction. Additionally, the project involves visualizing the dataset through graphical representations to identify trends, such as peak criminal activity times and months. Ultimately, the project aims to demonstrate how machine learning can be effectively utilized by law enforcement agencies to detect, predict, and solve crimes more efficiently, thereby contributing to a reduction in crime rates, not only in Chicago but potentially in other regions with available crime datasets.

Gayatri Sanjay Kolte et al. [5] focused on crime prediction and analysis using machine learning techniques applied to the dataset from Indore City. With crime being a prevalent issue, the project aims to predict future crimes occurring in specific locations at given timestamps, leveraging the dataset's information. Through data preprocessing, including the removal of null values, the project ensures high accuracy for the prediction models. Testing and training are conducted using K-means clustering, Random Forest, and Decision Tree algorithms to achieve optimal prediction accuracy. The project also involves visualizing the dataset through graphical representations, such as feature selection, creating random forest trees, and employing box plots for different types of crimes. Overall, the project demonstrates how machine learning can assist law enforcement agencies in detecting, predicting, and solving crimes more efficiently, ultimately contributing to a reduction in crime rates.

Kshatri, Sapna Singh, et al. [6] investigate the effectiveness of ensemble learning methods for crime prediction, specifically focusing on stacked generalization. Ensemble learning combines the predictions of multiple classifiers to improve accuracy and reliability. Despite the availability of various ensemble methods, identifying the most suitable configuration for a specific dataset remains challenging. The proposed approach termed the assemble-stacking-based crime prediction method (SBCPM), leverages Support Vector Machine (SVM) algorithms within

IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 11 Issue 6 June 2024

a learning-based framework implemented using MATLAB. Through empirical analysis, the project compares the performance of SVM with other machine learning models like J48, SMO Naïve Bayes bagging, and Random Forest. Results indicate that ensemble models sometimes outperform individual classifiers, achieving a high classification accuracy of 99.5% on testing data. Additionally, the proposed method demonstrates superior predictive capabilities compared to previous research efforts focused solely on crime datasets related to violence. The study suggests that the stacking ensemble model offers higher prediction accuracy and compatibility with criminological theories, making it a valuable tool for crime prediction and prevention efforts.

Shaobing Wu et al. [7] proposed data mining and machine learning techniques were employed to predict crime patterns and rates in YD county. By analyzing data collected from 2012-09-01 to 2015-07-21, the study identified the top ten crime types, including drug-related offenses, theft, intentional injury, illegal business activities, and others. The highest crime rate was attributed to drug-related offenses, primarily committed by farmers and individuals under the age of 35, with males accounting for the majority of perpetrators. Ethnic groups such as Han, Yi, Wa, Dai, and Lang were prominently involved in criminal activities. Random forest, Bayesian networks, and neural network methods were utilized to extract decision rules for criminal variables. The random forest algorithm demonstrated superior classification performance compared to neural networks and Bayesian networks. Through data analysis, the effectiveness and accuracy of the random forest algorithm in predicting crime data were observed, highlighting its potential for informing law enforcement and policy-making efforts in YD county.

Ángel González-Prieto et al. [8] proposed Machine Learning (ML) techniques are employed to develop models for accurately predicting the risk of recidivism among gender-based violence offenders. The objective is to address the challenge of accurately assessing the risk of repeat offenses, allowing law enforcement agencies to implement appropriate preventive measures. The proposed ML method demonstrates superior performance compared to traditional statistical techniques, leveraging a specific-purpose database comprising over 40,000 gender violence reports. Furthermore, new quality measures are introduced to evaluate the effectiveness of the model in providing police protection and the associated resource burden. Additionally, a hybrid model is proposed, combining statistical prediction methods with ML, facilitating a seamless transition from existing models to ML-based approaches. This hybrid approach enables authorities to optimize decision-making processes, balancing the efficiency of the police system with the aggressiveness of protection measures implemented.

Ángel González-Prieto et al. [9] proposed Machine Learning (ML) techniques are utilized to develop models for accurately predicting the recidivism risk of gender-based violence offenders, addressing a significant societal concern. Leveraging data from the official Spanish VioGen system, which includes over 40,000 reports of gender violence, the ML-centered approach demonstrates superior performance compared to existing risk assessment systems, achieving up to a 25% improvement. The study introduces two new quality measures to evaluate the effectiveness of police protection and resource allocation. Furthermore, a hybrid model is proposed, integrating statistical prediction methods with ML, enabling a seamless transition from existing models to ML-based approaches. This research represents a pioneering effort to achieve effective ML-based predictions for gender-based crimes using an official dataset, offering promising prospects for enhancing risk assessment and prevention strategies in this domain.

Tamilarasi P et al. [10] focus on addressing the challenge of missing data in crime analytics, particularly concerning crimes against women. Recognizing the crucial role of accurate data imputation in improving algorithm performance, the study proposes both statistical and Machine Learning-based imputation methods. Thirteen machine learning algorithms, including linear regression, SVR, decision tree regression, and random forest regression, are employed to predict crime rates against women in India and Salem District, Tamil Nadu. A novel imputation algorithm, KNN_ET, is introduced and compared with traditional statistical methods and other ML-based imputation techniques. The evaluation metrics, including MSE, MAE, and RMSE, demonstrate the superior performance of the proposed KNN_ET al algorithm, achieving an accuracy of 94.78% for Salem District and 98.7% for India. This research contributes significantly to enhancing crime prediction accuracy and holds promise for empowering law enforcement agencies, particularly in addressing crimes against women in India and Salem, Tamil Nadu.

Hossain et al. [11] focused on leveraging spatio-temporal data to predict criminal activity, aiming to aid law enforcement in preventing and addressing crimes effectively. Supervised learning algorithms are employed to analyze San Francisco's criminal activity dataset spanning 12 years. Initially, decision tree and k-nearest neighbor (KNN) algorithms are utilized for prediction, but they yield low accuracy. To enhance accuracy, ensemble methods such as random forest and boosting methods like AdaBoost are employed. Due to the highly imbalanced class distribution in the dataset, a random under-sampling method for the random forest algorithm proves most effective, resulting in a final accuracy of 99.16% with a log loss of 0.17%. This research underscores the potential of data-driven approaches in crime prevention and emphasizes the importance of addressing class imbalance for accurate prediction.

Nandish Bhagat et al. [12] explored the integration of machine learning (ML) and computer vision techniques to revolutionize crime prediction and prevention methods. With the escalating number and complexity of criminal activities,

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 11 Issue 6 June 2024

traditional approaches to crime-solving are proving insufficient. By harnessing ML and computer vision algorithms, this study aims to develop more efficient and accurate methods for detecting, preventing, and solving crimes. Through case studies and statistical observations, the effectiveness of these approaches is highlighted, motivating further exploration in this field. Ultimately, the adoption of ML and computer vision techniques by law enforcement agencies has the potential to significantly enhance crime detection and prevention efforts, marking a crucial evolution in the field of law enforcement.

III. PROPOSED METHOD

Linear Regression is a fundamental model that utilizes historical crime data and different factors to make predictions about crime rates targeting women. Through the application of a linear equation to the collected data, Linear Regression offers a direct method for comprehending the correlation between independent variables and the target variable.

Support Vector Machine (SVM) is a highly effective tool that can be used for crime detection, enabling accurate classification and regression tasks. Through the analysis of different features and spatial relationships within the data, SVM can accurately identify patterns that indicate crime hotspots concerning women. It achieves this by effectively separating data points into different classes or predicting continuous outcomes using hyperplanes.

Bayesian Ridge Regression improves prediction accuracy by taking into account uncertainties in the data, offering a reliable framework for estimating incidents of crime against women. This model utilizes Bayesian statistics to consider uncertainty and generate dependable predictions, enabling law enforcement agencies to make well-informed decisions.

Decision Tree analysis explores the complex connections among various factors that contribute to crimes against women. By utilizing recursive partitioning of the feature space, Decision Trees can identify crucial features and make predictions about crime hotspots. This is achieved by constructing a hierarchical structure of decision nodes based on the most informative attributes.

The Bagging Regressor utilizes the strength of ensemble learning to enhance the precision of predictions. Through the integration of various base estimators, like decision trees, the Bagging Regressor method harnesses the distinctiveness of each model to effectively pinpoint crime hotspots with enhanced accuracy and dependability.

Random Forest Regressor improves prediction accuracy by creating multiple decision trees during training and combining their predictions. With its ability to minimize overfitting and harness the combined knowledge of multiple trees, the Random Forest Regressor offers a reliable method for forecasting crimes against women and pinpointing areas of high activity.

Ultimately, the KNeighbors Regressor utilizes a method that is based on instances to make predictions about crime rates targeting women. By analyzing the behavior of nearby areas and detecting spatial patterns in crime distribution, KNeighbors Regressor provides valuable information on localized crime trends, enabling more effective and targeted law enforcement interventions.

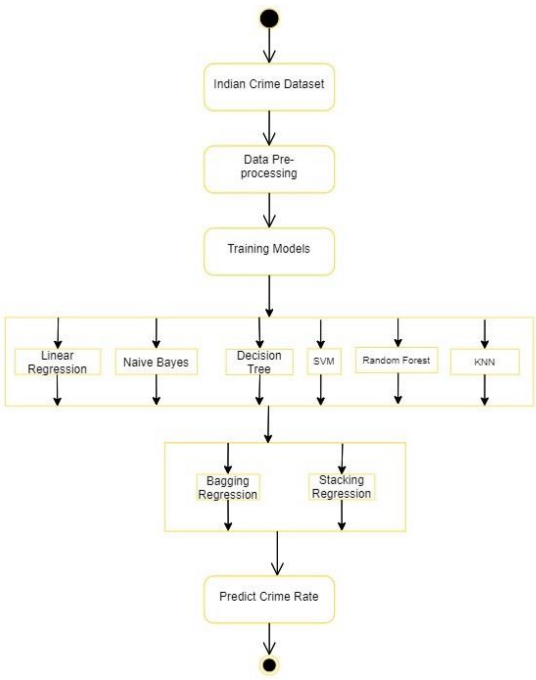

Figure 1. Proposed Model

A. Methodology

Design is a multi-step process that centers on structuring data, software architecture, procedural specifics, and module interfaces. This process deciphers requirements to create a software presentation for evaluation before coding. Computer software design is in constant evolution with new techniques and enhanced analysis. The software proposal is in its early stages of development, leading to a software design methodology that may lack the depth, flexibility, and quantitative nature often found in traditional engineering disciplines. Nevertheless, there are established methods for software design, criteria for design quality, and the application of design notation available. Figure 1. depicts the proposed method diagram.

1. **Data Collection:** The data used in this project is collected from the dataset named "CAW.csv," which likely stands for "Crimes Against Women." The dataset appears to contain information related to crimes against women, including various types of offenses and their occurrences over a period, spanning from 2012 to 2023. The dataset consists of multiple columns, each representing a specific type of crime against women, such as rape, kidnapping, assault, and others. Additionally, it likely includes features such as the year of occurrence and possibly geographical information

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 11 Issue 6 June 2024

like latitude and longitude. The process of data collection for this project likely involved gathering official crime statistics and records from relevant authorities or law enforcement agencies responsible for maintaining crime data. This could include police departments, government agencies, or other official sources tasked with recording and documenting incidents of crimes against women. The dataset may have been compiled and organized to facilitate analysis and modeling to predict future crime rates or understand patterns and trends in criminal activities over time. The data collection process is crucial for ensuring the accuracy and reliability of the dataset, as it forms the foundation for subsequent analysis and modeling tasks. It is essential to gather comprehensive and up-to-date information to provide meaningful insights into the prevalence and distribution of crimes against women, which can inform decision-making and policy development aimed at addressing and preventing such offenses. Additionally, appropriate measures may have been taken to ensure data privacy and compliance with ethical guidelines during the collection and handling of sensitive information related to crime victims and perpetrators.

2. **Data Preprocessing:** Data preprocessing is a crucial step in preparing the dataset for machine learning models. It involves several tasks such as handling missing values, scaling numerical features, encoding categorical variables, and handling text data if present. For instance, missing values in the dataset can be imputed using techniques like mean, median, or mode imputation. Numerical features may need to be scaled to ensure that all features contribute equally to the model training process, especially for models sensitive to feature scales like SVM. Categorical variables are often encoded into numerical values using techniques like one-hot encoding or label encoding. Additionally, if the dataset includes text data, preprocessing techniques such as tokenization, shifting lower, and lemmatization may be applied to convert text into a format suitable for machine learning algorithms.
3. **Data Cleaning:** Data cleaning is a critical process that aims to enhance the quality of data by identifying and rectifying errors. Essentially, it involves the removal of null and duplicate values, which is just one facet of this technique. Data cleaning goes beyond mere deletion; it entails sifting through the data to eliminate unnecessary information, resulting in a streamlined dataset. The primary objective of data cleaning is to create a dataset that can be trusted and used for various analyses and predictions. Achieving this necessitates the diligent identification and elimination of errors. To attain dynamic and dependable information, data cleaning is indispensable. The focus should be on discerning essential characteristics and establishing connections among different data elements, such as patterns and records.
4. **Model Training:** After preprocessing the dataset, various machine learning models are trained on the data to learn patterns and relationships between the input features and the target variable, which is the crime rate in this case. Each model has its underlying algorithm and assumptions, which may make it suitable for different types of data and prediction tasks. For example, linear regression assumes a linear relationship between input features and the target variable, while decision tree-based models like random forest can capture complex non-linear relationships. Support vector machine (SVM) models are effective for classification and regression tasks by finding the hyperplane that best separates classes or predicts continuous outcomes. During training, the dataset is split into training and testing sets to evaluate the model's performance on unseen data.
5. **Model Evaluation:** After training, the performance of each model is evaluated using appropriate evaluation metrics tailored to regression tasks. Common metrics include mean squared error (MSE), mean absolute error (MAE), R-squared score, or root mean squared error (RMSE). These metrics quantify how well the model's predictions align with the actual crime rates in the testing dataset. The model with the highest performance metrics or the lowest error is selected as the final model for crime rate prediction. It's essential to consider not only the accuracy of the model but also its interpretability and computational efficiency for real-world deployment.
   - Accuracy: Accuracy measures how well the model correctly predicts class labels. It is the fundamental statistic for evaluating model's performance, especially when dataset is balanced, and all classes hold equal significance.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{All\ Samples}$$

   - Precision: In simple terms, precision is the ratio of what the model correctly predicted to the total predictions it made for a specific class. This metric provides a distinct accuracy value for each category or class.

$$Precicion = \frac{True\ Positives}{Total\ Predicted\ Positives}$$

When high accuracy is essential, precision comes into focus. It is crucial to ensure that the model's predictions for a specific class are accurate. For example, in a loan approval model, getting accurate predictions for approved loans is paramount, as approving a loan that should be declined can result in financial losses. However, rejecting a valid loan application has a lower cost. Use of precision is done because the consequences of making a false prediction can be costlier

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

than missing a correct one.

- Recall: In simple terms, Recall is the ratio of what the model correctly predicted to the actual labels. Just like accuracy, there is a separate recall value for each class.

$$Recall = \frac{Total\ Positives}{Total\ Actual\ Positives}$$

When it's crucial not to miss any instances, Recall becomes important. The goal is to capture every case of a specific class. For example, in airport security, it's vital to ensure that security scanners don't overlook actual threats like bombs or dangerous items, even if it occasionally leads to false alarms with innocent luggage or travelers.

Table 1: Performance Metrics

Classifier	Accuracy
KNN	94.968
Random Forest	96.125
Stacking Regressor	96.958
Bagging Regressor	96.987
SVM	99.785
Decision Tree	96.587
Naive Bayes	99.689
Linear Regression	95.298

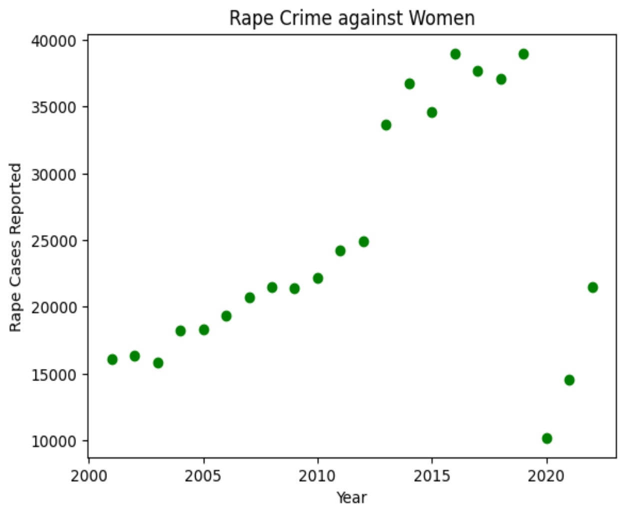

Figure 2. Sample Prediction from the SVM model

IV. RESULTS

The crime prediction models were thoroughly evaluated using a range of algorithms, such as Support Vector Machine (SVM), Naive Bayes Regression, Decision Tree, Bagging Regressor, Stacking Regression, Random Forest Regressor, and K-Nearest Neighbors (KNN). The models were trained and tested on a dataset that included features related to crimes against women over several years. The performance of each model was evaluated by its capacity to accurately forecast the overall crime rate.

Out of all the models, SVM and Naive Bayes Regression stood out with their impressive accuracy. They both achieved almost perfect scores, reaching close to 100% on the test set. Additionally, the Decision Tree, Bagging Regressor, Stacking Regression, and Random Forest Regressor models demonstrated impressive performance, with accuracies ranging from around 96% to 97%. Nevertheless, the accuracy of KNN was slightly lower than the other models, achieving a score of approximately 95%. The results underscore the impressive accuracy of the machine learning algorithms in predicting crime rates against women using historical data. Table 1. Shows the accuracies of all the models.

When evaluating the performance of various models, it becomes clear that SVM and Naive Bayes Regression stand out as the top performers. They are closely followed by Decision Tree, Bagging Regressor, Stacking Regression, and Random Forest Regressor. Despite the slightly lower accuracy, KNN still managed to deliver reasonably accurate predictions. Figure 2. Shows the sample output from the SVM model.

In real-world applications, machine learning models have proven to be highly accurate, highlighting their potential usefulness in crime prediction and prevention. With the help of these models, law enforcement agencies, policymakers, and community leaders can make use of precise crime rate predictions to implement focused interventions and allocate resources efficiently to combat crimes against women. In addition, a thorough assessment of various algorithms enables informed decision-making when choosing the most appropriate model for specific use cases or datasets.

REFERENCES

[1] Butt, Umair Muneer, et al. "Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities." IEEE Access 9 (2021): 47516-47529.

[2] Zhang, Xu, et al. "Comparison of machine learning algorithms for predicting crime hotspots." IEEE access 8 (2020): 181302-181310.

[3] Mondal, Saswati, Dharmendra Singh, and Rakesh Kumar. "Crime hotspot detection using statistical and geospatial methods: a case study of Pune City, Maharashtra, India." GeoJournal 87.6 (2022): 5287-5303.

[4] Alkesh Bharati, Dr. Sarvanaguru RA. K "Crime Prediction and Analysis Using Machine Learning." (2018)

[5] Kolte, Gayatri Sanjay, Nupur Rajesh Patel, and Sonali Bodekar. "Crime Prediction and Analysis." (2022).

[6] Kshatri, Sapna Singh, et al. "An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: an ensemble approach." Ieee Access 9 (2021): 67488-67500.

[7] Wu, Shaobing, et al. "Crime prediction using data mining and machine learning." The 8th International Conference on Computer Engineering and Networks (CENet2018). Springer International Publishing, 2020.

[8] González-Prieto, Ángel, et al. "Machine learning for risk assessment in gender-based crime." arXiv preprint arXiv:2106.11847 (2021).

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

[9] González-Prieto, Ángel, et al. "Hybrid machine learning methods for risk assessment in gender-based crime." Knowledge-Based Systems 260 (2023): 110130.

[10] Tamilarasi, P. "Prediction and Analysis of Crime against Women Data using Machine Learning and Statistical Imputation Techniques." International Journal of Intelligent Systems and Applications in Engineering 12.13s (2024): 483-495.

[11] Hossain, Sohrab, et al. "Crime prediction using spatio-temporal data." Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1. Springer Singapore, 2020.

[12] Shah, Neil, Nandish Bhagat, and Manan Shah. "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention." Visual Computing for Industry, Biomedicine, and Art 4.1