**Mini Project Report on**



**CRIME RATE PREDICTION**

**USING**

**MACHINE LEARNING TECHNIQUES**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**RUDRAKSH AGARWAL 2021907**

***Under the Mentorship of***

**Mr. Vivek Tomar**

**Assistant Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**July-2024**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“ CRIME RATE PREDICTION USING MACHINE LEARNING TECHNIQUES ”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Mr. Vivek Tomar, Assistant Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name University Roll No.

RUDRAKSH AGARWAL 2021907

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | **1-2** |
| Chapter 2 | Literature Survey | **3-4** |
| Chapter 3 | Methodology | **5-7** |
| Chapter 4 | Result and Discussion | **8-9** |
| Chapter 5 | Conclusion and Future Work | **10** |
|  | References | **11** |

**Chapter 1**

**Introduction**

The Purpose of this project is to develop a model for Crime Rate Prediction Using Machine Learning Techniques. Being a Regression Model, it should be able to predict the crime rate of the given data with the minimum mean squared error possible.

**1.1 Introduction**

Crime Rate Prediction involves using statistical methods and machine learning algorithms to forecast the likelihood and frequency of criminal activities in a given area over a specific period. This Predictive analysis aims to identify patterns and trends based on historical crime data and other relevant constraints. Here, in this project we have focused on implementing Crime Rate Prediction model using various Machine Learning Techniques. Crime Rate Prediction has many applications, some of them are:

**1.1.1 Law Enforcement**

Predictive Models can assist law enforcement agencies in locating areas where crime is more likely to occur, which helps them allocate resources better. Police can also plan patrol routes and schedules based on anticipated crime outcomes to maximize their presence in high-risk areas. It can also aid in the development and execution of targeted crime prevention measures.

**1.1.2 Urban Planning and Infrastructure**

The use of Crime Rate Prediction Models helps the planner in constructing safer urban areas by installing street lights or signaling cameras on vehicles, as well as installing public spaces.

**1.1.3 Commercial and Business Use**

Businesses, particularly retail and real estate, can use crime forecasts to access the risk of operating in specific locations and make appropriate choices regarding these locations.

For Businesses like Insurance Companies, take it as a benefit in their marketing plans, by creating premium and coverage plans in various areas.

**1.1.4 Policy and Public Safety**

Based on Projected Trends and underlying factors, government agencies can develop policies aimed at reducing crime. The Government can also initiate awareness programs in areas with \

high crime rates.

**1.1.5 Research and Academic Research**

Academics can use predictive models to study crime patterns, understand the causes of crime, and evaluate the effectiveness of different interventions. Researchers also analyze the impact of various public policies on crime rates and suggest improvements based on predictive analysis.

**1.2 Objective**

The Main Objectives of Crime Rate Prediction using Machine Learning Techniques are:

* To design & develop a crime rate prediction using machine learning algorithms.
* To Evaluate the performance of different models on a given dataset.
* To analyze the outputs of all the models and find the best model.

**Chapter 2**

**Literature Survey**

A literature survey on Crime Rate Prediction using Machine Learning Techniques includes research and implementation of various models on various datasets used in some previous studies stated that crime rate prediction has a vast scope in various fields. Some of them are:

An earlier case study by Wajiha Safati, Sohail Asghar and Saira Andleeb Gillani [1]applied various machine learning algorithms to improve security in large cities through accurate crime prediction and forecasting. Models such as logistic regression, SVM, Naive Bayes, KNN, decision tree, MLP, random forest, XGBoost, LSTM and ARIMA were used in the study. Among them, LSTM performed well in terms of RMSE and MAE. The analysis revealed more than 35 types of crime, with a year-over-year decrease in crime in Chicago and a slight increase in Los Angeles. Future projections showed moderate increases in crime in Chicago and sharp decreases in Los Angeles, identifying specific hot spots. The results were intended to improve early crime detection and inform law enforcement strategies with better predictive accuracy.

The research conducted by Sakib Mahmud, Musfika Nuha, and Abdus Sattar, which is a part of the book series Advances in Intelligent Systems and Computing, Vol 1248[2], focuses on analyzing crime patterns and trends to improve community safety using data mining techniques. It addresses security concerns such as hijackings, kidnappings and harassment during everyday travel. The study uses clustering methods and the K-Nearest Neighbor (KNN) algorithm to analyze crime rates in Bangladesh. Examining both primary and secondary data, the study predicts crime levels in different locations and uses these predictions to identify safer routes. The goal is to help people be aware of high crime areas and find safe routes to their destinations.

In a Research by Shraddha Ramdas Bandekar and C. Vijayalakshmi, part of Procedia Computer Science, Vol. 172 Released in 2020 [3], focuses on leveraging machine learning algorithms to reduce crime in India by identifying patterns and correlations in crime data. The study highlights the importance of maintaining a detailed criminal database for future reference. Through analysis and visualization, the study aims to predict crime types by location and develop proactive measures to improve public safety. The methodology includes data cleaning, transformation and application of clustering and optimization algorithms for statistical analysis. The ultimate goal is to identify risk factors and implement strategies to protect society.

Lawrence McClendon and Natarajan Meghanathan's research, published in the March 2015 issue of Machine Learning and Applications [4], focuses on using the open-source data mining software WEKA to compare violent crime patterns from community and crime anomalous datasets with actual crime. The study evaluates linear regression, additive regression and decision algorithms using consistent features of the dataset, highlighting the superior performance of linear regression in predicting crime patterns. The purpose of the research is to highlight the effectiveness of machine learning algorithms in enhancing crime detection and prevention strategies.

**Chapter 3**

**Methodology**

**3.1 Dataset**

Here, we have used the Indian Crime Rate Dataset in which we have state-wise crime record for 12 consecutive years from 2001-2012, for which we have to analyze the dataset and implement multiple Regression models. So that, we can store their outputs and analyze them.

**3.1.1 Shape of Dataset**

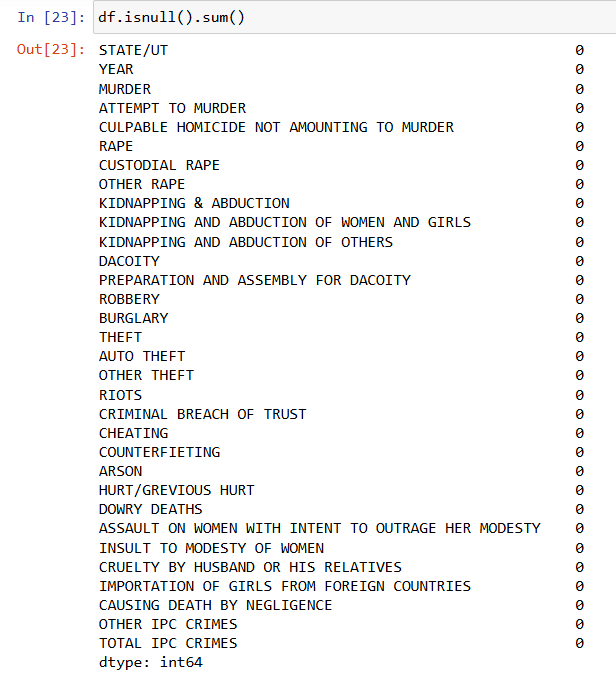
Here, we have considered a dataset containing 420 Rows and 32 Columns as shown in Fig. 3.1, which itself is a large dataset having columns containing both numerical as well as categorical values.

****

**Fig. 3.1**

**3.1.2 Checking for Null Values**

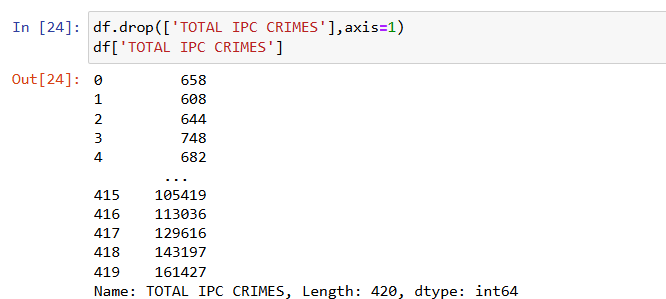
Being a Dataset with numerical values in a large number the dataset has no null values as shown in figure 3.2

****

**Fig 3.2**

**3.2 Preprocessing**

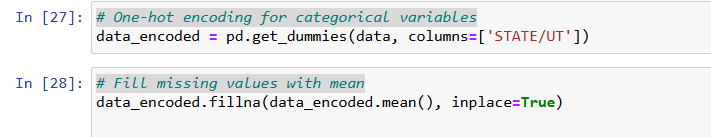
Data Preprocessing is the process of making changes in the dataset for making it suitable to use in a model, here, we have dropped the column TOTAL IPC CRIMES as shown in Fig. 3.3. and applied Encoding.

****

**Fig. 3.3**

**3.2.1 One Hot Encoding for Categorical Variables**

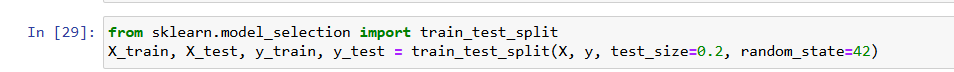
Now, initially we have to preprocess the dataset to make it more suitable for the model. For which we have firstly applied One-Hot Encoding for all Categorical Variables, as shown in Fig. 3.4.



**Fig. 3.4**

**3.3 Splitting of Data**

We have made a dataset split between testing data and training data, i.e., 80% Training Dataset & 20% Testing Dataset, as shown in Fig. 3.5.



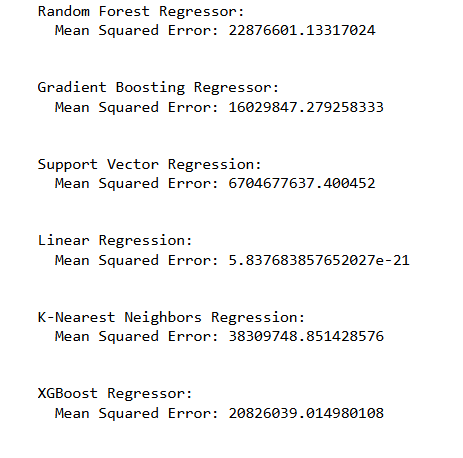
**Fig. 3.5**

**3.4 Model Selection**

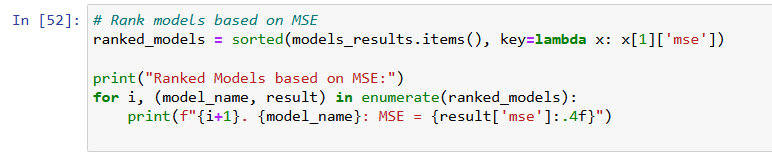
Now, we have selected six different regression algorithms namely Random Forest Regressor, Gradient Boosting Regressor, Support Vector Regression, Linear Regression, K-Nearest Neighbors Regression and XGBoost Regressor.

**3.5 Training & Evaluation**

We have to trained our model on training data using all six regression models. After that we have calculated the Mean Squared Error (MSE) for the testing data as shown in Fig. 3.6. Later we have analyzed the Mean Squared Error for each model and rank the models on the basis of Mean Squared Error as shown in Fig. 3.7.

****

**Fig. 3.6**

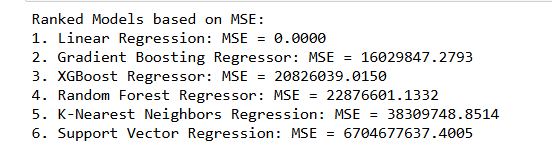
****

**Fig. 3.7**

**Chapter 4**

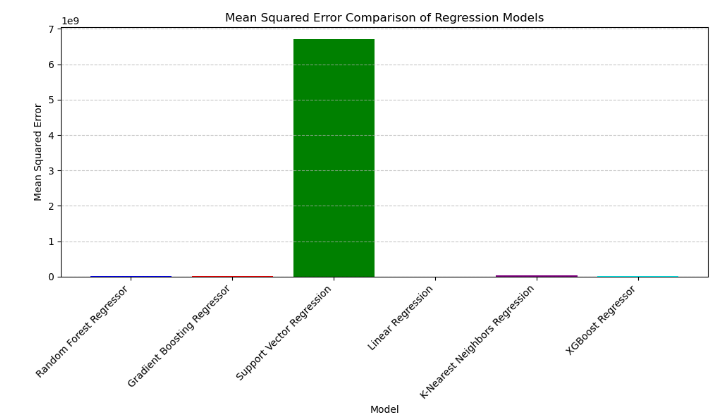
**Result and Discussion**

The Results for the Crime Rate Prediction Model is: The Models are Ranked on the Basis of their Mean Squared Error as shown in Fig. 4.1.



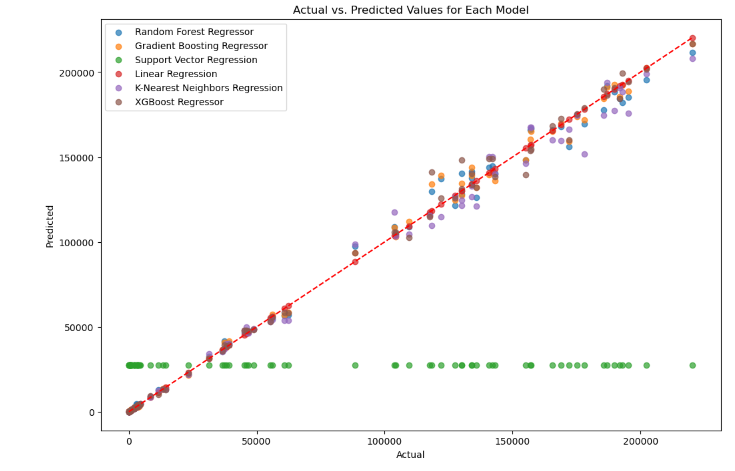
**Fig 4.1**

The Graph for the Comparison of the Mean Squared Error of all Regression Models as shown in Fig. 4.2:

****

**Fig 4.2**

Now, we have displayed a scatter plot for Actual Values and Predicted Values for all the models we have used as shown in Fig 4.3

****

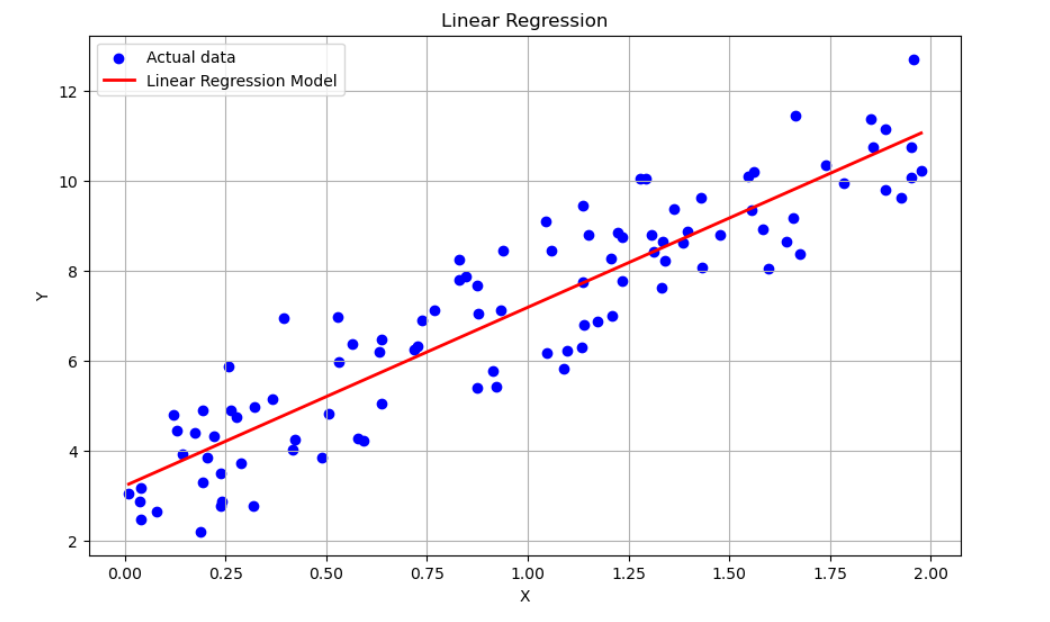
**Fig 4.3**

We have calculated a Cross-Validation Score for the Linear Regression Model which was ranked highest among all the other models as shown in Fig. 4.4.



**Fig 4.4**

Now, we have also plotted a scatter plot for Actual data and Linear Regression Model to verify its performance as shown in Fig 4.5.

****

**Fig 4.5**

**Chapter 5**

**Conclusion and Future Work**

**5.1 Conclusion**

In this project, a Crime Rate Prediction model developed using various Machine Learning techniques provides the best performance on the dataset using a linear regression model. The model’s Mean Squared Error is the parameter on which we have ranked all the models used. Among all of them Linear Regression is the most suitable for use in a variety of applications like Predictive Policing, Crime Analytics, Law Enforcement Forecasting, Crime Hotspot Detection, etc.

This crime rate prediction project used machine learning models such as random forest regressor, gradient boosting regressor, support vector regression, linear regression, K-nearest neighbor regressor and XGBoost regressor and ranked them by mean squared error (MSE). Linear regression achieved the lowest MSE, indicating better prediction accuracy. This approach improves applications such as proactive policing, optimization of police resources, informed decision making and proactive crime prevention, effectively enhancing overall public safety efforts.

**5.2 Future Work**

In the future, the accuracy and applicability of the models can be improved in predicting crime. Investigating ensemble methods combining multiple algorithms can potentially improve predictive power beyond the performance of a single model. In addition, the inclusion of more diverse and comprehensive datasets, including real-time data streams and socio-economic factors, could provide deeper insights into crime patterns. The integration of advanced deep learning techniques such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) can provide better temporal and spatial analysis of crime trends. In addition, refining the interpretability of models and ensuring ethical considerations in the use of information are crucial steps in implementing robust, open and fair predictive systems in crime prevention and law enforcement.

**References**

[1] W. Safat, S. Asghar and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," in IEEE Access, vol. 9, pp. 70080-70094, 2021, doi: 10.1109/ACCESS.2021.3078117.

[2] Mahmud, S., Nuha, M., Sattar, A. (2021). Crime Rate Prediction Using Machine Learning and Data Mining. In: Borah, S., Pradhan, R., Dey, N., Gupta, P. (eds) Soft Computing Techniques and Applications. Advances in Intelligent Systems and Computing, vol 1248. Springer, Singapore. <https://doi.org/10.1007/978-981-15-7394-1_5>

[3] Bandekar, S. R., & Vijayalakshmi, C. (2020). Design and Analysis of Machine Learning Algorithms for the reduction of crime rates in India. *Procedia Computer Science*, *172*, 122–127. <https://doi.org/10.1016/j.procs.2020.05.018>

[4] McClendon, L., & Meghanathan, N. (2015). Using Machine Learning Algorithms to Analyze Crime Data. *Machine Learning and Applications*, *2*(1), 1–12. <https://doi.org/10.5121/mlaij.2015.2101>