iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50310

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Crime Analysis and Prediction Using Machine Learning

Prakash Maurya[1], Tahir Shaikh[2], Imran Ahmed[3], Amaan Firdosi[4], Prof. Kiran Deshmukh[5]
[1, 2, 3, 4]*B.E. Student,* [5]*Assistant Professor, Dept of Information Technology, VPPCOE&VA, Mumbai, Maharashtra, India*

Abstract: ***The aim of this study is to develop a machine learning-based application that can analyze crime data across different districts in India and categorize them as high, moderate, or low based on the frequency of crimes. Based on the frequency of particular crimes in particular districts we will suggest necessary preventive measures and also recommend some precautions before visiting a particular crime hotspot. The methodology involved using a Logistic regression model for crime classification, followed by k-means clustering to group districts based on their crime rates. The results of the study demonstrated the efficacy of the machine learning model in accurately classifying crimes. The original contribution of this research lies in the development of an application that provides users with valuable insights into the crime rates in different districts.***
Keywords: Machine Learning, Crime Analysis, Linear Regression, Logistic Regression, K-Means clustering etc.

I. INTRODUCTION

In India, the incidence of crimes has been on a steady rise, and people are often unaware of the types of crimes that occur frequently in their districts or the places they visit. Unfortunately, there are no specific applications that provide users with detailed insights into crime rates in their area. To address this issue, we have developed a project that leverages data from the NCRB website, an open web source that stores records of crimes across India. The data is fetched from the website and then cleaned to meet the project's requirements. The project utilizes two algorithms: Logistic Regression and K Means algorithm. Logistic regression predicts the crime rate for the current year based on the previous year's records, and K Means is used to create clusters based on low, moderate, and high crime rates. The findings of this analysis will be presented in an Android application, which will provide users with a better understanding of crime rates in their area.

II. PROPOSED SYSTEM

A. *Problem Statement*

Criminal activity in India is on the rise and has a significant negative impact on society. The recent surge in crime has left many wondering what the future holds. Instances of murder, abduction, rape, and fatal accidents have seen a sharp increase. It is essential to raise awareness of this issue and make people understand the severity of the situation. Machine learning advancements and deep learning algorithms can help identify new patterns in different data sets and uncover previously unknown information. Crime prediction and the identification of criminals are crucial problems that need to be addressed by the police department, given the vast amount of crime-related data available. A technological solution is urgently needed to speed up the process of solving cases.

B. *Proposed Methodolgy*

1) *Web sources and Excel:* For our project, we sourced data from the NCRB website, which is a trusted government repository for crime data and is open to all. The data was then fetched into Excel and processed to meet the project's requirements. Unnecessary data was removed, and the relevant data was formatted appropriately. Finally, the processed data was stored in CSV format for further analysis.
2) *Android Application:* The Android application we have developed for our project has three pages. The first page provides users with a detailed analysis of crimes across various districts in India, including a pie chart that highlights the prevalence of different types of crimes in each district. Users can easily search for their district and access the analysis. The second page provides an analysis of crime rates across districts, classified as low, moderate, or high. The data is visualized through a map to aid in better understanding. The third page offers views and suggestions on how to tackle crimes and ensure the safety of families and society. It serves as a platform for users to share their thoughts and suggestions to combat crime.

3) Model Development

a) Step1: Data Preprocessing

We start by Pre-processing the data from the dataset. Removing redundant spaces, filling empty spaces and also removing unwanted values from rows and columns. After doing this we do outlier analysis to detect anomalous values.

After this preprocessing, we transform the dataset into necessary format for applying algorithms for model building.

Divide the training data into two sets: train and validation.

In the current stage of the project, the focus has shifted from data pre-processing and feature engineering to the creation and assessment of the machine learning model. In order to effectively evaluate the performance of the model, it is crucial to divide the available training data into two distinct sets: a train set and a validation set. A validation set is a portion of the data that is held out from the training process and is used to evaluate the performance of the model.

In this particular case, the validation set was created by randomly selecting 20% of the training data. This split ensures that the model is trained on a majority of the data and is tested on a relatively smaller portion of the data, giving a better estimate of its performance on unseen data.

b) Step 2: Model Creation

During this step, a machine learning model was selected to be trained on both the train-set and validation set. Linear regression is an algorithm that establishes a linear connection between an independent variable and a dependent variable, making it possible to forecast the outcome of future events. It is a statistical technique utilized in data science and machine learning for predictive analysis.

On the other hand, logistic regression is a supervised machine learning algorithm commonly used for classification tasks, where the objective is to predict the probability of an instance belonging to a given class or not. It is a type of statistical algorithm that examines the correlation between a set of independent variables and the dependent binary variables, making it a powerful tool for decision-making. An example of its application is email spam identification.

In this case, logistic regression was applied to the data set for training purposes, and subsequently, it was employed to predict the crime rate for a year in which data was already available to verify the accuracy. Logistic regression was preferred over other algorithms due to its higher level of precision.

c) Step 3: Predicting the current year data

In this step trained machine learning model is used to predict the current year data. The model gives all the crime rates categorized in different crimes which we will use in further analysis.

d) Step 4: Clustering the predicted data

During this step, K-means clustering is utilized to cluster the data. K-means is an unsupervised clustering algorithm that aims to divide unlabeled data into a specific number of distinct groups, referred to as "K". Essentially, K-means identifies observations that share significant attributes and groups them together into clusters. A successful clustering solution is one that identifies clusters in which the observations within each cluster are more similar than those in other clusters. The districts are then categorized into three clusters labeled as low (0), moderate (1), and high (2).

e) Step 5:

In this step, we will integrate the cluster data with the map. Once we have created clusters of data for specific crimes, we will generate csv files for them. These csv files will then be uploaded to Google Maps as input data. We will then plot the cluster values of 0, 1, and 2 as low, moderate, and high.

4) Visualizing output in Google Maps: Visualizing the data on maps will give a better insight to user to understand the crime rate in different districts.

C. System Architecture

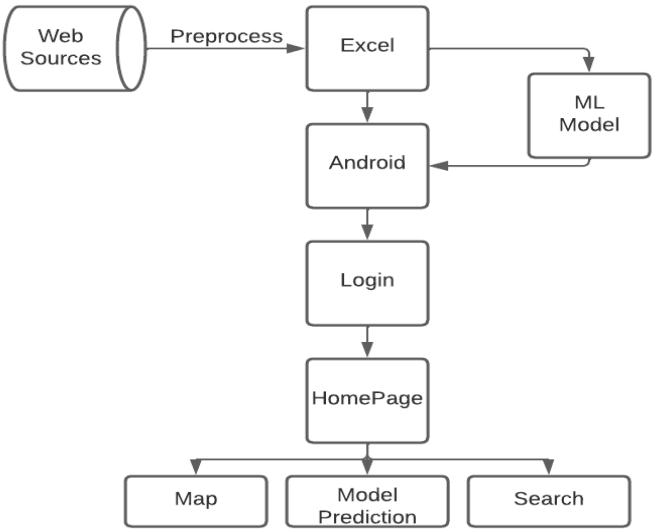


Fig -1: System Architecture

III. RESULT

We have prepared a map containing the anticipated crime data of the district. This map is then integrated into the end-user application and is accessible to users through the application. They can identify the primary crime hotspots and learn which crimes are most prevalent in the district. Based on the severity and types of crime, the application suggests necessary preventive measures to the users. Additionally, some precautions are also provided to consider before visiting a particular crime hotspot district.

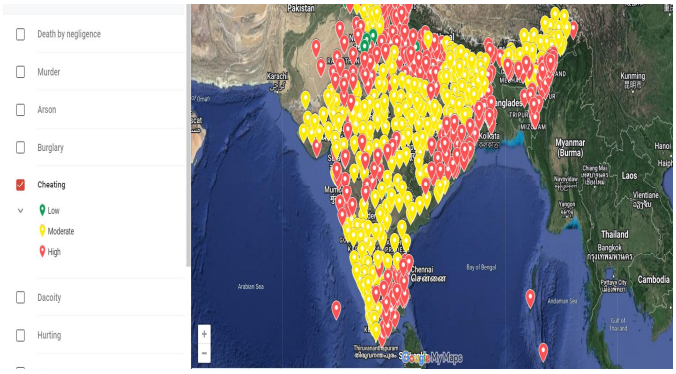


Fig -2: Cheating Plot

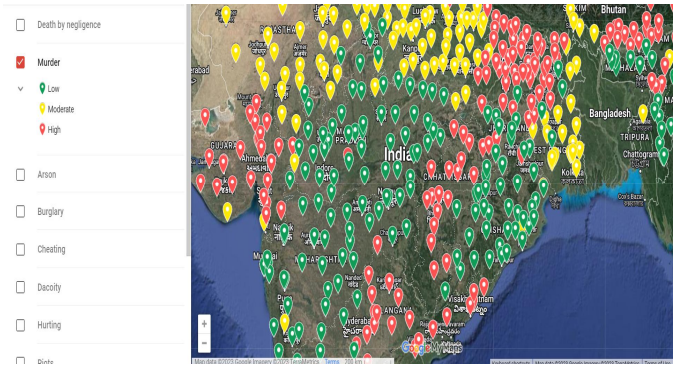


Fig -3: Murder Plot

IV. FUTURE SCOPE

There are several potential future applications for the crime data and map developed in the above project.

One potential future scope could be to integrate real-time crime data into the map, allowing users to stay updated with the latest crime trends and hotspots in their district. This could help users avoid potentially dangerous areas and make more informed decisions about where to go and when.

Additionally, the project could be expanded to include more types of crime data, such as cybercrime, white-collar crime, and environmental crime.

This would require additional data sources and analysis methods but could provide a more comprehensive understanding of crime trends in the district or region.

Overall, there are many potential future applications for the crime data and map developed in the above project, and further research and development could lead to even more valuable insights and applications in the future.

V. CONCLUSIONS

In conclusion, the project aimed to develop a map of predicted crime data for a particular district. Through data collection, cleaning, and analysis, we were able to identify crime hotspots and classify them into three categories based on severity. The resulting map was integrated into an end-user application, allowing users to access information about the most prevalent crimes in their district and take necessary precautions before visiting potentially dangerous areas.

Overall, the project demonstrates the potential of data analysis and mapping technologies to improve public safety and inform decision-making. By using data to identify crime hotspots and trends, we can take proactive measures to prevent crime and improve public safety. While there is still much work to be done in refining the accuracy and scope of the project, it represents an important step towards using data-driven approaches to address complex social issues.

VI. ACKNOWLEDGEMENT

We would like to express our heartfelt appreciation to Professor Kiran Deshmukh, Assistant Professor of the Information Technology Department, for his valuable guidance and encouragement throughout the course of this project. We are also grateful to the faculty members of our department for their assistance and support during the completion of this project.

Furthermore, we extend our sincere thanks to the Principal of Vasantdada Patil Pratishthan's College of Engineering for providing us with the opportunity to undertake this project and for their continued support throughout its development.

REFERENCES

[1] Llaha, "Crime Analysis and Prediction using Machine Learning," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, pp. 496-501, doe: 10.23919/MIPRO48935.2020.9245120.

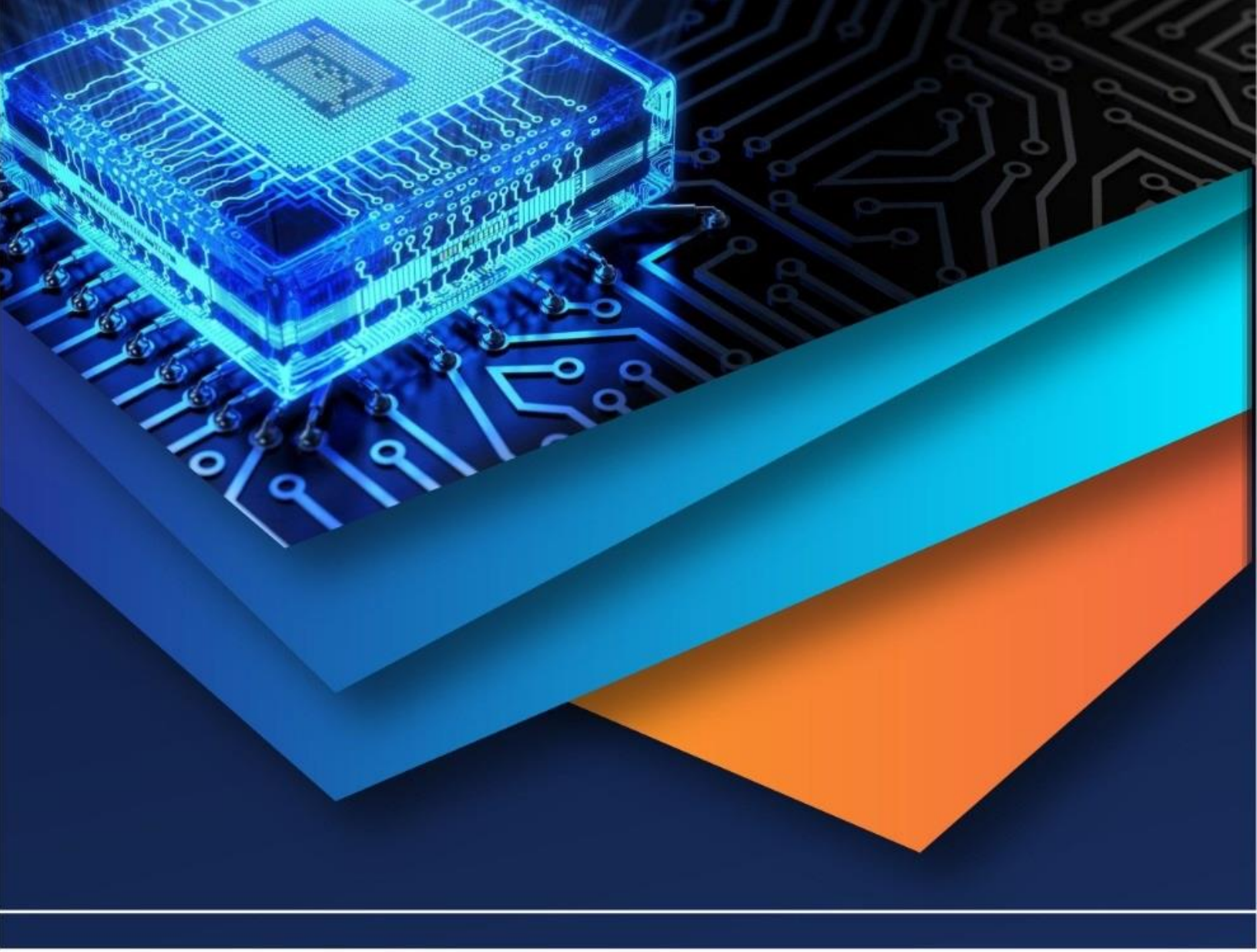
[2] A. Shukla, A. Katal, S. Raghuvanshi and S. Sharma, "Criminal Combat: Crime Analysis and Prediction Using Machine Learning," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-5, doi: 10.1109/CONIT51480.2021.9498397

[3] Pratibha, A. Gahalot, Uprant, S. Dhiman and L. Chouhan, "Crime Prediction and Analysis," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170731

[4] A. Mary Shermila, A. B. Bellarmine and N. Santiago, "Crime Data Analysis and Prediction of Perpetrator Identity Using Machine Learning Approach," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018, pp. 107-114, doi: 10.1109/ICOEI.2018.8553904.

[5] S. Kim, P. Joshi, P. S. Kalsi and P. Taheri, "Crime Analysis Through Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 415-420, doi: 10.1109/IEMCON.2018.8614828.

[6] J. Kiran and K. Kaishveen., "Prediction Analysis of Crime in India Using a Hybrid Clustering Approach," 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on, 2018, pp. 520-523, doi: 10.1109/I-SMAC.2018.8653744.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY