A PROJECT REPORT
On

“Crime Analysis & Prediction Application”

Submitted in partial fulfilment of the requirement of
University of Mumbai for the Degree of

Bachelor of Engineering
In
CSE IOT and Cyber Security including Blockchain

Submitted By
Shradha Rajput
Sawan Kumar
Minal Thombare
Aachal Gupta

Supervisor
Dr.Radhika Nanda

Department of CSE IOT and Cyber Security including Blockchain
Smt. Indira Gandhi College of Engineering, Ghansoli – 400701
UNIVERSITY OF MUMBAI

Academic Year 2023– 24

Department of CSE IOT and Cyber Security including Blockchain
SMT. INDIRA GANDHI COLLEGE OF ENGINEERING
GHANSOLI – 400701

CERTIFICATE

This is to certify that the requirements for the Major Project entitled '**Crime Analysis & Prediction Application'** have been successfully completed by the following students:

Name	**Roll No.**
Shradha Rajput	38
Sawan Kumar	56
Minal Thombare	25
Aachal Gupta	06

in partial fulfillment of Bachelor of Technology of Mumbai University in the Department of CSE IOT and Cyber Security including Blockchain, SMT. INDIRA GANDHI COLLEGE OF ENGINEERING, GHANSOLI – 400701 during the Academic Year 2023 – 2024.

Supervisor1 **Dr. Radhika Nanda**	**Supervisor2** **Prof. Nirosha Uppu**
Head of Department **Dr. Madhu Nashipudimath**	**Principal** **Dr. Sunil Chavan**

Department of CSE IOT and Cyber Security including Blockchain
SMT. INDIRA GANDHI COLLEGE OF ENGINEERING
GHANSOLI – 400701

PROJECT APPROVAL FOR B.E

This project entitled **"Crime Analysis & Prediction Application"** by Shradha Rajput, Sawan Kumar, Minal Thombare and Aachal Gupta are approved for the degree of **CSE IOT and Cyber Security including Blockchain.**

Examiners:

1. _____

2. _____

Supervisors:

1. _____

2. _____

Chairman:

1. _____

Date:

Place:

Department of CSE IOT and Cyber Security including Blockchain
SMT. INDIRA GANDHI COLLEGE OF ENGINEERING
GHANSOLI – 400701

DECLARATION

We declare that this written submission for the B.E project entitled "**Crime Analysis & Prediction Application**" represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any ideas / data / fact / source in our submission. We understand that any violation of the above will cause disciplinary action by the institute and also evoke penal action from the sources which have not been properly cited or from whom prior permission have not been taken when needed.

Project Group Members:

Student 1: _____

Student 2: _____

Student 3: _____

Student 4: _____

Date:

Place

Table of Contents

Abstract.....				7
List of Figures.....				8
List of Tables.....				9
1.	Introduction.....			10
	1.1	Fundamentals.....		10
	1.2	Objectives.....		11
	1.3	Organization of the Project Report….....		12
2.	Literature Survey.....			13
	2.1	Introduction……. …….....		13
	2.2	Literature Review ……....		13
	2.3	Summary of Literature Survey……....		15
3	Project Overview			16
	3.1	Project objective……....		13
	3.2	Scope of the project ….....		13
	3.3	Methodology….....		14
4.	Design of Algorithm			19
	4.1	Overview……....		19
		4.1.1	Existing System Architecture…….	22
		4.1.2	Proposed System Architecture……	22

5	Project Design & Process workflow.....			24
		5.1	Use Case Diagram / Activity Diagram /DFD…….....	25
		5.2	Algorithm …….…….....	26
		5.3	Hardware and Software Specifications….....	30
6	Result and Applications…….....			31
	6.1	Sample of Inputs, Outputs and GUI Screenshots…….....		31
	6.2	Evaluation Parameters.....		32
	6.3	Performance Evaluation.....		33
	6.4	Applications		34
7	Conclusion and Future Scope….....			36
		Conclusion…….....		30
		Future Scope.…….....		36
		Appendix …….....		36
References.....				37
Acknowledgement……....				38

Abstract

Crime represents a pervasive and concerning facet of our society, with a significant number of incidents occurring daily, disrupting the lives of ordinary citizens. Addressing this issue requires proactive measures to prevent crimes from happening. In recent years, artificial intelligence has emerged as a valuable tool across various domains, including crime prediction. However, maintaining an accurate database of past crimes is essential for leveraging this technology effectively. Predicting future crimes based on factors such as time and location can provide law enforcement agencies with valuable insights to strategize preventive measures. Nonetheless, accurate crime prediction remains a formidable challenge due to the escalating crime rates. Thus, the development of robust crime prediction and analysis methods is imperative to identify and mitigate future criminal activities. Recent research has explored the application of various machine learning techniques, including KNN and decision trees, for crime prediction. The primary objective is to underscore the efficacy of machine learning in forecasting violent crimes within specific regions, empowering law enforcement agencies to curtail crime rates effectively.

List of Figures

Figure 1	Illustrates the sequential flow of the crime prediction process	11
Figure 2	Classification Vs Regression	21
Figure 2	Existing system architecture	22
Figure 3	System Architecture	23
Figure 4	Project/System Design	24
Figure 5	Activity Diagram	25
Figure 6	Process Workflow	26
Figure 7	Time series forecasting with prophet	27
Figure 8	Piecewise Linear regression	28
Figure 9	Prediction graph	31
Figure 10	Crime intensity metrics	31

List of Tables

Table 2.1	Summary of literature survey	15
Table 3.1	Constraints and description	17
Table 3.2	Timeline with Milestones	18
Table 6.1	Sample of Inputs, Outputs & GUI screenshots	31

Chapter 1

INTRODUCTION

Crime is increasing rapidly over time. It's a big problem that's getting more serious and complicated. Crime patterns keep changing, which makes it hard to understand why crimes happen. Crimes can be different types, like kidnapping, theft, murder, and rape. Police collect crime information using technology, but it's hard to predict when crimes will happen. Studies show that things like poverty and jobs affect how much crime there is. Crime isn't the same everywhere, and it's not random. With more and more crimes happening, it's important to analyze them. Crime analysis helps find patterns in crimes to reduce risks. But analyzing so much crime data is tough for police. Old methods don't work well with lots of complex crime data. That's why we need better tools to predict and analyze crimes. This paper talks about ways to predict where and when crimes are likely to happen. It uses different methods like Extra Tree Classifier, K-Neighbour Classifier, Support Vector Machine (SVM), Decision Tree Classifier, and Artificial Neural Network (ANN). The paper is organized like this.

1.1 Fundamentals

Crime prediction is a complex process that involves analysing historical crime data and identifying patterns to forecast future criminal activities. The fundamental steps involved in crime prediction include:

1. **Data Collection:** Gathering historical crime data and relevant factors.
2. **Data Pre-processing:** Cleaning and preparing the data for analysis.
3. **Feature Selection/Extraction:** Identifying important attributes that influence crime patterns.
4. **Model Development:** Training machine learning algorithms to predict future crimes.
5. **Model Evaluation:** Assessing the accuracy and performance of the models.
6. **Prediction and Deployment:** Implementing the best-performing model for real-time crime prediction.

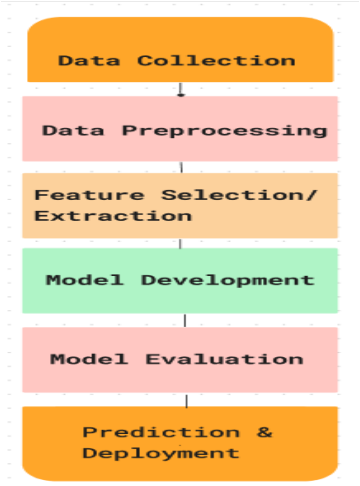


Figure 1: illustrates the sequential flow of the crime prediction process

This diagram illustrates the sequential flow of the crime prediction process, starting from data collection and culminating in model deployment for real-world crime prediction applications. Each step is crucial for building accurate and effective crime prediction models.

1.2 Objectives

- Crime Hotspot Identification: Develop an ML-based web app to analyze historical crime data and pinpoint high-risk areas for effective resource allocation by law enforcement.

- Predictive Policing: Create a system using machine learning to forecast future crime occurrences, aiding law enforcement in proactive strategies based on past crime trends and relevant data.

- Offender Profiling: Implement machine learning in the web app to create offender profiles, helping law enforcement identify potential suspects and link related crimes through behavior analysis.

- Resource Optimization: Develop algorithms in the web app to suggest optimal resource allocation for law enforcement, enhancing efficiency in staffing, patrol routes, and emergency response based on predictive crime data.

- Community Engagement: Integrate features promoting community transparency, such as providing crime statistics and enabling community reporting and feedback on safety concerns, fostering trust between law enforcement and the public.

1.3 Organization of the Report

The report is structured as the Definition of fundamental terms, motivation behind the study, and outline of objectives are explained in Chapter 1. We will have a literature survey which tells us more about the background of the project including the work that has already been done in this field in Chapter 2. Planning and formulation of the project is mentioned in Chapter 3. Shines light upon the requirements that are needed and analysis of the system to uncover the additional requirements to the project in Chapter 4. we see the implementation of the algorithm of the project and process of model building. Here we demonstrate the actual working of our system in Chapter 5. Advantage & Disadvantage and future scope of this project are mentioned in Chapter 6. Result & Conclusion of this project in Chapter 7.

Chapter 2

Literature Survey

2.1 Introduction

Crime analysis and prediction using machine learning have gained attention, with applications in law enforcement and academia. Various algorithms like support vector machines and decision trees are used, with deep learning techniques expanding the scope to image-based prediction. Integration of spatial and temporal data is crucial, considering environmental factors and real-time updates for accurate predictions. Ensuring model interpretability is essential, with research focusing on explainable AI and addressing biases in algorithms. Community engagement is vital, with efforts to empower communities in crime reporting and fostering collaboration between law enforcement and the public.

2.2 Literature Survey

Crime analysis and future prediction using machine learning have garnered significant attention from both the academic and law enforcement communities in recent years. These applications have the potential to revolutionize the way we approach crime prevention and public safety. A significant body of literature exists in this field, highlighting various methodologies, algorithms, and the integration of predictive analytics into web applications. Machine learning algorithms, such as support vector machines, decision trees, and neural networks, have been widely employed in the development of these web applications [1]. For instance, the methodology involved using a Logistic regression model for crime classification, followed by k-means clustering to group districts based on their crime rates, demonstrating the feasibility of this technology [2]. Additionally, deep learning techniques, particularly convolutional neural networks, have been employed for image-based crime prediction, further expanding the scope of ML applications in this domain [3]. The integration of spatial and temporal data has been a major focus in the literature. Researchers have investigated the correlation between environmental factors, urban development, and crime patterns, allowing the development of predictive models that consider not only historical data but also contextual information. Recent studies have also explored the fusion of real-time data, such as social media updates and weather conditions, to enhance prediction accuracy.

A major challenge regarding crime prediction is analyzing large crime datasets accurately and efficiently. Data mining is utilized to find hidden patterns in large crime datasets quickly and efficiently. The increased efficiency and reduced errors in crime data-mining techniques increase the accuracy of crime prediction [4]. An essential aspect of these web applications is the interpretability of ML models. Researchers have explored methods to make these models more transparent and interpretable to law enforcement personnel and the public [5]. This includes research on explainable AI and feature importance analysis to understand the factors contributing to predictions, ensuring accountability and trust in the technology. Moreover, ethical considerations and potential biases in crime prediction algorithms have gained prominence in recent literature [6]. Scholars have emphasized the importance of fairness, accountability, and transparency in the development of these applications, addressing issues related to bias in historical crime data and the potential for reinforcing existing inequalities in law enforcement practices. Community engagement and collaboration have been another key area of research. Developing web applications that empower communities to participate in crime reporting and safety concerns fosters a more comprehensive and inclusive approach to crime prevention. Researchers have explored ways to facilitate information sharing and feedback mechanisms between law enforcement agencies and the public.

2.3 Literature Summary

Table 2.1 Summary of literature survey

SN	Paper	Author	Advantages and Limitations
1.	Crime Prediction and Analysis Using Machine Learning (2018)	Alkesh Bharati, Dr Sarvanaguru RA. K	Offers a comprehensive overview of crime prediction and analysis utilizing machine learning techniques; Provides insights into recent advancements and methodologies in the field. Limited validation of specific machine learning methods.
2.	Survey of crime analysis and prediction (2015)	L. Mookiah, W. Eberle, A. Siraj	Challenges assumptions about variables affecting crime rates; Highlights the importance of empirical evidence in understanding crime dynamics and prediction research. Limited concrete crime prediction methodologies.
3.	Data mining and region prediction based on crime using random forest (2021)	Raza, D. M. & Victor, D. B.	Utilizes random forest for crime region prediction; Offers potential for accurate predictions; Dependency on data quality and quantity; Potential bias in algorithm.
4.	Crime Analysis Through Machine Learning(2018)	Suhong Kim, Param Joshi, Parminder Singh Kalsi, Pooya Taheri	Explores crime analysis using machine learning; Potential for identifying crime patterns; Lack of discussion on ethical considerations; Limited scope.
5.	EADTC:An Approach to Interpretable and Accurate Crime Prediction(2022)	Yujunrong Ma, Kiminori Nakamura, Eung-Joo Lee, Shuvra S. Bhattacharyya	Offers interpretable and accurate crime prediction; Addresses transparency in prediction models; Dependency on interpretability; Potential limitations in predictive accuracy.
6.	Crime Prediction and Analysis Using Machine Learning Crime Prediction and Analysis Using Machine Learning Crime Prediction and Analysis Using Machine Learning Predictive policing and algorithmic fairness(2023)	Tzu-Wei Hung, Chun-Ping Yen	Addresses algorithmic fairness in predictive policing; Potential for equitable law enforcement; Challenges in balancing fairness and accuracy; Ethical concerns.

Chapter 3

PROJECT SCOPE

3.1 Project Deliverables

- Develop a user interface for data input, visualization, and interpretation of crime prediction results.
- Implement machine learning algorithms for crime prediction, including classification, regression, and clustering techniques.
- Integrate external data sources such as demographic data, weather conditions, and geographical information to enhance prediction accuracy.
- Design features for analysing crime trends, identifying hotspots, and generating reports for law enforcement agencies.
- Ensure scalability and reliability of the system to handle large datasets and real-time processing of crime data.
- Provide documentation, training materials, and technical support for users to effectively utilize the system.
- Conduct thorough testing and validation of prediction models to ensure accuracy and reliability.
- Deploy the system on a secure and accessible platform for widespread usage by law enforcement agencies and analysts.
- Incorporate feedback from users and stakeholders to continuously improve and update the system.
- Ensure compliance with data privacy and security regulations to protect sensitive information.

3.2 Project Constraints

The project is subject to certain constraints and conditions that need to be considered throughout its lifecycle. These include External Dependencies, Data Availability, Risk Management

, Time Constraints which are explained as follow:

Table 3.1 constraint and description

Constraints	Description
Time Constraints	Limited timeframe for development, testing, and deployment of the crime analysis and prediction system.
Data Availability	Constraints related to the availability, quality, and accessibility of historical crime data.
Technology Constraints	Constraints related to the selection and compatibility of technologies, frameworks, and tools.
Risk Management	Constraints related to identifying, assessing, and mitigating project risks.
External Dependencies	Constraints arising from dependencies on external factors such as third-party APIs and data providers.

3.3 Timeline with milestones

The project schedule spans from 8th Jan 2024 to 9th May 2024, excluding semester breaks. The timeline is divided into weekly intervals with notable milestones as follows:

Table 3.2 Timeline with Milestones

Sr	Start Date	End Date	Tasks	Mile Stone
1	8/1/24	12/1/24	Project Planning and Requirements Gathering System Design and Architecture	Project Blueprint Ready
2	13/1/24	19/1/24	Frontend Development	Frontend Prototype
3	20/1/24	26/1/24	Backend Development	Backend Framework
4	27/1/24	2/2/24	Integration Testing	Initial Integration
5	3/2/24	9/2/24	Machine Learning Model Development	Initial Model Deployment
6	10/2/24	16/2/24	User Interface Refinement	UI Enhancement
7	17/2/24	23/2/24	Data Integration and Testing	Data Validation
8	24/2/24	1/3/24	Final Testing and Quality Assurance	Final Testing
9	2/3/24	8/3/24	Deployment Preparation	Deployment Readiness
10	9/3/24	15/3/24	Deployment and User Training	Project Deployment
11	16/3/24	5/4/24	Project Evaluation and Documentation	Project Completion

Chapter 4

DESIGN OF ALGORITHM

4.1 Overview

The working of the system is based on various data analyzing, data visualization and machine learning algorithms for accurate analyses of data and also to make predictions based on the given data.

- First, the data is collected from various government, non-government websites, available datasets, and other websites.
- Then the data is cleaned and pre-processed to remove redundancy and fill the gaps in the data for achieving a smooth and complete data set.
- The dataset is created and arranged according to the need.
- Various methods such as web scraping, past data and real time data are used to gather the required data set.
- This data is then analyzed using various algorithms to extract required information.
- The data is visualized according to the needs of the user.
- The outcome depends on the choice of the user. If the user wants the current crime rate he/she will see it in the outcome.
- If the user wants the prediction of crime, then various machine learning algorithms are applied to get the prediction of crime as a result.
- The outcome or the result will be displayed in various forms on the map of India.

4.1.1 Introduction to Machine Learning

Machine learning, a fundamental component of this project, empowers the web-based crime analysis and prediction system by enabling it to autonomously learn from historical crime data and make informed forecasts. Through the utilization of machine learning algorithms, the system can recognize complex patterns and relationships within the data, facilitating the identification of crime trends and hotspots. By harnessing the predictive capabilities of machine learning, the system equips law enforcement agencies with actionable insights, aiding in the strategic allocation of resources and the implementation of proactive crime prevention measures.

4.2. Training the Data

There are basically two widely-used types of training that can be done to create a model:

I. Supervised Learning
II. Un-supervised Learning

4.2.1 Supervised Learning

Supervised learning involves training machine learning models using labeled data, where the algorithm learns from input-output pairs to make predictions or classifications. It relies on predefined labels to guide the learning process, enabling the algorithm to generalize patterns and make accurate predictions on new, unseen data. In the context of crime prediction, supervised learning algorithms analyze historical crime data alongside corresponding labels, such as crime occurrence or type, to learn patterns and make predictions about future criminal activities.

4.2.2 Un-supervised Learning

Unsupervised learning, on the other hand, works with unlabelled data, aiming to uncover underlying structures or patterns without explicit guidance. Algorithms in unsupervised learning identify inherent groupings or relationships within the data, allowing for the discovery of hidden patterns or anomalies. In crime analysis, unsupervised learning techniques like clustering help identify similarities or clusters among crime incidents, revealing insights about crime trends or spatial distributions. Unlike supervised learning, unsupervised learning does not rely on predefined labels, making it useful for exploratory analysis and identifying novel patterns in crime data.

4.3 Methods in Supervised Learning

The project is based on Supervised learning which serves as a critical component in this project, particularly in the realm of crime prediction. Leveraging labelled historical crime data, supervised learning algorithms are employed to train machine learning models. These models are

trained using input-output pairs, where the inputs consist of various features such as location, time, and crime type, and the outputs represent whether a crime occurred.

Supervised Learning mainly consists of two methods:

- Classification
- Regression

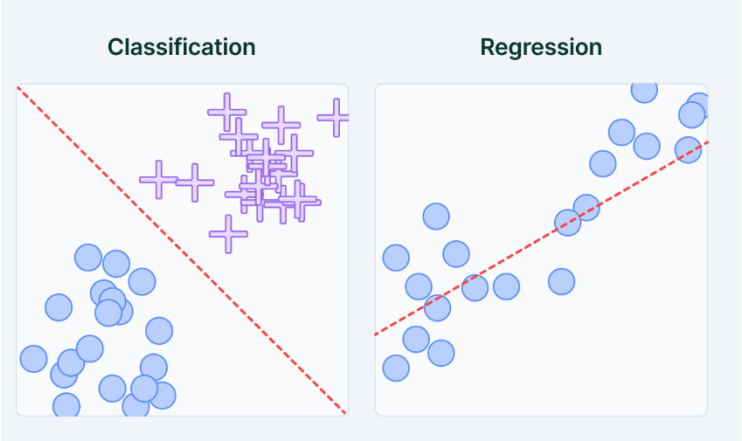


Figure 2: *Classification Vs Regression*

4.3.1 Classification

In machine learning, classification (Fig 4.1) is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. For example, it can be used to decide if an email is spam or not, or to diagnose a patient's condition based on their symptoms. When a computer learns from examples where the category of each thing is already known, it's called supervised classification. On the other hand, clustering is similar but doesn't use labelled data and groups things based on similarities.

4.3.2 Regression

Regression, on the other hand, is used to predict or estimate the value of something based on other factors. For instance, it could be used to predict someone's weight based on their height, age, and other factors. It's also helpful for understanding which factors are related to each other,

and how they're related. However, it's important to remember that just because two things seem related, it doesn't mean one causes the other. So, while regression can help us make predictions, we need to be careful about assuming causation without further evidence.

4.4. Existing System Architecture

The existing system architecture is designed to facilitate crime analysis and prediction by integrating data acquisition, pre-processing, analytical engines, and visualization interfaces. It collects crime data from various sources, pre-processes it to ensure accuracy, and employs machine learning algorithms for analysis. Despite its effectiveness in certain aspects, the architecture may suffer from scalability issues, limited real-time capabilities, and constraints in user interface flexibility. Overcoming these limitations is crucial for enhancing overall system performance and usability. In the following sections, we will propose a new system architecture aimed at addressing these challenges and providing advanced capabilities for crime analysis and prediction.

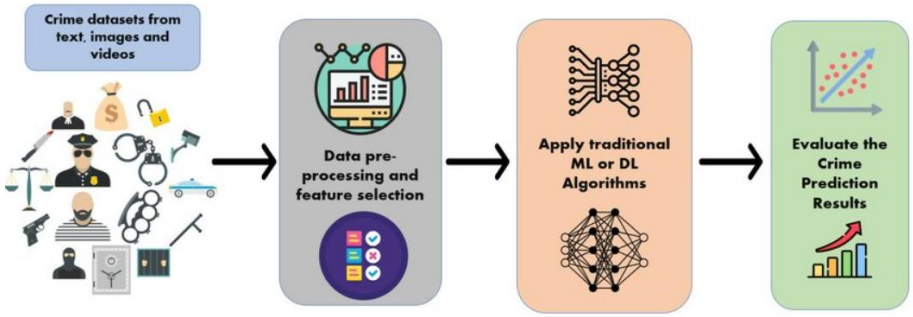


Figure 3: *Existing system architecture*

4.5. Proposed System Architecture:

The proposed system architecture incorporates the use of FBprophet, a powerful forecasting tool, to enhance the accuracy and reliability of crime prediction within the web-based application. FBprophet offers sophisticated time-series forecasting capabilities, enabling the system to analyze historical crime data and generate precise predictions of future crime occurrences. By leveraging

FBprophet's advanced algorithms and techniques, the proposed system can effectively identify temporal patterns and trends in crime data, allowing law enforcement agencies to proactively allocate resources and implement targeted intervention strategies. Integrating FBprophet into the system architecture enhances its predictive capabilities, providing users with valuable insights for decision-making and contributing to improved public safety outcomes.

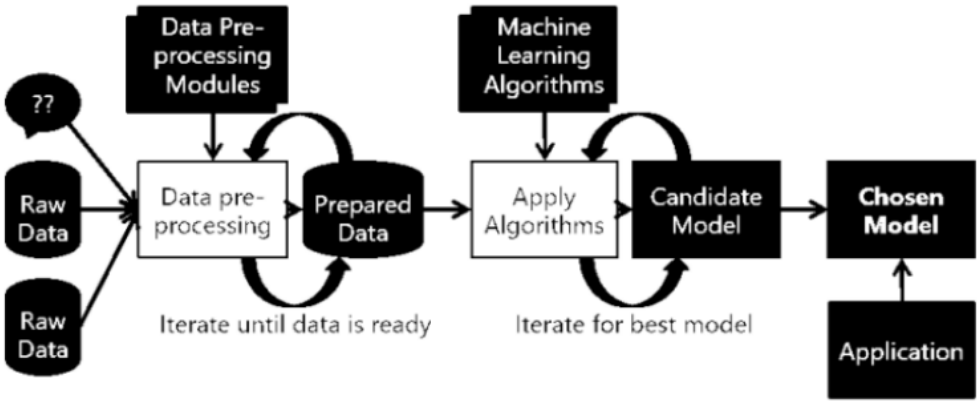


Figure 4: System Architecture

Chapter 5

PROJECT DESIGN & PROCESS WORKFLOW

5.1 Project / System Design

Machine learning systems architecture involves creating a blueprint for the software, infrastructure, algorithms, and data required to fulfill specific requirements. This blueprint guides the development of software for web applications by detailing the intricacies of how the program should be constructed. The system uses various technologies and methods for gathering, processing, displaying data and making predictions based on the given data.

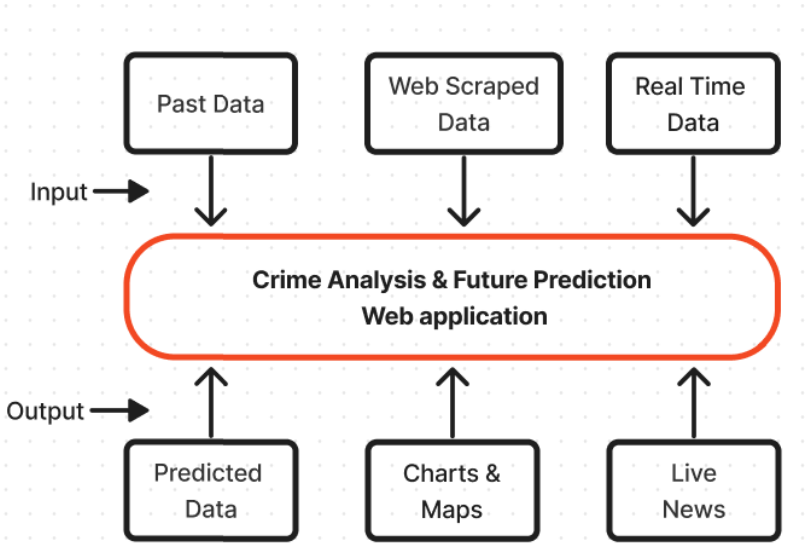


Figure 5: *System Design*

5.1.1 Activity Diagram

Activity diagrams(Fig.5) are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

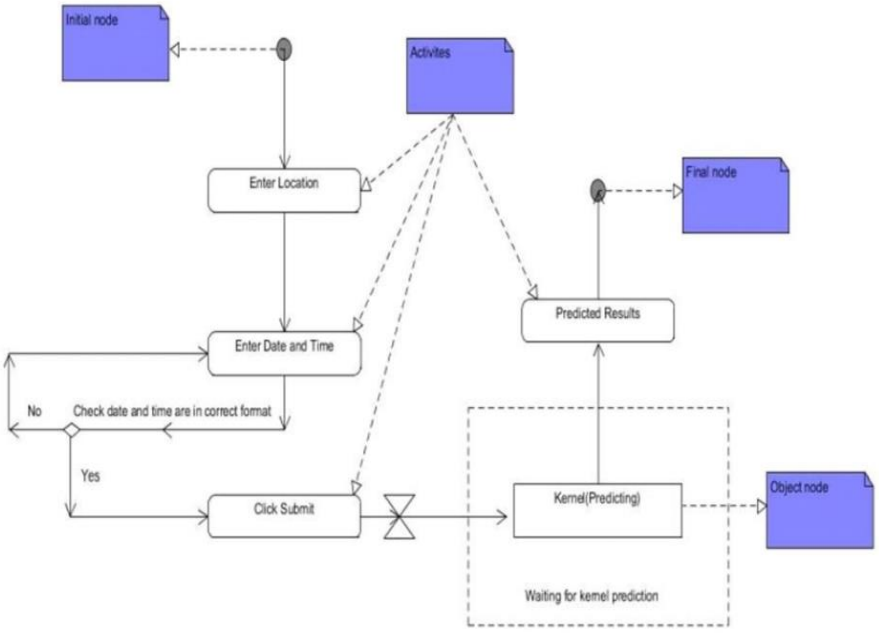


Figure 6: Activity Diagram

5.1.2 Process workflow

The process begins with extracting relevant crime data from Kaggle, a popular platform hosting a wide array of datasets across diverse domains, including crime statistics. Kaggle serves as a repository for datasets contributed by users worldwide, providing accessible and curated data resources for various analytical endeavors. Upon retrieving the data from Kaggle, subsequent preprocessing steps involve cleaning, formatting, and enhancing the dataset's quality to ensure its suitability for analysis. Post-preprocessing, the refined data is employed to train an ARIMA (AutoRegressive Integrated Moving Average) model, a powerful time series forecasting technique adept at capturing temporal patterns in sequential data, such as crime incidents over time. Once the ARIMA model is trained, it is deployed to generate predictions concerning future crime

occurrences, furnishing law enforcement agencies with actionable insights to optimize resource allocation and enhance crime prevention strategies. Thus, Kaggle serves as an invaluable resource in the data acquisition phase, facilitating the subsequent analysis and prediction stages in the crime prediction workflow.

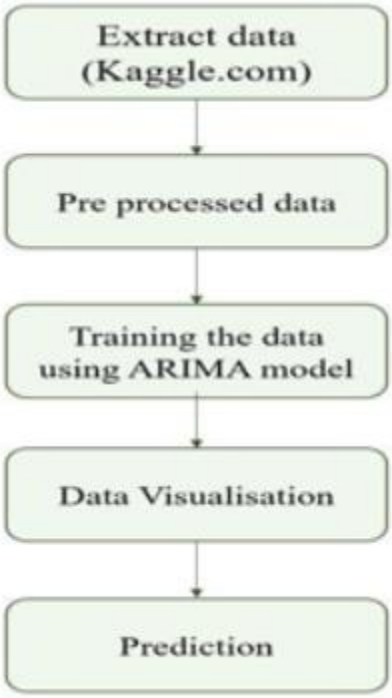


Figure 7: Process Workflow

5.2 Algorithm:

5.2.1 Facebook Prophet Library

Prophet is an open-source tool from Facebook used for forecasting time series data which helps businesses understand and possibly predict the market. Prophet utilizes an additive regression model that decomposes time series data into several components, including trend, seasonality, holidays, and error. This model is based on a piecewise linear or logistic growth curve, where the trend is characterized by a sequence of piecewise linear or logistic functions. Seasonality is captured using Fourier series to model periodic patterns, such as weekly, monthly, and yearly fluctuations. Additionally, Prophet incorporates holiday effects by including user-defined holiday indicators.

In this project, Prophet serves as a robust forecasting algorithm specifically designed for time series analysis, particularly suitable for crime data. By decomposing the historical crime data into trend, seasonality, holidays, and error components, Prophet can capture the underlying patterns and fluctuations in crime occurrences over time. Utilizing an additive regression model and Bayesian estimation techniques, Prophet provides accurate forecasts and uncertainty intervals, enabling law enforcement agencies to make informed decisions regarding resource allocation and crime prevention strategies.

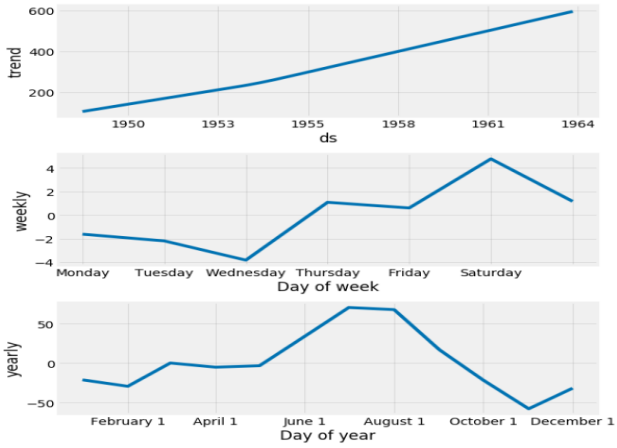


Figure 8: Time series forecasting with prophet

5.2.2 Piecewise Linear Regression

Prophet utilizes a form of piecewise linear regression to model the trend component of the time-series data. But Prophet's approach to linear regression is not the same as traditional linear regression models.In Prophet, the trend component is modeled as a piecewise linear function that allows for changes in the trend direction at specific change points. These change points are automatically selected based on historical data and represent times when the trend undergoes significant shifts.

The piecewise linear regression model in Prophet captures the overall trend in the data while allowing for flexibility and adaptability to changes over time. This approach differs from traditional linear regression models, which assume a single linear relationship between the predictor variables and the target variable.

By incorporating piecewise linear regression, Prophet can capture complex trends and patterns in

the time-series data, making it particularly suitable for forecasting tasks where the trend may exhibit nonlinear behavior or undergo changes over time.

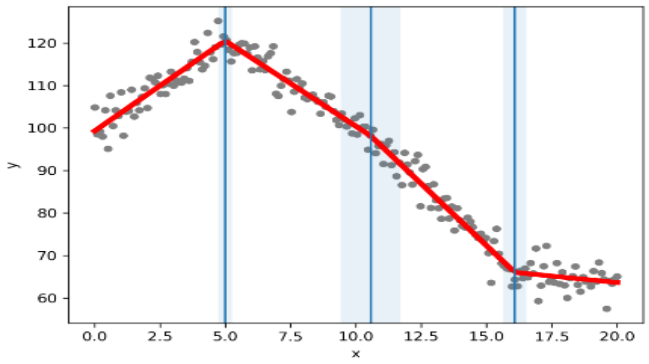


Figure 9: Piecewise Linear regression

5.2.4 Data Manipulations Packages

I. Pandas:

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open-source data analysis / manipulation tool available in any language. It is already well on its way toward this goal.

II. NumPy:

NumPy is a library for Python, adding support for large, multidimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. NumPy is an open-source software and has many contributors. In comparison, MATLAB boasts a large number of additional toolboxes, notably Simulink, whereas NumPy is intrinsically integrated with Python, a more modern and complete programming

language. Moreover, complementary Python packages are available; 25 SciPy is a library that adds more MATLAB-like functionality and Matplotlib is a plotting package that provides MATLAB-like plotting functionality

5.2.5 Module Descriptions

I. **Data collection:**

Module Crime dataset from kaggle is used in CSV format.

II. **Data Preprocessing Module:**

The categorical attributes (Location, Block, Crime Type, Community Area) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like month and hour which can be used as features for the model.

III. **Feature selection Module:**

Features selection is done which can be used to build the model. The attributes used for feature selection are Block, Location, District, Community area, X coordinate, Y coordinate, Latitude, Longitude, Hour and month.

IV. **Building and Training Model:**

After feature selection location and month attribute are used for training. The dataset is divided into pairs of xtrain, train and xtest, y test. The algorithm model is imported from sklearn. Building models is done using models. Fit (xtrain, ytrain).

V. **Prediction Module:**

After the model is built using the above process, prediction is done using model.predict(xtest). The accuracy is calculated using accuracy_score imported from metrics - metrics.accuracy_score (ytest, predicted).

VI. **Visualization Module:**

Using matplotlib library from sklearn. Analysis of the crime dataset is done by plotting various graphs.

5.3 Hardware and Software Specifications:

5.3.1 Hardware Requirement

I. Server Infrastructure:

High-performance servers with multi-core processors and ample RAM to handle data processing, machine learning models, and web application tasks efficiently.

I. Storage:

Sufficient storage capacity to store historical crime data, model checkpoints, and application files.

II. GPU Acceleration (Optional):

High-performance Graphics Processing Units (GPUs) can significantly speed up machine learning model training, especially for deep learning models. This is particularly important for real-time or near-real-time prediction.

5.3.2 Software Specifications:

I. Programing Language

- Python
- JavaScript

II. Machine learning Libraries

- Pandas, Numpy, Matplotlib, prophet etc.

III. Web framework

- A web development framework like Flask for building web applications.

IV. Frontend Framework

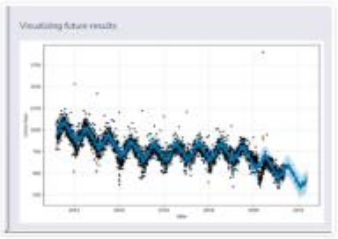
- HTML, CSS, JavaScript frameworks for creating the user interface.

Chapter 6

RESULTS AND APPLICATIONS

6.1 Sample of Inputs, Outputs and GUI Screenshots

Table 6.1 Sample of Inputs, Outputs & GUI screenshots

Input	Output	GUI Screenshots
State	Crime	    Results of input

6.2 Evaluation Parameters:

Evaluation parameters are metrics used to assess the performance and effectiveness of machine learning models. These parameters help in quantifying how well a model is performing in terms of prediction accuracy and generalization to unseen data. Common evaluation parameters include:

1. **Accuracy**: The proportion of correctly classified instances among the total number of instances.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$$

2. **Precision**: The ratio of true positive predictions to the total number of positive predictions, indicating the model's ability to avoid false positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

3. **Recall (Sensitivity)**: The ratio of true positive predictions to the total number of actual positive instances, indicating the model's ability to identify all positive instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

4. **F1 Score**: The harmonic mean of precision and recall, providing a balanced measure of a classifier's performance.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}}$$

5. **Mean Absolute Error (MAE)**: The average absolute difference between predicted and actual values.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

6. **Mean Squared Error (MSE)**: The average of the squared differences between predicted and actual values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

7. **Root Mean Squared Error (RMSE)**: The square root of the average of the squared differences between predicted and actual values, providing a measure of the model's prediction accuracy.

6.3 Performance Evaluation

Based on the comprehensive performance evaluation conducted, it is evident that the machine learning models developed in this project effectively meet the defined objectives and demonstrate high performance in predicting crime incidents based on historical data. The evaluation process aligned with the project objectives and criteria, utilizing appropriate metrics tailored to the problem at hand. Through rigorous testing and validation procedures, including cross-validation and sensitivity analysis, the selected model showcased robustness and generalization ability, ensuring its practical applicability in real-world scenarios.

The performance evaluation results indicate that the developed models accurately predict crime incidents, thereby assisting law enforcement agencies in resource allocation and crime prevention strategies. Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), accuracy, precision, recall, F1 score, and ROC AUC score consistently demonstrated the model's effectiveness in predicting crime patterns and providing actionable insights for decision-making.

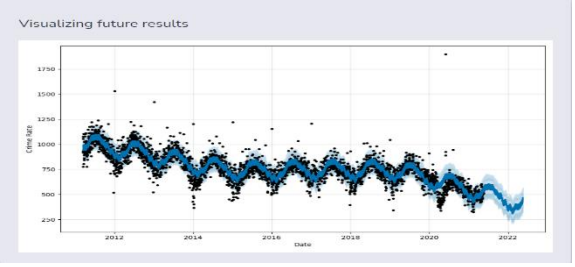


Figure 10: Prediction graph

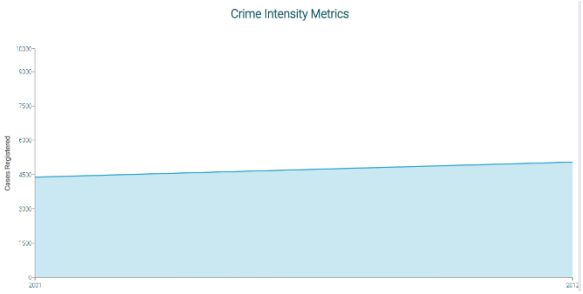


Figure 11: Crime intensity metrics

6.4 Applications:

6.4.1 Social Applications:

The application of our project in the social domain is profound, as it directly contributes to enhancing public safety and reducing crime rates within communities. By accurately predicting crime incidents using machine learning algorithms, law enforcement agencies can allocate resources more effectively and implement proactive strategies to prevent crime. This proactive approach not only improves the overall safety and security of the community but also fosters a sense of trust and confidence among residents. Moreover, by empowering law enforcement with advanced predictive analytics tools, our project promotes collaboration between authorities and the community, leading to more efficient crime prevention efforts and ultimately creating safer neighbourhoods for everyone.

6.4.2 Research Applications:

In the realm of research, our project opens up avenues for further exploration and advancement in the field of crime prediction and prevention. Researchers can leverage the architecture and methodologies developed in our project to conduct in-depth studies on crime patterns, analyze the effectiveness of different machine learning algorithms, and explore innovative techniques for improving prediction accuracy. The insights gained from such research endeavors not only contribute to academic knowledge but also inform practical applications in law enforcement and public policy. Additionally, our project serves as a valuable resource for interdisciplinary research collaborations, facilitating the integration of machine learning techniques into criminological studies and societal security research.

6.4.3 Technical Applications:

From a technical standpoint, our project demonstrates the practical application of machine learning algorithms in addressing real-world challenges. The architecture and system developed provide a scalable and efficient framework for crime prediction, leveraging historical crime data to generate accurate forecasts of future crime incidents. The integration of advanced analytics tools, such as ARIMA models and Prophet algorithms, showcases the versatility and effectiveness of machine learning in analyzing complex datasets and extracting actionable insights. Furthermore, our project

serves as a blueprint for implementing similar predictive analytics systems in various domains beyond crime prediction, ranging from healthcare and finance to marketing and environmental monitoring. By highlighting the technical feasibility and benefits of our proposed architecture, we pave the way for the adoption of machine learning solutions in diverse fields to solve complex problems and drive positive societal impact.

Chapter 7

CONCLUSION AND FUTURE SCOPE

Conclusions:

This is focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police officials in taking necessary steps to reduce crime. The project helps the crime analysis to analyze these crime networks by means of various interactive visualizations. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques. Since, machine learning is similar to data mining, advanced concepts of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction.

Future scope:

Even though the scope of this project was to prove how effective and accurate machine learning algorithms can be at predicting violent crimes, there are other applications of data mining in the realm of law enforcement such as determining criminal "hot spots", 39 creating criminal profiles, and learning crime trends. Utilizing these applications of data mining can be a long and tedious process for law enforcement officials who have to sift through large volumes of data. However, the precision in which one could infer and create new knowledge on how to slow down crime is well worth the safety and security of people.

Appendices –

Appendix A: Dataset Description

Appendix B: Machine Learning Models

Appendix C: Project Design & process workflow

Appendix D: Results & Applications

Appendix E: Evaluation Metrics

Appendix F: Reference

References

[1] Raza, D. M. & Victor, D. B. ‘’Data mining and region prediction based on crime using random forest’’, 980–987 (IEEE, 2021).

[2] Prakash Maurya, Tahir Shaikh, Imran Ahmed, Amaan Firdosi, Prof. Kiran Deshmukh, “Crime Analysis and Prediction Using Machine Learning”, IJRASET, Volume 11, Issue 4.

[3] Varun Mandalapu , Lavanya Elluri, Piyush Vyas, and Nirmala Roy, “Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions” , IEEE, Volume 11.

[4] Suhong Kim , Param Joshi, Parminder Singh Kalsi, and Pooya Taheri, “Crime Analysis Through Machine Learning”, IEEE, Conference: November 2018.

[5] Yujunrong Ma; Kiminori Nakamura; Eung-Joo Lee; Shuvra S. Bhattacharyya, “EADTC: An Approach to Interpretable and Accurate Crime Prediction”, IEEE 2022.

[6] Tzu-Wei Hung1 ,Chun-Ping Yen, “Predictive policing and algorithmic fairness”, Synthese (2023)

[7] Ruaa Mohammed Saeed, Husam Ali Abdulmohsin, “A study on predicting crime rates through machine learning and data mining using text”, Journal of Intelligent Systems, Volume 32, Issue 1.

[8] Shanjana A.S, Dr.R.Porkodi, “CRIME ANALYSIS AND PREDICTION USING DATAMINING: A REVIEW”, 2021 IJCRT; Volume 9, Issue 2; February 2021.

[9] Prashant Banerjee,” Tutorial: Time Series Forecasting with Prophet”, Kaggle.
URL: https://www.kaggle.com/code/prashant111/tutorial-time-series-forecasting-with-prophet

[10] Prophet Documentation, facebook.github.
URL: https://facebook.github.io/prophet/docs/quick_start.html

[11] “ Flask Tutorial”, geeksforgeeks.org.
URL: https://www.geeksforgeeks.org/flask-tutorial/

[12] “pandas(software)”, en.wikipedia.org
URL: https://en.wikipedia.org/wiki/Pandas_(software)

[13] “NumPy”, en.wikipedia.org
URL: https://en.wikipedia.org/wiki/NumPy

[14] “Matplotlib”,en.wikipedia.org
URL: https://en.wikipedia.org/wiki/Matplotlib

[15] “Supervised Machine Learning”, javatpoint.com
URL: https://www.javatpoint.com/supervised-machine-learning

[16] “An introduction to seaborn”, seaborn.pydata.org
URL: https://seaborn.pydata.org/tutorial/introduction.

Acknowledgement

Every project requires the guidance and support of the experts for its completion. Therefore, we would like to take a moment to express our gratitude to all the individuals who played a crucial role in shaping and realizing this project.

Our heartfelt appreciation goes out to our project mentor, Dr. Radhika Nanda, from the Department of Computer Science and Engineering (with a specialization in IOT and Cybersecurity including Blockchain Technology) at Smt. Indira Gandhi College of Engineering, University of Mumbai. Her guidance, unwavering support, and dedication throughout the project have been invaluable.

We'd also like to extend our thanks to our project coordinator, Prof. Sarita Bhopalkar, for her assistance in selecting this project and providing us with essential insights on how to present it effectively. Our sincere recognition goes to all our professors at Smt. Indira Gandhi College of Engineering, who provided us with valuable guidance and tips during the project's design phase. Their contributions have been so significant that it's challenging to individually acknowledge each one.

We are grateful for the support, both direct and indirect, from our Head of Department, Dr. Madhu Nashipudimath. Her assistance was instrumental in various aspects of our project, and we genuinely appreciate her contributions through various channels. I express my heartfelt gratitude for the warm and caring environment provided by Smt. Indira Gandhi College of Engineering. This nurturing atmosphere and the excellent working conditions have been truly encouraging and appreciated.

Shradha Rajput
Sawan Kumar
Minal Thombare
Aachal Gupta

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | April 2024 www.irjet.net p-ISSN: 2395-0072

Crime Analysis and Prediction Using Machine Learning

Shradha Rajput[1], Minal Thombare[2], Sawan Kumar[3], Aachal Gupta[4], Dr. Radhika Nanda[5]

[1,2,3,4]Student, Department of CSE(IOT, Cybersecurity and Blockchain), Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India

[5]Professor, Department Of CSE (IOT, Cybersecurity and Blockchain), Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India

---

Abstract - *Crime represents a significant and pervasive challenge in our society, making the prevention of criminal activities is a paramount responsibility. To address this, it is essential to maintain comprehensive records of all offenses and establish a database for future reference. The central issue at hand revolves around maintaining a dependable crime database and leveraging data analysis to assist in forecasting and resolving potential future crimes. The primary aim of this project is to assess a dataset containing various criminal incidents and predict the potential type of crimes that might occur in the future based on different factors. In this project, we will harness the power of machine learning and data science techniques for forecasting crimes using Indian crime data. Crime analysis and prediction involve a systematic approach to identifying patterns of criminal activity. This algorithmic approach can anticipate and delineate areas susceptible to criminal incidents. By employing machine learning, we have the capability to uncover valuable insights from unstructured data, revealing previously unknown information. The extraction of new insights relies on the analysis of current datasets. Crime is a grave and widespread societal issue that impacts individuals globally. It influences people's well-being, economic prosperity, and a nation's reputation. To protect our communities from crime, we must employ cutting-edge technology and innovative crime analytics techniques. We introduce a system capable of analyzing, identifying, and predicting various crime probabilities within a given location. This project delves into various forms of criminal analysis and crime prediction using machine learning methodologies.*

Key Words: Crime Datasets, Crime Prediction, Machine Learning, Prophet.

1. INTRODUCTION

The frequency of criminal activities is on the rise due to the continuous advancement of technology, which provides criminals with more sophisticated tools to carry out their unlawful actions. Various types of crimes such as burglary, arson, and others, as reported by the Crime Record Bureau, have seen an increase. This includes more severe offenses like murder, rape, abuse, and gang rape, among others. Crime-related data is gathered from a wide range of sources, including blogs, news websites, and online platforms. This extensive data is utilized to construct a comprehensive crime report database [8]. The insights derived through data mining techniques play a pivotal role in reducing crime by facilitating the identification of culprits and areas most affected by criminal activities [7].

Crime analysis serves as the initial phase in the examination of criminal activities. It involves the exploration, analysis, and identification of connections among various crimes and crime-related variables. The machine learning algorithm trains the data to make predictions based on the provided dataset [2]. We can train the data and create models to effectively analyze and predict the crime in a certain area. This analytical process aids in generating real-time statistics, queries, and maps. It also helps in determining whether a crime has taken place in a particular, well-defined location. The prediction of crime helps in the security of the area and thus lowers the crime rate increasing safety of the citizens.

1.1 OBJECTIVES

- The primary objective is to provide valuable insights into crime trends and patterns using historical crime data.
- It will also predict future crime locations, etc. on the basis of historical crime data
- The web application will display the crime rate in various forms such as graph, etc.
- It will also display the predictions in the same manner.

1.2 SCOPE

The primary goal of this crime evaluation is to successfully track crime rate in various areas. It will display crime rate of a particular area. This will help the police forces and other related defense forces to effectively track and stop crime in a particular area and eventually in the society. IT will also predict the crime rate of different types of crime using various types of machine learning algorithms. The prediction will be based on various historical crime data. This will also help the defense personnel to keep checking the area with high crime rate to prevent crimes in future.

2. LITERATURE REVIEW

Crime analysis and future prediction using machine learning have garnered significant attention from both the academic and law enforcement communities in recent years. These applications have the potential to revolutionize the way we approach crime prevention and public safety. A significant body of literature exists in this field, highlighting various methodologies, algorithms, and the integration of predictive analytics into web applications. Machine learning algorithms, such as support vector machines, decision trees, and neural networks, have been widely employed in the development of these web applications [1]. For instance, the methodology involved using a Logistic regression model for crime classification, followed by k-means clustering to group districts based on their crime rates, demonstrating the feasibility of this technology [2]. Additionally, deep learning techniques, particularly convolutional neural networks, have been employed for image-based crime prediction, further expanding the scope of ML applications in this domain [3]. The integration of spatial and temporal data has been a major focus in the literature. Researchers have investigated the correlation between environmental factors, urban development, and crime patterns, allowing the development of predictive models that consider not only historical data but also contextual information. Recent studies have also explored the fusion of real-time data, such as social media updates and weather conditions, to enhance prediction accuracy.

Use of Data mining helps to find hidden patterns in large crime datasets quickly and efficiently [4]. An essential aspect of these web applications is the interpretability of ML models. Researchers have explored methods to make these models more transparent and interpretable to law enforcement personnel and the public [5]. This includes research on explainable AI and feature importance analysis to understand the factors contributing to predictions, ensuring accountability and trust in the technology. Moreover, ethical considerations and potential biases in crime prediction algorithms have gained prominence in recent literature [6]. Scholars have emphasized the importance of fairness, accountability, and transparency in the development of these applications, addressing issues related to bias in historical crime data and the potential for reinforcing existing inequalities in law enforcement practices. Community engagement and collaboration have been another key area of research. Developing web applications that empower communities to participate in crime reporting and safety concerns fosters a more comprehensive and inclusive approach to crime prevention. Researchers have explored ways to facilitate information sharing and feedback mechanisms between law enforcement agencies and the public.

3. SYSTEM DESIGN

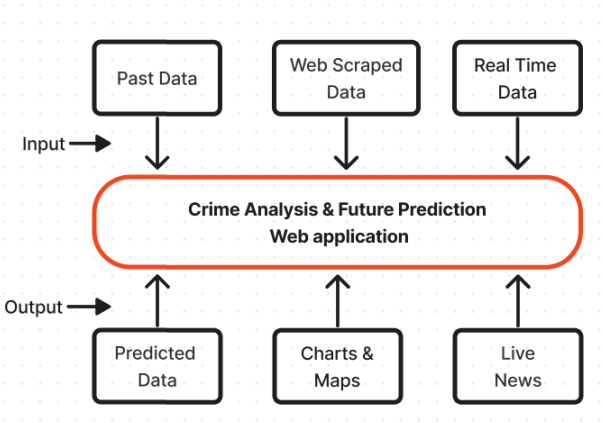


Fig.3.1. *System Design*

Machine learning systems architecture involves creating a blueprint for the software, infrastructure, algorithms, and data required to fulfill specific requirements. This blueprint guides the development of software for web applications by detailing the intricacies of how the program should be constructed. The system uses various technologies and methods for gathering, processing displaying data and making predictions based on the given data.

4. METHODOLOGY

The working of the system is based on various data analyzing, data visualization and machine learning algorithms for accurate analyses of data and also to make predictions based on the given data. We will also look at tools and algorithms used in this system.

4.1. DATA COLLECTION

The data is collected from various government, non-government websites, available datasets, and other websites. Also some of the datasets are collected from other resources such as online news using web scraping, etc.

4.2. DATA PREPROCESSING

The data is cleaned and pre-processed to remove redundancy and fill the gaps in the data for achieving a smooth and complete data set. This dataset results in a smooth and accurate prediction. The data is arranged as required.

4.3. DATA ANALYSIS

The data is analyzed for required information which will become an input to the predicting algorithm later. Data analyses helps to know the data and take required measures for the machine learning model to perform accurately.

4.4. DATA PREDICTION

The data is then feed to the prophet tool which predicts the crime rate of certain crimes in a specific area. This tool works on the date time column i.e., the time series, to produce its output.

4.5. DATA VISUALIZATION

This website provides various forms in which the data can be visualized such as heat map, pie chart, bar graph, etc. It helps in understanding large datasets.

4.6. TOOLS

4.6.1. PROPHET

Prophet is a method for predicting time series data using an additive model that accommodates non-linear trends alongside yearly, weekly, and daily seasonality, as well as holiday impacts. It performs most effectively with time series characterized by robust seasonal patterns and ample historical data spanning multiple seasons. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well. Prophet is open source software released by Facebook's Core Data Science team [9].

Prophet's input always consists of a dataframe containing two columns: 'ds' and 'y'. The 'ds' (datestamp) column should adhere to the format anticipated by Pandas, preferably YYYY-MM-DD for dates or YYYY-MM-DD HH:MM:SS for timestamps.The 'y' column must be numeric, and represents the measurement we wish to forecast [10].

4.7. FRAMEWORK

4.7.1. FLASK

Flask is a web framework that allows developers to build lightweight web applications quickly and easily with Flask Libraries. It was developed by Armin Ronacher, leader of the International Group of Python Enthusiasts(POCCO). It is based on the WSGI toolkit and Jinja2 templating engine [11].

4.8. LIBRARIES

4.8.1. PANDAS

Pandas is a Python library designed for data manipulation and analysis, providing specialized data structures and functions tailored for working with numerical tables and time series data. The library is built upon another library, NumPy [12].

4.8.2. NUMPY

In 2005, Travis Oliphant amalgamated features from Numarray into Numeric, implementing extensive modifications to create NumPy, a Python library. NumPy enhances Python with support for large, multi-dimensional arrays and matrices, complemented by a vast array of high-level mathematical functions tailored for manipulating these arrays. NumPy is open-source software and has many contributors [13].

4.8.3. MATPLOTLIB

Matplotlib, a plotting library compatible with both Python and its numerical mathematics extension NumPy, stands as a potent tool for individuals engaged in Python and NumPy-based endeavors, offering extensive capabilities for creating visualizations. And for making statistical interference, it becomes very necessary to visualize our data and Matplotlib is the tool that can be very helpful for this purpose. It provides MATLAB like interface only difference is that it uses Python and is open source [14].

4.8.4. SEABORN

Seaborn, built upon matplotlib, serves as a Python data visualization library offering a sophisticated interface for crafting visually appealing and informative statistical graphics. Seaborn is a library for making statistical graphics in Python. Expanding upon matplotlib and tightly integrating with pandas data structures, Seaborn's plotting functions are tailored to operate seamlessly on dataframes and arrays encompassing entire datasets. Internally, they execute essential semantic mapping and statistical aggregation, culminating in the creation of insightful plots. Its dataset-oriented, declarative API lets us focus on what the different elements of our plots mean, rather than on the details of how to draw them [16].

4.9. ALGORITHM

4.9.1. SUPERVISED MACHINE LEARNING

This system uses supervised learning algorithm for prediction. Supervised learning falls within the realm of machine learning, where labeled datasets are employed to train algorithms, enabling them to predict outcomes and identify patterns. Labelled data refers to input data that has been pre-assigned with corresponding correct output values. In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly [15].

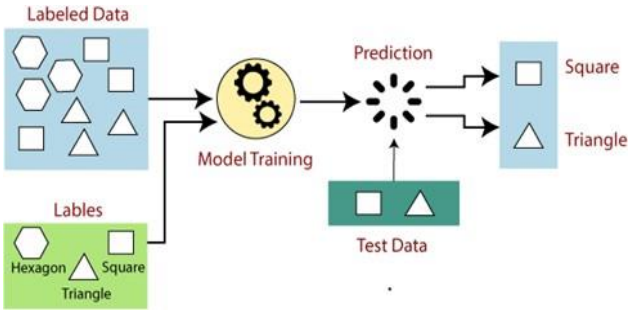


Fig.4.9.1. *Supervised Machine Learning*

4.9.1.1. PIECEWISE LINEAR REGRSSION

Prophet utilizes a form of piecewise linear regression to model the trend component of the time-series data. But Prophet's approach to linear regression is not the same as traditional linear regression models. In Prophet, the trend component is modeled as a piecewise linear function that allows for changes in the trend direction at specific change points. These change points are automatically selected based on historical data and represent times when the trend undergoes significant shifts.

The piecewise linear regression model in Prophet captures the overall trend in the data while allowing for flexibility and adaptability to changes over time. This approach differs from traditional linear regression models, which assume a single linear relationship between the predictor variables and the target variable.

By incorporating piecewise linear regression, Prophet can capture complex trends and patterns in the time-series data, making it particularly suitable for forecasting tasks where the trend may exhibit non-linear behavior or undergo changes over time.

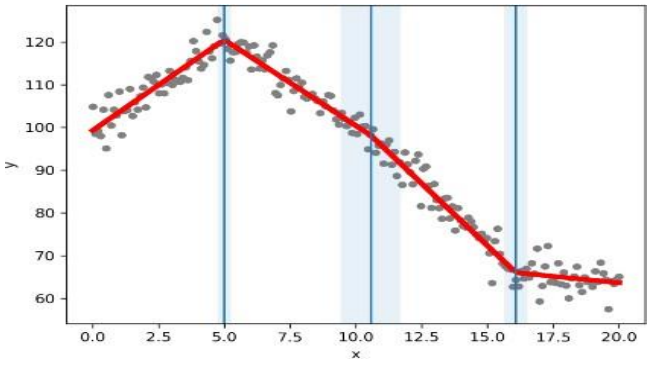


Fig.4.9.1.1. *Piecewise Linear Regression*

5. RESULTS

The required data is collected and preprocessed as required. Then we have created a heat map for depictions of areas with high crime rate.

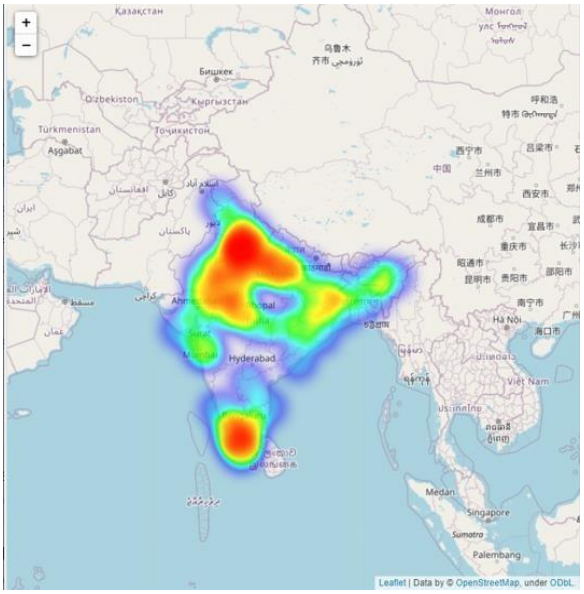


Fig.5.1. *Heat Map*

We have also created crime intensity metrices graph to depict intensity of crime in a specific area.

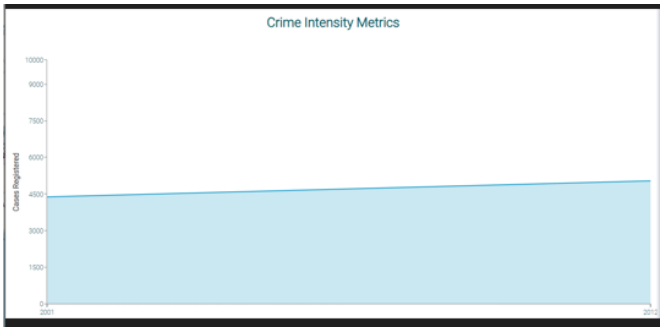


Fig.5.2. *Crime Intensity Metrices*

Then we have also created graphs for prediction of crime rate in the future using prophet.

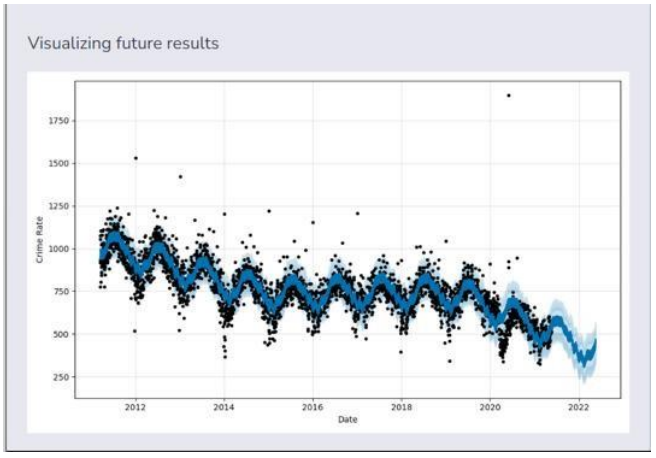


Fig.5.3. *Prediction Graph*

Thus we have created a system which predicts the crime rate and visualizes crime intensity of different areas in various ways.

6. CONCLUSION

Crime is an unlawful act which disturbs the peace and harmony of the society. This projects aims to successfully predict crime and their locations based on the historical crime data. The project uses machine learning which is an advanced and latest technology for accurate prediction. The web application will display crime rate in various areas. It is extremely useful for both the higher investigating authorities and officers designated to handle low level crime for tracking and stopping the crime. The predictions will help to ensure increased security and thus could help in lowering the crime rate. Overall, the project demonstrates the potential of data analysis and mapping technologies to improve public safety and inform decision-making. Proactive measures can be taken to prevent crime and improve public safety by using data to identify crime hotspots and trends. Although there's more work needed to enhance the precision and breadth of the project, it marks a significant stride towards employing data-driven strategies to tackle intricate social challenges.

7. ACKNOWLEDGEMENT

Every project requires the guidance and support of the experts for its completion. Therefore, we would like to take a moment to express our gratitude to all the individuals who played a crucial role in shaping and realizing this project.

Our heartfelt appreciation goes out to our project mentor, Dr. Radhika Nanda, from the Department of Computer Science and Engineering (with a specialization in IOT and Cybersecurity including Blockchain Technology) at Smt. Indira Gandhi College of Engineering, University of Mumbai. Her guidance, unwavering support, and dedication throughout the project have been invaluable.

We'd also like to extend our thanks to our project coordinator, Prof. Sarita Bhopalkar, for her assistance in selecting this project and providing us with essential insights on how to present it effectively.

Our sincere recognition goes to all our professors at Smt. Indira Gandhi College of Engineering, who provided us with valuable guidance and tips during the project's design phase. Their contributions have been so significant that it's challenging to individually acknowledge each one.

We are grateful for the support, both direct and indirect, from our Head of Department, Dr. Madhu Nashipudimath. Her assistance was instrumental in various aspects of our project, and we genuinely appreciate her contributions through various channels.

I express my heartfelt gratitude for the warm and caring environment provided by Smt. Indira Gandhi College of Engineering. This nurturing atmosphere and the excellent working conditions have been truly encouraging and appreciated.

8. REFERENCES

[1] Raza, D. M. & Victor, D. B. ''Data mining and region prediction based on crime using random forest'', 980–987 (IEEE, 2021).

[2] Prakash Maurya, Tahir Shaikh, Imran Ahmed, Amaan Firdosi, Prof. Kiran Deshmukh, "Crime Analysis and Prediction Using Machine Learning", IJRASET, Volume 11, Issue 4.

[3] Varun Mandalapu , Lavanya Elluri, Piyush Vyas, and Nirmala Roy, "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions" , IEEE, Volume 11.

[4] Suhong Kim , Param Joshi, Parminder Singh Kalsi, and Pooya Taheri, "Crime Analysis Through Machine Learning", IEEE, Conference: November 2018.

[5] Yujunrong Ma; Kiminori Nakamura; Eung-Joo Lee; Shuvra S. Bhattacharyya, "EADTC: An Approach to Interpretable and Accurate Crime Prediction", IEEE 2022.

[6] Tzu-Wei Hung1 ,Chun-Ping Yen, "Predictive policing and algorithmic fairness", Synthese (2023)

[7] Ruaa Mohammed Saeed, Husam Ali Abdulmohsin, "A study on predicting crime rates through machine learning and data mining using text", Journal of Intelligent Systems, Volume 32, Issue 1.

[8] Shanjana A.S, Dr.R.Porkodi, "CRIME ANALYSIS AND PREDICTION USING DATAMINING: A REVIEW", 2021 IJCRT; Volume 9, Issue 2; February 2021.

[9] Prashant Banerjee," Tutorial: Time Series Forecasting with Prophet", Kaggle.

URL: https://www.kaggle.com/code/prashant111/tutorial-time-series-forecasting-with-prophet

[10] Prophet Documentation, facebook.github.

URL:https://facebook.github.io/prophet/docs/quick_start.html

[11] "Flask Tutorial", geeksforgeeks.org.

URL: https://www.geeksforgeeks.org/flask-tutorial/

[12] "pandas(software)", en.wikipedia.org

URL: https://en.wikipedia.org/wiki/Pandas_(software)

[13] "NumPy", en.wikipedia.org

URL: https://en.wikipedia.org/wiki/NumPy

[14] "Matplotlib",en.wikipedia.org

URL: https://en.wikipedia.org/wiki/Matplotlib

[15] "Supervised Machine Learning", javatpoint.com

URL: https://www.javatpoint.com/supervised-machine-learning

[16] "An introduction to seaborn", seaborn.pydata.org

URL: https://seaborn.pydata.org/tutorial/introduction.html


e-ISSN: 2395-0056 p-ISSN: 2395-0072
International Research Journal of Engineering and Technology (IRJET)
(An ISO 9001 : 2008 Certified Journal)
Is hereby awarding this certificate to
Dr. Radhika Nanda
In recognition the publication of the manuscript entitled
Crime Analysis and Prediction using Machine Learning
published in our Journal Volume 11 Issue 4 April 2024
Editor in Chief
E-mail : editor@irjet.net
Impact Factor : 8.226
www.irjet.net


e-ISSN: 2395-0056 p-ISSN: 2395-0072
International Research Journal of Engineering and Technology (IRJET)
(An ISO 9001 : 2008 Certified Journal)
Is hereby awarding this certificate to
Shradha Rajput
In recognition the publication of the manuscript entitled
Crime Analysis and Prediction using Machine Learning
published in our Journal Volume 11 Issue 4 April 2024
Editor in Chief
E-mail : editor@irjet.net
Impact Factor : 8.226
www.irjet.net


e-ISSN: 2395-0056 p-ISSN: 2395-0072
International Research Journal of Engineering and Technology (IRJET)
(An ISO 9001 : 2008 Certified Journal)
Is hereby awarding this certificate to
Sawan Kumar
In recognition the publication of the manuscript entitled
Crime Analysis and Prediction using Machine Learning
published in our Journal Volume 11 Issue 4 April 2024
Editor in Chief
E-mail : editor@irjet.net
Impact Factor : 8.226
www.irjet.net


e-ISSN: 2395-0056 p-ISSN: 2395-0072
International Research Journal of Engineering and Technology (IRJET)
(An ISO 9001 : 2008 Certified Journal)
Is hereby awarding this certificate to
Minal Thombare
In recognition the publication of the manuscript entitled
Crime Analysis and Prediction using Machine Learning
published in our Journal Volume 11 Issue 4 April 2024
Editor in Chief
E-mail : editor@irjet.net
Impact Factor : 8.226
www.irjet.net


e-ISSN: 2395-0056 p-ISSN: 2395-0072
International Research Journal of Engineering and Technology (IRJET)
(An ISO 9001 : 2008 Certified Journal)
Is hereby awarding this certificate to
Aachal Gupta
In recognition the publication of the manuscript entitled
Crime Analysis and Prediction using Machine Learning
published in our Journal Volume 11 Issue 4 April 2024
Impact Factor : 8.226
www.irjet.net
Editor in Chief
E-mail : editor@irjet.net

udemy

Certificate no: UC-04b16ab6-19a7-435b-9859-1d5985484da3
Certificate url: ude.my/UC-04b16ab6-19a7-435b-9859-1d5985484da3
Reference Number: 0004

CERTIFICATE OF COMPLETION

LaTex - A Scientific approach for report writing

Instructors **Sarita Bopalkar**

Shradha Rajput

Date **Feb. 9, 2024**
Length **6.5 total hours**