Crime Prediction using Naïve Bayes Algorithm

Mohammad Tahir
connectotahir@gmail.com
Raja Rajeswari College of Engineering, Bengaluru, Karnataka

Md Asif Anwar
mdasifanwarrrce@gmail.com
Raja Rajeswari College of Engineering, Bengaluru, Karnataka

Ruma Afsha Sultana
afshasultana33@gmail.com
Raja Rajeswari College of Engineering, Bengaluru, Karnataka

Manjunath S. R.
manjunathsr@rrce.org
Raja Rajeswari College of Engineering, Bengaluru, Karnataka

Fardeen Khan
fardeenkhan7411@gmail.com
Raja Rajeswari College of Engineering, Bengaluru, Karnataka

ABSTRACT

This paper presents the detection of the crimes happening in India. The criminal offences lead to certain punishment according to Indian Penal Code (IPC). For particular crime particular sections are assigned to punish the criminal or convicts with jail terms and fine. On this pre-processed data sets, by applying Naïve Bayesian algorithm we create a predictive model which analyze the data and helps to predict the crime type in a near future. We are using a dataset to apply Naïve Bayes algorithm to predict the crimes in India.

Keywords— *Naïve Bayes algorithm, Dataset*

1. INTRODUCTION

In today's Data In data mining, large pre-existing databases are evaluated, analyzed, and interpreted to produce new information that may be essential to the organization. The data mining process uses existing datasets to predict new information. In data mining, many approaches have been employed for analysis and prediction. However, very little effort has been made in the criminology field. Very few have made the effort to compare the information that all of theseapproaches produce. A police station and other criminal justice agencies typically have large databases of information that can be used to predict or analyze criminal activity involvements in society. Criminals can also be predicted from crime data. The paper identifies several Data Mining tools and approaches that can be used to analyze and predict crime in the telecom industry. The proposed Fraud detection methods can utilize either data mining techniques or machine learning algorithms

as suggested by different research results. This research proposes a new approach to solving different classification problems with deep learning techniques using objects, images, linguistic data, and dimensionality reduction Deep learning algorithms use deep mechanisms to compute deep memory models Through the hierarchical progression of learning features from bottom to top from bottom to top without previous knowledge of ant rule, which becomes even more challenging when dealing with huge data sets, multiple layers can be used to extract features from raw data. By avoiding feature engineering tasks that are time-consuming and resource-intensive, deep learning significantly reduces complexity.

2. REVIEWS OF LITERATURE

One of the Bakura, N. Suleiman, and I. Yusuf (2014), "Improved method of crime prediction classified by algorithms", International Symposium on Biometrics and Security Technologies (ISBAST), IEEE.

Kim, Param, Kalsi, and Taheri, (2019), "Crime Analysis Through Machine Learning", IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), DOI: 10.1109/IEMCON.2018.8614828.

Lavanya M, Kartika Padmanabhan AS & Lalitha S D, (2019), Prediction of Crime Rate Analysis Using Supervised Classification Machine Learning Approach, International Research Journal of Engineering and Technology (IRJET), Vol. 6771-6775, vol. 6.

An Analysis Framework Based on K-means Mining and Decision Trees to Analyze Crime Data by Kadhim Benjamin. Swadi Al-Janabi (2011, May), Vol. The first issue of Volume 1, Number 3, pages 8-24.

Murad, A. Mustapha, R. Iqbal. Shariat Panahy, P.H., and N.P. A current experimental study of a crime prediction algorithm

for Indian crime" Khanahmadliravi, (March 2013). Applied Sciences. Vol. 12 of Applied Science and Technology, p. Pages 4219-4225 in Vol. 6, No. 3.

This is (2005). Using SVMs for classification and regression. This is an international journal of pure and applied mathematics. The 100th issue. 811-820.

3. PROPOSED METHODOLOGY
System Analysis
Researchers are increasingly concerned with predicting terrorism activities. Due to the large number of events, it is difficult to predict terrorist groups responsible for certain acts of terrorism.

The current research aims to determine the correlation between terrorism and its causes. It is not possible to predict the future with existing efforts. Using machine learning approaches, it is possible to predict the likelihood of a terrorist attack, provided the required data is available. By taking relevant and effective measures, the results of this research can assist security agencies and policy makers in eradicating terrorism.

Therefore, there is a way of analyzing terrorist behavior patterns by analyzing the terrorism region and country with machine learning techniques and terrorism specific knowledge.

Functional Requirement
User interfaces are particularly I important. Customers are the outside clients. This product can be used by all of the customers for ordering and looking. Hardware Interfaces: Customers' PCs are utilized as outside equipment interfaces for ordering and viewing. Since the web association provided will be remote, the PCs may be portable PCs with remote LANs. Software Interfaces: Frameworks can be any version of Windows. Performance Prerequisites: With the goal that they can provide optimal performance, the PCs must have at least a Pentium 4 processor.

SYSTEM DESIGN
Design of a system is the definition of a system's architecture, components, modules, interfaces, and data in order to meet specific requirements. It could be viewed as the application of systems theory to product development. System analysis, system architecture, and system engineering share some similarities. A broader definition of product development is blending marketing, design, and manufacturing into a single approach to product development. If that is the case, then design is the act of taking the marketing information and designing the product to be manufactured. In other words, systems design is the process of defining and developing systems to meet the specifications of the user.
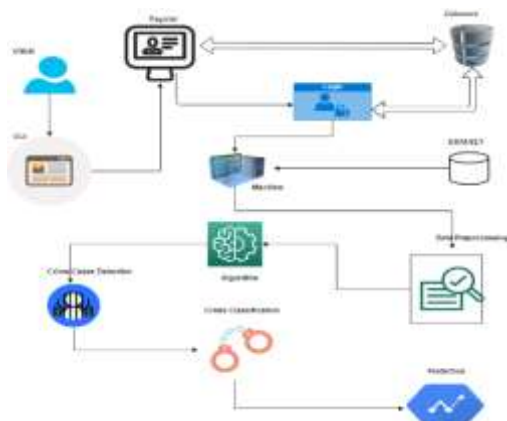

Fig: System Design Architecture

GUI: The user interacts with it to login to the project. User must register before logging in.

Database (MySQL): A user is registered with his or her details through it. GUIs are available for users to run the code and access the project.

Dataset: Data sets of crimes are imported into the Naive Bayes algorithm.

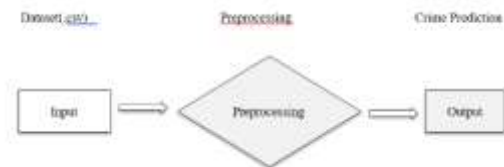Data Preprocessing: Data from the dataset is pre-processed and converted to a clean dataset.

Applying algorithm: The Naïve Bayes and random forest algorithm is used.

Crime Detection: Crime cases detected at this stage.

Crime Classification: Here are the types of crimes classified in this stage.

Prediction: Finally, the crime type is predicted.

Input/Output Design:

Input: We import the crime dataset. The raw dataset is fed into our machine. A .csv file contains the dataset. After loading the data, it proceeds to preprocessing

Preprocessing: Preprocessing is done on raw datasets. The dataset preprocessing is done to convert raw data into clean data. Output: We apply The Naïve Bayes algorithm. Finally, crime type is prediction is done.

Class diagram
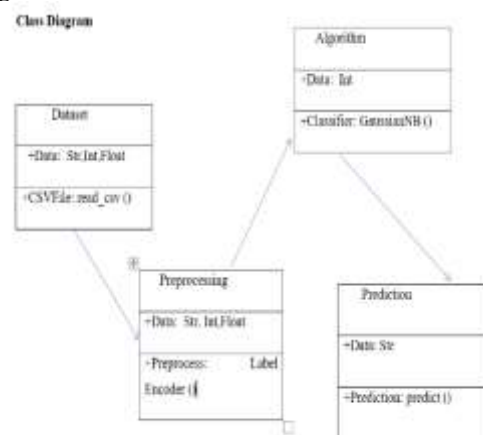
Dataset: The dataset is having the string, integer, float types of data. Hence the same dataset is read in this step.

Preprocessing: The data is imported to our Preprocessing step to process the raw data into numerical labels.

Algorithm: We apply Naïve Bayes algorithm to our project using inbuilt Naïve Bayes classifier function i, e GaussianNB ().
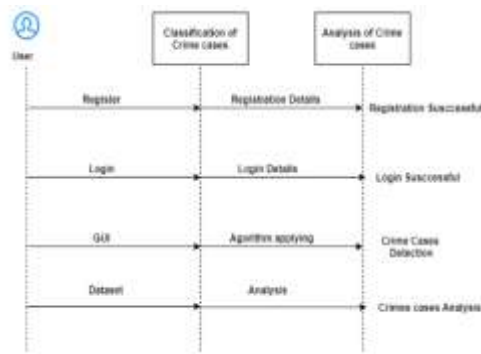Prediction: We predict the type of crime in this step.

Use Case Diagrams

In its simplest form, a use case diagram represents a user's interaction with a system and features the specifications of a use case. The use case diagram can depict the different types of users of a system and the various ways in which they interact with the system.
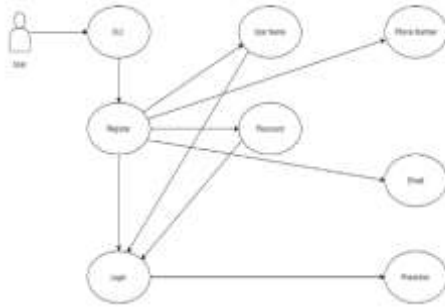
Sequence Diagram

User: GUI and MySQL will be used to access the project. GUI stands for Graphical User Interface. A database like MySQL is used to store all the user credentials. Upon registering, the user can access the project with a username and password.

Classification: Our algorithm is trained on the crimes dataset. In this stage, the crime is classified.

Crimes are classified using the Naïve Bayes algorithm Prediction and Analysis: The type of crime is analyzed and predicted.

Activity Diagram

Activity Diagrams are used to illustrate the flow of control in a system and to show the steps involved in executing a use case. Activity diagrams are used to model sequential and concurrent activities. An activity diagram is a visual representation of workflows. An activity diagram shows the conditions of flow and the order in which they occur. An activity diagram describes or depicts what causes a particular event
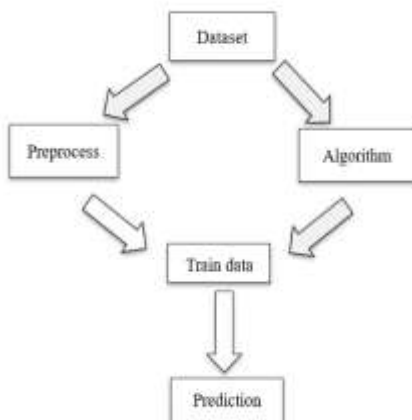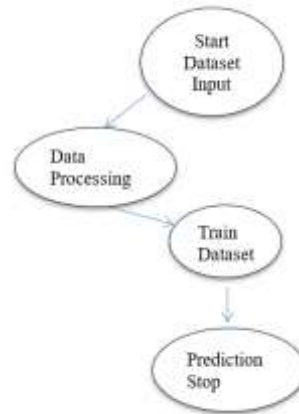
Dataset: The dataset of crimes is imported to our Naïve Bayes algorithm.

Data Preprocessing: The dataset data is preprocessed and converted into clean dataset.

Applying algorithm: The Naïve Bayes algorithm is applied.
Prediction: The crime type is predicted finally.

State Transition Diagram

In state-transition diagrams, all states of an object are described, as well as the points at which an object changes states (transitions), the conditions that must be met before a transition occurs (guards), and the activities an object undertakes during its lifetime (actions). State-transition diagrams can be used to describe the behavior of individual objects over the full range of use cases that affect them. The state-transition diagram cannot depict the collaboration between objects that causes a transition.

Dataset Input: The dataset of crimes is imported to our Naïve Bayes algorithm.
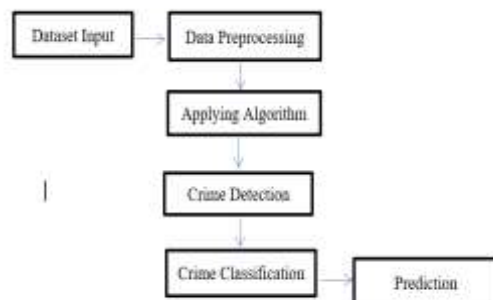
Data Preprocessing: The dataset data is preprocessed and converted into clean dataset.

Train Dataset: The Naïve Bayes algorithm is applied to train the dataset.

Prediction: The crime type is predicted finally, and process stops.

Data Flow Diagram

A data flow diagram is a graphical representation of the "flow" of data through an information system, representing various aspects of its processes. These are often preliminary steps used to create an overview of the system that can be elaborated later. A data flow diagram can also be used to visualize data processing (structured design).DFDs are also known as bubble charts. A system can be represented using graphical formalism by considering the input data, the process conducted on the data, and the output data generated by the system.

Dataset: The dataset of crimes is imported to our Naïve Bayes algorithm.

Data Preprocessing: The dataset data is preprocessed and converted into clean dataset.
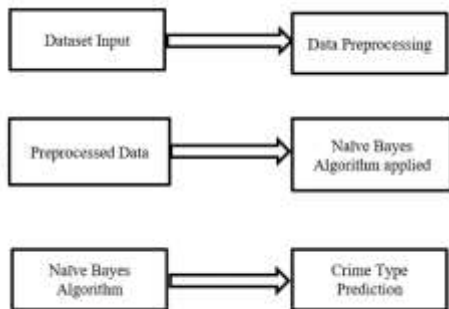
Applying algorithm: The Naïve Bayes algorithm is applied.

Crime Detection: The Crime cases detected here in this stage.

Crime Classification: The type of crimes is classified here in this stage.

Prediction: The crime type is predicted finally.

Algorithm

Naive Bayes: is a supervised classification algorithm method which is based on Bayes' theorem.

Based on Bayes' Theorem with an assumption of independence among predictors, it is a classification technique.

Naive Bayes assumes that the presence of a particular feature in a class has no effect on the presence of any other feature.

Example: When a fruit is red, round, and about 3 inches in diameter, it can be considered an Apple. The fact that all of these properties contribute to the probability that this fruit is an apple is why it is referred to as "Naive", even if they depend on each other or on the existence of other features.

Bayes' Theorem is given by below equation:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Were,

A, B = events

P(A|B) = probability of A given B is true

P(B|A) = probability of B given A is true

P(A), P(B)= the independent probabilities of A and B

Advantages of Naïve Bayes algorithm:
It is the simplest and most effective classification algorithm

A fast and accurate method for predicting the future.

Work efficiently with large datasets

Disadvantages:

By using NB, all variables are considered independent factors contributing to probability.

Applications:
Sentiment analysis,
Medical data classification,
Text Classification, Real time Prediction

4.SYSTEM IMPLEMENTATION
Modules
1. Registration Module
2. Login Module
3. Feature Selection
4. Prediction Module

Module description
1. Registration Module:
The user data is saves on the MySQL database.

The user data is to be saved for the login into the project.Once registration is successful the user can access the project.

2. Login Module:
After successful registration the user needs to login. After the login is successful the project can be predicted.

3. Features selection: It is done which can be used to build the model. The attributes used for feature selection are City, Year, Crime Type, Incidences, Rate. After feature selection the dataset is divided into pair of xtrain, ytrain and xtest, ytest. The algorithms model is imported form sklearn. Building model is done using Fit (xtrain, ytrain)

4. Prediction Module: Finally, the user can run the code for the results. The algorithm is run using GUI part. Once the algorithm is selected in the GUI the user can run the code for the successful prediction of the project

Methodology:
We use a crime dataset (.csv file):
➠ We train our dataset to the machine.
➠ We apply Naïve Bayes algorithm to the dataset.
➠ We classify the crimes.
➠ We finally predict the output.

Applications:
1.Detecting of high-profile crimes.
2.Analysis of the serious crimes in CBI.
3. Data Analysis of crime dataset.

System Testing
Testing is critical to ensure the quality and effectiveness of the proposed system in (satisfying) meeting its objectives. System testing occurs at various stages of the System design and implementation process with the objective of developing a transparent, flexible, and secure system. Software development is incomplete without testing, in a way, certifies whether the product, which is developed, meets the standards, to which it was designed. The testing process involves creating test cases that will be used to test the product.

Test Cases
The Test cases in unit testing are as follows:
Table I: Unit Test Case 1

Test Case ID	Unit Test Case 1
Description	User registered to the system
Input	User entered registration details.
Expected output	Registered successfully.
Actual Result/Remarks	Got the expected output.
Passed(?)	Yes.

Table II: Unit Test Case 2

Test Case ID	Unit Test Case 2
Description	User login to the system
Input	User enter login details.
Expected output	Login into the system.
Actual Result/Remarks	Got the expected output.
Passed(?)	Yes

Table III: Unit Test Case 3

| Test Case ID	Unit Test Case 3
Description	Click on the prediction and GUI code option to go to the next window
Input	Click the prediction button.
Expected output	Got next window to display the predicted output and GUI.
Actual Result/Remarks	Got the expected output.
Passed(?)	Yes.

Table III: Unit Test Case 4

Test Case ID	Unit Test Case 4
Description	The output window appears and displays the analyzed output.
Input	The final output window is displayed.
Expected output	The crime prediction is done and prints the accuracy.
Actual Result/Remarks	Got the expected output.
Passed(?)	Yes.

5. RESULTS AND DISCUSSION

Home Page:

In a main program open it will be displayed by on a register and login button.

Registration Page:
Once you click on the register button, the register form appears (username, password, emails, and phone number). Filling in these fields allows you to sign up. After registering, it will open the login page.

Log in Page:
The login page will be displayed then we should enter a username and password (wrong password) I will show error, we must enter an exact username and password program will be executed.
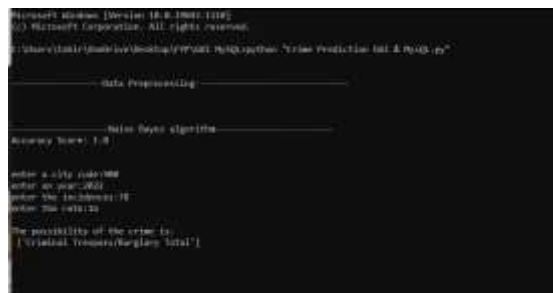
Result Page:
Based on the dataset with parameters such as city, crime type, and year, the prediction is made. Based on the year and the city, the crime type is predicted. We used the Naïve Bayes algorithm to predict the crime type.

6. CONCLUSION

A growing research advancement aims to reduce crime rates by using machine learning and data mining to detect crime. In this study, we study the distinct types of crimes and their occurrences at different times and places. In this project, we analyzed and predicted crimes in India based on a dataset.

7. FUTURE ENHANCEMENT

As a future extension of our work, we plan to apply more classification models to increase crime prediction accuracy and to improve overall performance.

8. REFERENCES

[1] Premalatha, M. & Vijayalakshmi, C.. (2019). SVM approach for non-parametric method in classification and regression learning process on feature selection with ε - insensitive region. Malaya Journal of Matematik. S. 276-279. 10.26637/MJM0S01/0051.

[2] Suhong Kim, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri,(2019), "Crime Analysis Through Machine Learning",DOI: 10.1109/IEMCON.2018.8614828 Conference: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

[3] Kirthika V, Krithika Padmanabhan A , Lavanya M & Lalitha S D,(2019), Prediction of Crime Rate Analysis Using Supervised Classification Machine Learning Approach, International Research Journal of Engineering and Technology (IRJET), Vol. 6, pp. 6771-6775.

[4] L. G. A. Alves, H. V. Ribeiro, and F. A. Rodrigues,(2018), "Crime prediction through urban metrics and statistical learning", Physica A, Vol. 505, pp. 435-443.

[5] S. Prabakaran and S. Mitra, (2018), "Survey of analysis of crime detection techniques using data mining and machine learning", Nat. Conf. on Math. Techn. and its Appl. (NCMTA 2018), IOP J. of Physics: Conf. Series, Vol. 1000.

[6] Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016). Crime prediction and forecasting in Tamilnadu using clustering approaches. 2016 International Conference on Emerging Technological Trends (ICETT). doi:10.1109/icett.2016.7873764 .

[7] M. V. Barnadas, Machine learning applied to crime prediction, Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, Sep. 2016.

[8] Premalatha, M. & Vijayalakshmi, C.. (2015). SVM approach for classification and regression with absolute value combination method for controlling complexity. International Journal of Pure and Applied Mathematics. 101. 811-820.

[9] Babakura, N. Sulaiman, and M. Yusuf, (2014), "Improved method of classification algorithms for crime prediction",International Symposium on Biometrics and Security Technologies (ISBAST) IEEE .

[10] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy & N. Khanahmadliravi,(2013, March) "An experimental study of classification algorithms for crime prediction", Indian J. of Sci. and Technol., Vol. 6, No. 3, pp. 4219-4225.

[11] J. Agarwal, R. Nagpal & R. Sehgal,(2013,December),"Crime analysis using k-means clustering", International Journal of Computer Applications, Vol. 83 – No.4.

[12] Chung-Hsien-Yu, Max W.Ward, Melissa Morabito & Wei Ding. "Crime Forecasting Using Data Mining Techniques" 11th International Conference on Data Mining pp. 779-786, IEEE 2011.

[13] Kadhim B. Swadi Al-Janabi,(2011,May), "A Proposed Framework for Analyzing Crime Data Set using Decision Tree and Simple K-means Mining Algorithms",Vol. 1- No. 3, pp. 8-24.

[14] Aravindan Mahendiran, Michael Shuffett, Sathappan Muthiah, Rimy Malla & Gaoqiang Zhang,(2011), "Forecasting Crime Incidents using Cluster Analysis and Bayesian Belief Networks".

[15] Jeroen S. de Bruin, Tim K. Cocx, Walter A. Kosters, Jeroen F. J. Laros & Joost N. Kok(2006),"Data mining approaches to criminal career analysis" ,In Proceedings of the Sixth International Conference on Data Mining (ICDM06) ,pp. 171-177

[16] Jain LC, Seera M, Lim CP, Balasubramaniam P (2014) A review of online learning in supervised neural networks. Neural Comput Appl 25:491–509.

[17] Usha D, Rameshkumar KA (2014) Complete survey on application of frequent pattern mining and association rule mining on crime pattern mining. Int J Adv Comput Sci Technol 3:264–275.

[18] Upadhyaya D, Jain S (2013) Hybrid approach for network intrusion detection system using k-medoid clustering and Naı̈ve Bayes classification. Int J Comput Sci Issues 10:231–236.

[19] Vural MS, Gök M, Yetgin Z (2014) Analysis of incident-level crime data using clustering with hybrid metrics. GAU J Appl Soc Sci 6:8–20.