International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Crime Detection Using Naive Bayesian Algorithm

[1]*Ashwini S S, Assistant Professor, CSE Dept, Sai Vidya Institute of Technology*
[2]*Amrutha P, CSE Dept, Sai Vidya Institute of Technology*
[3]*Madhushree D, CSE Dept, Sai Vidya Institute of Technology*
[4]*Sushmitha S, CSE Dept, Sai Vidya Institute of Technology*
[5]*Deepak Jaiswal, CSE Dept, Sai Vidya Institute of Technology*

Abstract - This essay discusses the identification of crimes committed in India. According to the Indian Penal Code, specific punishments are associated with the criminal offences (IPC).To sentence the offender or offenders to jail time and a fine, specific sections are assigned for each type of crime. By using the Naive Bayesian algorithm on these pre-processed data sets, we develop a forecasting model that analyses the data and aids in the prediction of the sort of crime that will occur in the near future.

***Key Words*:** Naïve Bayes algorithm, Dataset.

1.INTRODUCTION

Present-day Data mining is the process of extracting new information from big pre-existing datasets through evaluation, analysis, and interpretation. Data mining is a method for predicting new information using already-existing datasets. Numerous techniques have been used in data mining for forecasting and analysis. However, criminology has seen very little action. Comparing the data that each of these methods generates has not been attempted by many people. Large informational databases that can be utilised to forecast or assess societal involvement in criminal behaviour are often available at police stations and other criminal justice institutions. Data about crime can also be used to anticipate criminals. The study outlines a number of data mining techniques and technologies that might be applied to assess and forecast criminal activity in the telecom sector. According to various research findings, the proposed fraud detection approaches may make use of machine learning The extraction of features from raw data can be done using many layers, which is more difficult when working with large data sets.Deep learning dramatically lowers complexity by avoiding time- and resource-consuming feature engineering operations. The extraction of features from raw data can be done using many layers, which is more difficult when working with large data sets. Deep learning dramatically lowers complexity by avoiding time- and resource-consuming feature engineering operations.

2. PROPOSED METHODOLOGY

System Analysis

As terrorism activities are increasing researchers are very much concerned. There are many events, it will be tough to predict terrorist groups who are responsible for various activity of terrorism. The purpose of the current study is to ascertain the relationship between terrorism and its root causes. With current methods, it is impossible to foretell the future. If the necessary data is available, it is possible to estimate the likelihood of a terrorist strike using machine learning techniques. The findings of this study helps eradicate terrorism by recommending pertinent actions.

As a result, using machine learning techniques and expertise in the field of terrorism, it is possible to analyse the patterns of terrorist behaviour in terrorism-prone countries and regions.

Functional Requirement

User interfaces are especially crucial. The outside clients are the customers. All of the clients can order and browse using this product. Hardware Interfaces: For ordering and viewing, customers' PCs are used as interfaces for external equipment. The given web connection will be remote, therefore the PCs may be mobile and connected to far LANs. Windows versions 0 through 10 can be utilised as frameworks for software interfaces. Performance requirements: For the PCs to give the best performance, they must have a Pentium 4 CPU at the very least.

System Scheme

The process of establishing an architecture, along with its the use of components, modules, interfaces, and data satisfy specific needs is known as system design. The systems theory of product development could be applied to it. There are several similarities among system engineering, system architecture, and system analysis. The combination of marketing, design, and production into a single strategy is part of a broader concept of product development. In this example, using marketing data to create a buildable product would be called design. The process of identifying and developing systems in order to meet user needs is known as systems design.The definition of a system's architecture, components,

modules, interfaces, and data in order to satisfy particular criteria is known as system design.

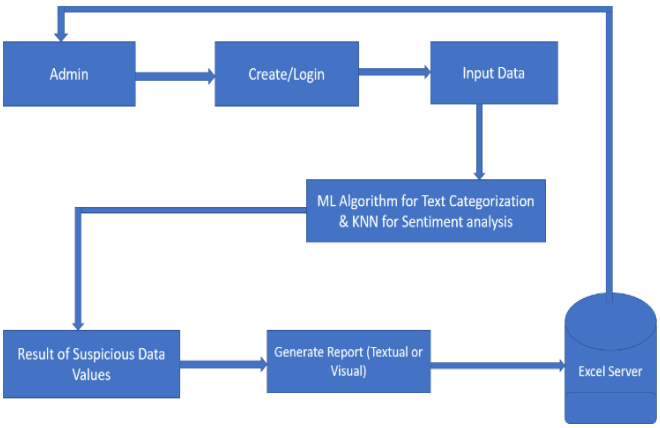

Fig 1: System Design Architecture

The GUI is used by the user to log into the project. Before logging in, the user needs register.

Database (ExcelServer): Through it, a user's information is registered. Users can access the project and run the code using GUIs.

The Naive Bayes technique imports data sets related to crimes.

Data preprocessing: The dataset's data is cleaned up and pre-processed before being used.

Algorithm used: The Random Forest and Naive Bayes algorithms.

Crime cases have been discovered thus far.

The several forms of crimes that are categorised in this level are listed below.

Prevision: Finally, if it is suspicious or non suspicious is

Predicted.

Input/Output Design:

Fig2: Input/out design

We import the crime dataset as input. Our system is provided the raw dataset. The dataset is contained in a.csv file. It starts with loading the data before preprocessing. Preprocessing: On unprocessed datasets, processing is done. To transform the raw to clean data, the dataset is first preprocessed. We use the Naive Bayes technique as our output. Finally, a prediction of the sort of crime is made.

Class diagrams

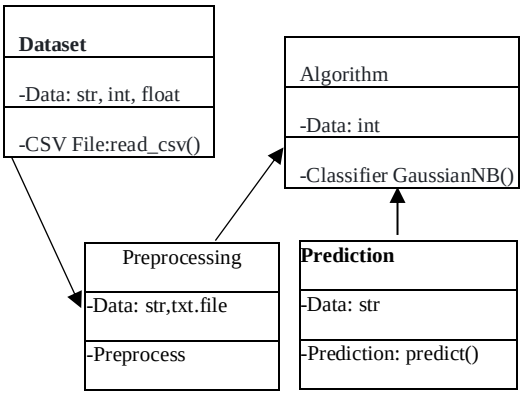

Fig 3: Class diagram

Preprocessing: The raw data is put into our Preprocessing stage where it is transformed into numerical labels.

The built-in Nave Bayes classifier function I e Gaussian NB is used to apply the Nave Bayes algorithm to our project ().

In this phase, we predict the type of crime that will occur.

Case Study Diagram

In its most basic form, a use case diagram depicts how a user interacts with a system and contains information about a use case. The various user types can be displayed in the use case diagram.

a system and the several ways people interact with it.

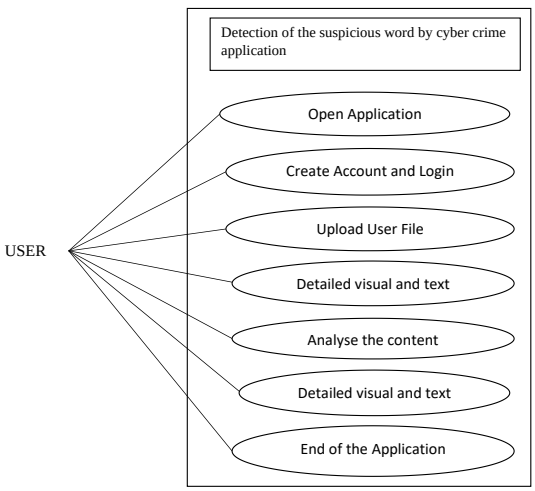

Fig 4: Use case diagram

Sequence Diagram

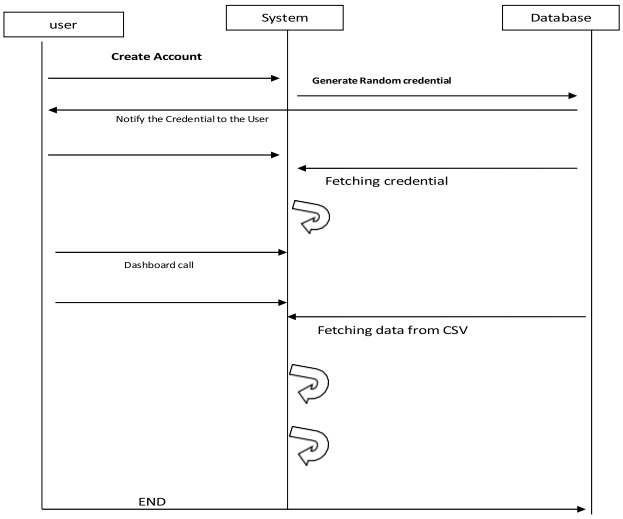

Fig 5: Sequence Diagram

Activity Schematic

Activity Diagrams are used to depict the system's control flow and the processes necessary to carry out a use case. Sequential and concurrent actions are both modelled using activity diagrams. Workflows are shown visually in activity diagrams. The circumstances of flow are displayed along with their occurrence sequence in an activity diagram. An activity diagram explains or illustrates the factors that lead to a specific event.

Dataset Input: Our Naive Bayes algorithm imports the dataset of crimes.Data cleaning and preprocessing are done on the dataset's data.

The algorithm is tested using the train dataset.

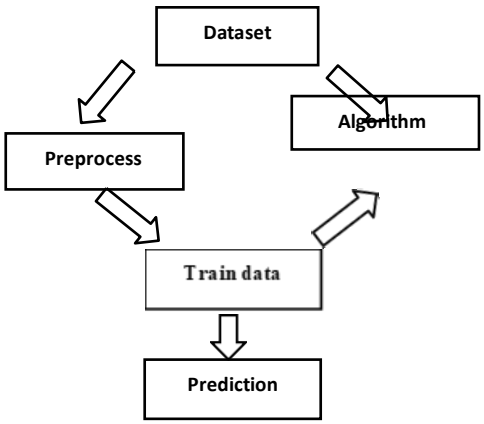

Fig 6: Activity Diagram

Dataset: Our Naive Bayes algorithm imports the dataset of crimes.

Preprocessing the data: The dataset is cleaned up and processed.

Application of algorithm: Naive Bayes is used.

Finally, a crime type prediction has been made.

State Transitions Diagram

All states of an object are described in state-transition diagrams. Transitions are the times when an item transforms between states, guards are the prerequisites that need to be fulfilled before a changeover can take place, and activities are the tasks that an object completes throughout the course of its lifespan (actions).Individual object behavior can be described using state-transition diagrams across all possible use cases. The cooperation between objects that results in a transition cannot be represented by the state-transition diagram.

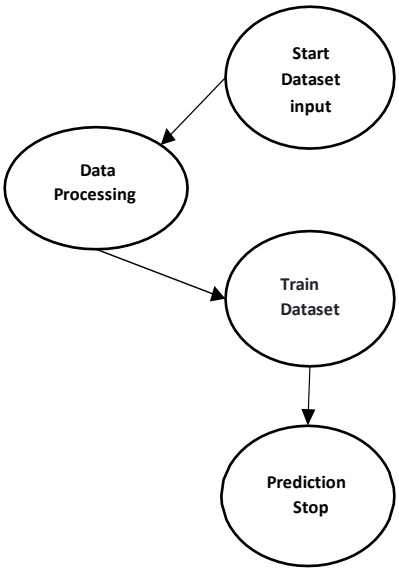

Fig 7: State Transitions Diagram

Finally, the crime type is anticipated, and the process comes to an end.

Diagram of a data flow

A data flow diagram illustrates how data "flows" through an information system, illustrating aspects of various operations. These are frequently first actions taken to develop a system overview that can be expanded upon later. To see how data is processed, a data flow diagram can be utilised (structured design) DFDs are also referred to as bubble diagrams. By taking into account A system can be graphically represented, this technique is used to process data, and production is generated by the system

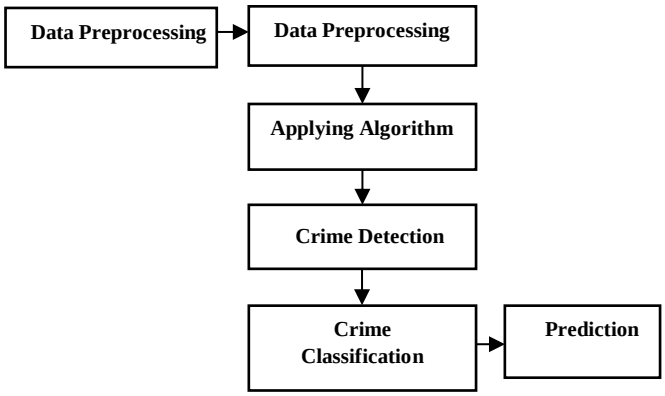

Fig 8: Data flow diagram

Algorithm

Naive Bayes is the name of a supervised classification algorithm built on the Bayes theorem.

The Bayes Theorem and the presumption of predictor independence serve as the foundation of this classification approach.

According to the Naive Bayes model, There is no connection between the presence of one feature in a class and the existence of any further features.

For instance, if a fruit is red, spherical, and around 3 inches in diameter, it can be classified as an apple.It is called "Naive" because all of these characteristics, even those that rely on other characteristics or the presence of other features, increase the likelihood that this fruit is an apple.

Bayes' s Theorem is given in below equation.

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

Where,

A,B = events

P(A|B) = likelihood that A is true if B true

P(B|A) = likelihood that B is true if A is true

P(A), P(B) = chances of A and B

Benefits of Naïve Bayes algorithm:

It is the most straightforward categorization method in use. a rapid and dependable technique for making future predictions. Work efficiently with large datasets.

Drawbacks of employing NB include the assumption that Every variable contributes to likelihood in an independent manner.

Applications:

Sentiment analysis, text categorization, medical data classification, and real-time prediction are examples of applications.

SYSTEM IMPLEMENTATION

Modules

1 .Registration

2. Login

3. Feature Selection

4. Prediction Description

Module description

1. Registration: The ExcelServer database stores the user data. For the purpose of logging into the project, the user data must be preserved. After the registration the project can be accessed.

2. Login Module: The user must login after successfully registering. The project can be forecast once the login has been successful.

3. Features are chosen that can be used to construct model. City, Year, Crime Type, Incidences, and Rate are the attributes that are used to choose the features. Following feature selection, the raw data is divided into pairs of xtrain , ytrain, and xtest.

4. Downloading the algorithm model from Sklearn. creating models utilising Fit (xtrain , ytrain). Prediction module: The user can execute the necessary code and see the results. The GUI component is used to carry out the algorithm. After choosing the algorithm in the GUI, the user can run the code to see if the project will be successful.

Methodology:

- It may use a dataset of crimes (.csv file)
- We train the machine with our dataset.
- The dataset is used in conjunction with the Naive Bayes technique
- We categorise the crimes.
- Finally, we forecast the outcome.

Applications:

1. Identifying prominent crimes.
2. Cbi's analysis of severe crimes.

3. Dataset on crime is analysed.

System Testing

Testing is essential to ensuring that the proposed system is high-quality and capable of (satisfactorily) achieving its goals.

To create a transparent, adaptable, and secure system, system testing is done the process of designing and implementing a system, at various phases. Testing, in a sense, validates whether the developed product satisfies the standards to which it was intended, and software development is incomplete without testing. The product will be tested using test cases as part of the testing process.

Case Studies

The following the test cases used in unit testing:

Table 1: First Test Suit 1

Test Case ID	Test Case 1
Interpretation	User registered to the system
Input	User accessed registration details
Expected Output	Registered Successfully
Actual Result State	Expected output is obtained
Passed(?)	Yes

Table 2: First Test Suit 2

Test Case ID	Test Case 2
Description	User login to the system
Input	User enter login details
Expected Output	Login into the system
Actual Result state	Expected output is Obtained
Progress(?)	Yes

Table 3: First Test Suit 3

Test Case ID	Test suit 3
Interpretation	Click on the prediction and GUI code option on to the new window
Input	Click the prediction button
Expected Output	Got next window to display the predicted output and GUI
Actual Result State	Expected output is obtained
Progress(?)	Yes

Table 4: First Test Suit 4

Test Case ID	Test Suit 4
Interpretation	Click on the prediction and GUI code option on to the new window
Input	Click the prediction button
Expected Output	Got next window to display the predicted output and GUI
Actual Result state	Got the expected output
Progress(?)	Yes

RESULTS AND DISCUSSION

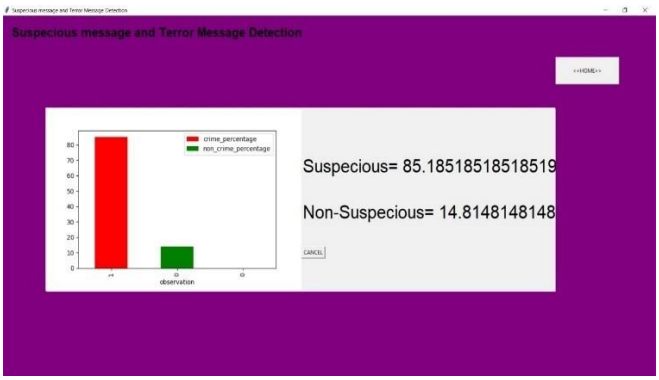

Fig 9: Expected output

Results page:

A prediction is performed using the dataset and parameters.

The output is in both graphical and textual form. That tell if the data is suspicious or not. And it also shows the percentage suspicious and non suspicious data.

FUTURE ENHANCEMENT

As a continuation of our study, we intend to use more categorization models to enhance performance overall and boost the accuracy of crime prediction

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 -sting

[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 predicition", Indian J. of Sci. and Technol., Vol. 6, No. 3, pp. 4219-4225.

[11] J. Agarwal, R. Nagpal , "Crime analysis using k- means clustering", International Journal of computer application, 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.

CONCLUSIONS

A growing body of research is attempting to lower crime rates by utilizing data mining and machine learning to identify criminal activity. This study examines the numerous criminal offences that take place at various dates and locations. In this research, we assessed and forecast crime in India. using a raw data.

ACKNOWLEDGEMENT

We are grateful to our colleagues from the Department of Computer Science and Engineering who offered advice that was very helpful to us while we worked on the project. We also like to express our gratitude to Prof. Ashwini S. S., an Assistant Professor at Sai Vidya Institute of Technology, for her assistance with this project and for her feedback on this paper. We also like to express our gratitude to Dr. H S Ramesh Babu, Principal of Sai Vidya Institute of Technology, and Dr. Shantakumar B Patil, Head of the Department of Computer Science and Engineering, for their consistent support and advice.