International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 3, May-June 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING

[1]Javvaji Venkata Krishna Rakesh Kumar, [2]Dr.M.Sreedevi
[1](Student Dept. of Master of Computer Applications, Amrita Sai Institute of Science and Technology, Paritala, Andhra Pradesh, 521180, India.
Email: javvaji.rakesh401@gmail.com)
[2] (Professor Dept. of Computer Science & Engineering, Amrita Sai Institute of Science and Technology, Paritala, Andhra Pradesh, 521180, India.
Email: m.sreedevi@amritasai.org.in)

----------**********----------

Abstract:

Recently, the rise of crime has made many people and the community experience disruption. To react properly to crime, we must analyse its patterns first. The research is centred on analysing crimes by looking at data from Kaggle. The main aim is to identify the crimes most often reported, by using details related to their type, how much time has passed and the area. Crime patterns were organised using machine learning algorithms, particularly Naïve Bayes, by the researchers. The study shows that the approach is accurate, matching or surpassing other studies on this topic.

Keywords *— Crime pattern analysis, Kaggle, machine learning, Naïve Bayes, crime prediction, classification, data analysis, crime trends.*

----------**********----------

1. Introduction

Experts say that crime is becoming an increasing danger to the public and the intensity of crimes is increasing as time goes on. A crime occurs when someone breaks a law and is seen as seriously wrongful. One important focus of criminology is to analyse crime patterns to discover what certain activities have in common among criminals. As more crime is happening, using advanced ways to predict and analyse crime trends is now especially crucial.

By blending machine learning approaches and past crime records, it's possible to predict the kinds of crime around different areas and times. More governments and agencies are using technology to fight and oversee crime. Crime analytics tools enable computing tools and programmes to support planning the best ways to stop anticipated crimes and share resources wisely.

Different analyses have investigated links among crime, where it takes place and the time it happens, to better explain criminal activity. When crime hotspots are found, authorized teams can deal with incidents and use prevention strategies more quickly. Our research uses machine learning to estimate what crimes will happen and where they will take place by studying both time and location. We hope to design classification models by studying open data on Kaggle which would help us discover where and when crime tends to occur.

The purpose of this study is to add to the efforts in crime pattern analysis, so that law enforcement agencies may better detect crime changes, assign their staff accordingly and increase public security. Analysing hotspots greatly helps police deal with crimes faster and use better ways to prevent them.

2. Literature Survey:

Both academics and industrial researchers have paid a lot of interest to crime prediction using

machine learning techniques and data analysis. Systems on the internet now access past crime statistics and let users input information for the analysis of and guessing of criminal actions.

Using Naive Bayes, SVM, Logistic Regression, Decision Trees, Random Forest, SGD and KNN models, crime data has been analysed with great accuracy by many researchers. Handy for finding patterns and crime hotspots, heat maps and correlation matrices are used to decide where police resources should be sent.

It has become valuable to use GIS and clustering techniques (such as K-Means and DBSCAN) to discover the locations of most crime. Recently, researchers are using LNNs and LSTMs to improve their predictions of crime data collected over time. Using precision, recall, F1-score and confusion matrices, we found that Gradient Boosting and Voting Classifiers have better predictive results than other methods.

3. Problem Analysis:

3.1 EXISTING SYSTEM:

Open-source datasets are gathered by the existing system which philtres these files to extract only necessary data. Crime patterns are identified and main features are discovered by making use of a decision tree on large data sets. As a result, classification becomes more straightforward going forward. Afterward, the classification team uses a DNN to sort crime patterns. The method assesses results with both datasets which makes it possible to predict crimes and help officers address them more rapidly.

DISADVANTAGES

- The present system experiences low accuracy, as using categories leads the classifier to favour nominal attributes with bigger values.
- With incomplete or poor-quality data, regions that use real-valued data find that classification techniques are not helpful.
- To improve the classifier's performance, its values have to be carefully tuned, so the optimal value must be chosen.

3.2 PROPOSED SYSTEM:

First, the proposed system performs pre-processing of the dataset with machine learning philtre and wrapper methods. At this point, you get rid of unnecessary and repeated data, so the dataset stays clean. Following this, the researchers divide the data so that some is assigned to training and some to testing. Subsequently, all crime-related data, including type, year, month, time, date and location, has an integer mapping process which supports smoother classification.

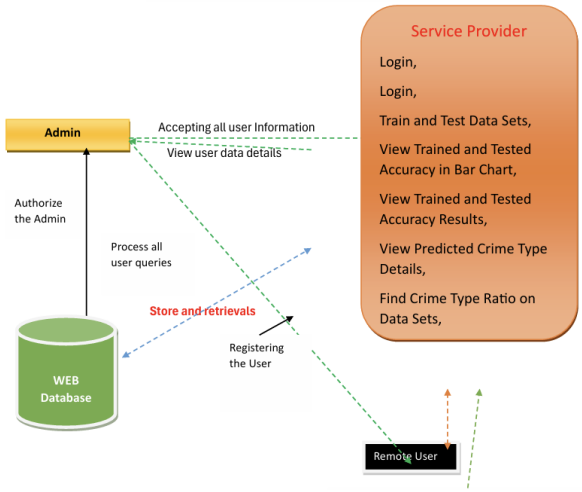
Then, the Naïve Bayes classifier, with the Bernoulli Naïve Bayes model, is used to analyse the relationships among the independent attributes. With government crime data, you can see how often crimes happen and when and where they are happening. Lastly, how well the model predicts is studied and Python is employed to put the prediction model into practise, running within the Colab environment for data analysis and machine learning.

Advantages:

1. The algorithm is designed well for crime pattern detection since its central features, time and location, play a big role in predicting crimes.
2. This problem is handled more effectively in this model than with previous ones.
3. This system means you don't need to figure out an optimal value on your own, because it works for numerical and descriptive data, even where there is less information.

4. The system we propose shows better accuracy than other models applied to prediction of crimes.

3.3 Architecture Diagram



3.4 System Requirements

Component	Specification
Operating System	Windows 7 Ultimate
Coding Language	Python
Front-End	Python
Back-End	Django-ORM
Designing	HTML, CSS, JavaScript
Database	MySQL (WAMP Server)

3.5 Hardware Requirements

Component	Specification
Processor	Pentium IV
RAM	4 GB (minimum)
Hard Disk	20 GB
Keyboard	Standard Windows Keyboard
Mouse	Two or Three Button Mouse
Monitor	SVGA

4. System Design and Development

4.1 Input Design

Accuracy and correctness of input data is important which means developers should pay close attention to input design in every software development process. The central objective of input design is to put data created by users into a form that a computer can read without making many errors. Input design that works well makes it so that input screens only allow data that is within the required limits.

The application includes user-friendly forms in all its modules, each of which cheques the entries automatically. The system's features detect mistakes during data entry, notify users by showing suitable messages and direct them to add the right data. Since error messages prompt the user, mistakes are quickly noticed and corrected, so the user can move ahead to further steps. Besides, the cursor automatically sets itself over the field the user should fill in, making sure they don't miss the target. Occasionally, users can select from preset alternatives which cuts the chance of entering wrong data.

Every part of data entry requires using validation procedures. Should an error take place, the system immediately shows a message to the user, so they have time to rectify the problem. Using this approach guarantees that the input process is clear and mistake-free, cutting down the chance of input problems in the application.

4.2 Output Design

The purpose of output design is to help the team, project leaders, administrators and clients talk to each other efficiently, inside the system. As a result of using the Virtual Private Network (VPN) system, project administration becomes easier for those in charge. Administrators are able to create new oreople, distribute assigned roles and manage each role's validity, while also offering clients with folder access in line with their projects. Once a

project is finished, administrators are able to give clients new projects with ease.

User auth is included right away so that administrators can oversee which users have access. A new user may be signed up by either the administrator or that user, though only the administrator can assign them and confirm their registration. The system was intended to be simple and clear, so that even novice users only need a short amount of training. When the server is started, the application starts on the LAN and the server manages things while all other systems are considered clients.

4.3 Module Design

There are distinct components within the system, each dealing with special responsibilities in the application. Because these modules depend on each other yet can be tested separately, development is clear and straightforward. The main topics included are these:

1. **User Authentication Module**: Here, we take care of user login, registration and permissions using roles to see if a user is authenticated and permitted before they use any system resources.

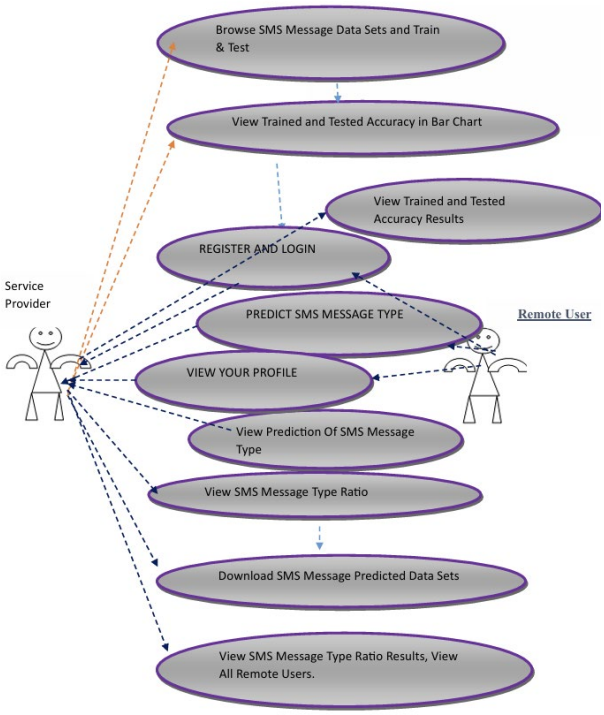
2. **Data Input Module**: Collects user information including the crime, its timing and the date and place where it took place. This section guarantees that data is traced correctly for useful actions afterward.

3. **Prediction Module**: Uses algorithms to study collected information and predict results such as from which locations crimes might happen the most.

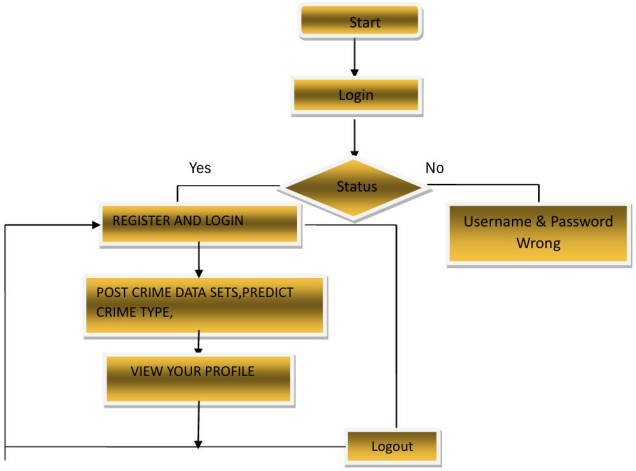
4. **Visualization Module**: Shows your data visually through charts, graphs, maps and tables. This module supports effective data analysis and choice making by organising information clearly.

5. **Admin Dashboard Module**: This application gives grants administrators tools to work with inputs, set up the system and look after user administration.

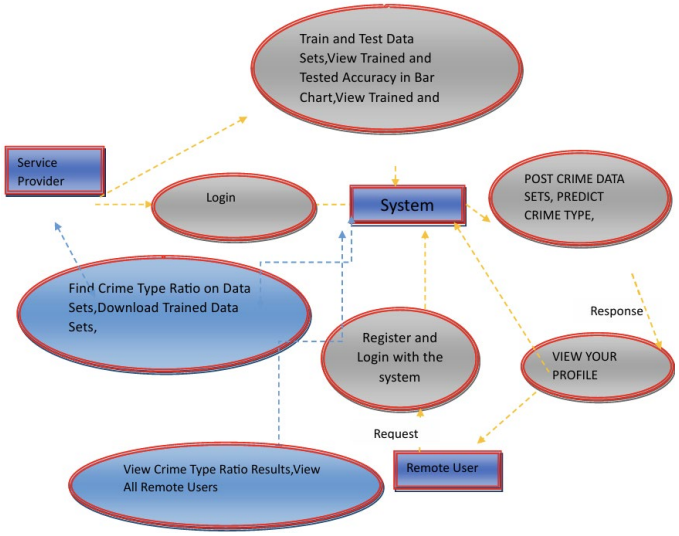
4.4 Use Case Diagram



4.5 Flow Chart : Remote User



4.6 Data Flow Diagram



5. SYSTEM TEST:

5.1 System Testing Overview

Errors are found in system testing when software is checked to see if it fits the requirements and what users hope for. Checking how everything functions should happen on the small element level as well as for the overall product. Here we will look at the important types of tests.

Types of Tests:

1. Unit Testing
Confirms that software modules generate predicted outputs and the internal mechanism works as intended.

2. Integration Testing
Brings all integrated parts together to behave as a whole system and discovers problems that may arise due to combinations of components.

3. Functional Testing
Cheques that the software carries out its functions satisfying business and technical needs with both valid and invalid inputs, producing the correct outcomes.

4. System Testing
Assures that all parts of the system are correct and function as specified, so that every connexion and partnership functions as designed.

5. White Box Testing
Tests the structure and rules inside the software by using code knowledge to review parts of the application not reached by black box tested test cases.

6. Black Box Testing
Analysis occurs from the outside, simply by comparing input and output, according to set specifications.

5.2 Unit Testing:

Most of the time, unit testing happens during the combined coding and testing phase of building software. Still, most of the time, coding and testing are performed separately.

Test Strategy and Approach:

Manual execution is planned for field testing which will be carefully detailed through functional testing.

Test Objectives:

- Every field on your form needs to be correctly operating.
- Cheque if clicking a link directs you to the right page.
- Cheque that users do not have to wait for entry screens, messages or for responses.

Features to be Tested:

- Validate that entries are in the correct format.
- Ensure that no duplicate entries are allowed.
- Confirm that all links lead to the correct page.

6. Conclusion:

Finally, the study confronts the problems caused by nominal data and continuous features by incorporating Multinomial Naïve Bayes (NB) and Gaussian Naïve Bayes (NB). The simple training these classifiers need means they are great for real-time crime prediction. The Naïve Bayesian Classification model is able to forecast and recognise the most common crimes due to addressing the issue of handling continuous target variables. The results are analysed using important metrics which are average precision, recall, F1 score and accuracy. Accuracy can be improved greatly if we use advanced machine learning techniques, but it can also improve even more with data from IoT surveillance, the weather and social media. Overall, this research suggests a useful way to handle crime intelligence, that could increase urban safety when it is connected to existing systems for managing crime.

Acknowledgement

I am thankful to the Management of Amrita Sai Institute of Science and Technology for giving me an opportunity to work with his project.
I would like to the thank **Dr. M. Sasidhar**, Principal, Amrita Sai institute of science and technology, for his constant encouragement and support during the progress of this work.
I am deeply grateful to **Dr. P. Chiranjeevi**, Professor and Head of the Department, for his valuable guidance and consistent support during the course of the project.
A special note of thanks to my internal guide, **Dr.M.Sreedevi**, for her exceptional guidance, constant motivation, and continuous encouragement, which played a crucial role in the successful completion of this project.

JAVVAJI VENKATA KRISHNA RAKESH KUMAR

References

[1] S. Kim, P. Joshi, P. S. Kalsi, and P. Taheri, "Crime analysis through machine learning," IEEE Trans., Nov. 2018.

[2] B. F. D. H. and A. Suruliandi, "Survey on crime analysis and prediction using data mining techniques," ICTACT J. Soft Comput., vol. 2, no. 1, pp. 1-7, Apr. 2012.

[3] S. S. Gosavi and S. S. Kavathekar, "A survey on crime occurrence detection and prediction techniques," Int. J. Manag. Technol. Eng., vol. 8, no. 12, pp. 3163-3174, Dec. 2018.

[4] A. Chandy, "Smart resource usage prediction using cloud computing for massive data processing systems," J. Inf. Technol., vol. 1, no. 2, pp. 108-118, 2019.

[5] R. Patil, M. Kacchi, P. Gavali, and K. Pimparia, "Crime pattern detection, analysis & prediction using machine learning," Int. Res. J. Eng. Technol., vol. 7, no. 6, pp. 2324-2327, Jun. 2020.

[6] U. M. Butt, S. Letchmunan, F. H. Hassan, M. Ali, A. Baqir, and H. H. R. Sherazi, "Spatio-temporal crime hotspot detection and prediction: A systematic literature review," IEEE Trans., Sep. 2020.

[7] Z. Nasiri, K. Zakikhani, K. Kimiya, and T. Zayed, "A failure prediction model for corrosion in gas transmission pipelines," Proc. Inst. Mech. Eng., Part O: J. Risk Reliab., vol. 234, no. 2, pp. 147-160, 2020.

[8] N. Dubey and S. K. Chaturvedi, "A survey paper on crime prediction technique using data mining," Corpus ID: 7997627, 2014.

[9] R. Ch., T. R. Gadekallu, M. H. Abdi, and A. Al-Ahmari, "Computational system to classify cybercrime offenses using machine learning," Sustainability, vol. 12, no. 10, pp. 3141-3156, May 2020.

[10] H. W. Kang and H. B. Kang, "Prediction of crime occurrence from multimodal data using deep learning," Peer reviewed journal, Apr. 2017.

