Emerging Trends in Digital Transformation

ISSN: 3107-4308 | Conference Issue 2025

International Conference on Emerging Trends in Engineering, Technology & Management (ICETM – 2025)

ETDT

ETDT

Emerging Trends in Digital Transformation

Special Issue - 2025 ISSN: **3107-4308** *Paper ID:* ***ETDT-SI-11***

International Conference on Emerging Trends in Engineering, Technology & Management (ICETM-2025)
Conducted by *Viswam Engineering College (UGC—Autonomous Institution)* held on 11th & 12th, April- 2025

CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING

[1]P.VISWANATHA REDDY, Associate Professor, Department of CSE, Viswam Engineering College
[2]Dr.N. KRISHNARAJ, Associate Professor, Department of Networking & Communication, Viswam Engineering College, SRMIST-KTR Campus

Abstract: In the fabric of contemporary society, the specter of crime looms large, posing significant challenges to the well-being and security of individuals and communities alike. The pervasiveness of criminal activities has disrupted the equilibrium within nations, exacerbating social tensions and undermining trust in public institutions. To confront this multifaceted issue, it is imperative to adopt proactive strategies that encompass comprehensive analysis and timely response mechanisms. This project embarks on a rigorous examination of crime patterns, drawing upon the wealth of open-source data available from platforms like Kaggle. By harnessing the power of advanced machine learning algorithms, particularly Naive Bayes classification, the endeavor aims to discern intricate nuances within crime data sets, identifying prevalent crime types, anticipating their occurrence, and uncovering temporal and geographical trends. The holistic approach adopted in this analysis transcends mere statistical observation, delving deeper into the underlying socio-economic factors and behavioral dynamics driving criminal activities. Through meticulous data preprocessing, feature engineering, and model refinement, the project endeavors to unlock actionable insights that can inform evidence-based policy formulation and resource allocation. By empowering law enforcement agencies and policymakers with timely and accurate information, it is envisaged that this initiative will contribute to the creation of safer, more resilient communities. Moreover, by fostering collaboration between academia, government, and civil society, the project aspires to catalyze a broader discourse on crime prevention and criminal justice reform.
Ultimately, the overarching goal is to foster a society where the rule of law prevails, and every individual can live free from the shadow of crime and insecurity.
***KEYWORDS*:** Crime pattern Analysis, Naive bayes, pervasiveness.

1. INTRODUCTION

Crime has become a major thread imposed which is considered to grow relatively high in intensity. An action stated is said to be a crime, when it violates the rule, against the government laws and it is highly offensive. The crime pattern analysis requires a study in the different aspects of criminology and in indicating patterns. The Government must spend a lot of time and work to imply technology to govern some of these criminal activities. Hence, use of machine learning techniques and its records is required to predict the crime type and patterns. It imposes the uses of existing crime data and predicts the crime type and its occurrence bases on the location and time.

Researchers undergone many studies that helps in analyzing the crime patterns along with their relations in a specific location. Some of the hotspots analyzed has become easier way of classifying the crime patterns. This leads to assist the officials to resolve them faster. This approach uses a dataset obtained from Kaggle open source based on various factors along with the time and space where it occurs over a certain period of time. We implied a classification algorithm that helps in locating the type of crime and hotspots of the criminal actions that takes place on the certain time and day.

In this proposed one to impose a machine learning algorithms to find the matching criminal patterns along with the assist of its category with the given temporal and spatial data.

2. SYSTEM ANALYSIS

EXISTING SYSTEM

In pre-work, the dataset obtained from the open source are first pre-processed to remove the duplicated values

DOI : https://doi.org/10.5281/zenodo.17277384

and features. Decision tree has been used in the factor of finding crime patterns and also extracting the features from large amount of data is inclusive. It provides a primary structure for further classification process. The classified crime patterns are feature extracted using Deep Neural network. Based on the prediction, the performance is calculated for both trained and test values. The crime prediction helps in forecasting the future happening of any type of criminal activities and help the officials to resolve them at the earliest.

Disadvantages of existing system

- The pre-existing works account for low accuracy since the classifier uses a categorical value which produces a biased outcome for the nominal attributes with greater value.
- The classification techniques do not suit for regions with inappropriate data and real valued attributes.
- The value of the classifier must be tuned and hence there is a need of assigning an optimal value.

PROPOSED SYSTEM

The data obtained is first pre- processed using machine learning technique filter and wrapper to remove irrelevant and repeated data values. It also reduces the dimensionality thus the data has been cleaned. The data is then further undergoing a splitting process. It is classified in t otest and trained data set. The model is trained by dataset both training and testing. It is then followed by

mapping. The crime type, year, month, time, date, place are mapped to an integer for ensuring classification easier. The independent effect between the attributes were analyzed initially by using Naïve Bayes. Bernoulli Naïve Bayes is used for classifying the independent features extracted. The crime features are labelled that allows to analyse the occurrence of crime at a particular time and location. Finally, the crime which occur the most along with spatial and temporal information is gained. The performance of the prediction model is found out by calculating accuracy rate. The language used in designing the prediction model is python and run on the Collab – an online compiler for data analysis and machine learning models.

Advantages of proposed system

- The proposed algorithm is well suited for the crime pattern detection since most of the featured attributes depends on the time and location.
- It also overcomes the problem of analyzing independent effect of the attributes.
- The initialization of optimal value is not required since it accounts for real valued, nominal value and also concern the region with insufficient information.
- The accuracy has been relatively high when compared to other machine learning prediction model.

3. BLOCKDIAGRAM

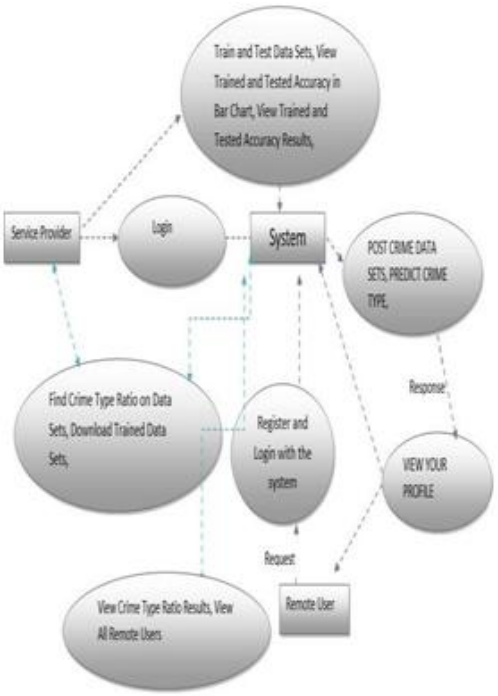

Fig: Data Flow diagram

DOI : https://doi.org/10.5281/zenodo.17277384

DESCRIPTION:

Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate as possible. So, inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designed to provide to have a validation control over the input limit, range and other related validations.

This system has input screens in almost all the modules. Error messages are developed to alert the user whenever he commits some mistakes and guides him in the right way so that invalid entries are not made. Let us see deeply about this under module design.

Input design is the process of converting the user created input into a computer-based format. The goal of the input design is to make the data entry logical and free from errors. The error is in the input are controlled by the input design. The application has been developed in user- friendly manner. The forms have been designed in such away during the processing the cursor is placed in the position where must be entered. The user is also provided within an option to select an appropriate input from various s alternatives related to the field in certain cases. Validations are required for each data entered. Whenever a user enters an erroneous data, error message is displayed and the user can move on to the subsequent pages after completing all the entries in the current page.

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining are cord of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself for a user can himself register as a new user but the task of assigning projects and validating a new user rest with the administrator only.

The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

4. CONCLUSION

In this paper, the difficulty in dealing with the nominal distribution and real valued attributes is overcome by using two classifiers such as Multi nominal NB and Gaussian NB. Much training time is not required and serves to be the best suited for real time predictions. It also overcomes the problem of working with continuous target set of variables where the existing work refused to fit with. Thus, the crime that occur the most could be predicted and spotted using Naïve Bayesian Classification. The performance of the algorithm is also calculated by using some standard metrics. The metrics include average precision, recall, F1 score, and accuracy are mainly concerned in the algorithm evaluation. The accuracy value could be increased much better by implementing machine learning algorithms.

FUTUTREENHANCEMENT

- Incorporating more sophisticated machine learning models such as deep learning or ensemble techniques could enhance the accuracy and efficiency of crime detection. These models can handle complex data
- patterns and adapt to evolving criminal behaviors.
- Implementing real-time data analysis capabilities would enable quicker response times to potential criminal activities. Integration with IoT devices, social media feeds, and other sources could provide valuable insights for proactive crime prevention.
- Enhancing the system with predictive analytics can help forecast crime trends and hotspots, allowing law enforcement agencies to allocate resources effectively. By analyzing historical data and identifying patterns, the system can anticipate where and when crimes are likely to occur.
- Combining various data sources such as video surveillance footage, sensor data, social media posts, and

DOI : https://doi.org/10.5281/zenodo.17277384

criminal databases through multimodal data fusion techniques can provide a more comprehensive understanding of criminal activities. This holistic approach improves the accuracy of crime detection and facilitates better decision-making for law enforcement.

- As the project deals with sensitive data, it's crucial to prioritize privacy and ethical considerations. Future enhancements should focus on implementing robust encryption techniques, anonymization methods, and strict access controls to safeguard individuals' privacy rights while still effectively detecting and preventing crimes. Additionally, incorporating transparency measures to ensure accountability and fairness indecision-making processes is essential.

REFERENCES

1. Chen, Ling, and Xu Lai. "Comparison between ARIMA and ANN models used in short-term wind speed forecasting." Power and Energy Engineering Conference (APPEEC), 2011 Asia- Pacific. IEEE, 2011.
2. Agarwal, Jyoti, Renuka Nagpal, and Rajni Sehgal."Crime analysis using K-
3. Means clustering."International Journal of Computer Applications 83.4 (2013).
4. Sathyadevan, Shiju, and Surya Gangadharan. "Crime analysis and prediction using data mining." Networks &Soft Computing (ICNSC), 2014 First International Conference on. IEEE, 2014
5. McClendon, Lawrence, and Natarajan Megha nathan. "Using machine learning algorithms to analyse crime data." Machine Learning and Applications: An International Journal (MLAIJ) 2.1 (2015).
6. Kiani, Rasoul, Siamak Mahdavi, and Amin Keshavarzi. "Analysis and prediction of crimes by clustering and classification." Analysis 4.8 (2015)
7. Suhong Kim, Param Joshi, Parminder Singh, Kalsi,Pooya Taheri, “Crime Analysis Through Machine Learning”, IEEE Transactions on November 2018.
8. Benjamin Fredrick David. H and A.Suruliandi, “Survey on Crime Analysis and Prediction using Data mining techniques”, ICTACT Journal on Soft Computing on April 2012.
9. Shruti S.Gosavi and Shraddha S.Kavathekar,“ A Survey on Crime Occurrence Detection and prediction Techniques”, International Journal of Management, Technology And Engineering , Volume 8,Issue XII, December 2018.
10. Chandy, Abraham, "Smart resource usage prediction using cloud computing for massive data processing systems" Journal ofInformationTechnology1,no.02
11. (2019):108-118.
12. Learning Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimparia, “Crime Pattern Detection, Analysis & Prediction using Machine”, International Research Journal of Engineering and Technology,(IRJET)e-ISSN:2395-0056,
13. Volume:07,Issue:06,June2020
14. Umair Muneer Butt, Sukumar
15. Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi, “Spatio-Temporal Crime Hotspot Detection and Prediction: A Systematic Literature Review”, IEEE Transactions on September 2020.
16. Nasiri, Zakikhani, Kimiya and Tarek Zayed,"A failure prediction model for corrosion in gas transmission pipelines", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, (2020).

DOI : https://doi.org/10.5281/zenodo.17277384