

CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING

Mrs. B. Haritha Lakshmi

Asst.Professor

Department of Information Technology Malla Reddy
Engineering College for Women(UGC-Autonomous)
Maisammaguda, Hyd-500100, Telangana, India.

S.Shresta

Student

Department of Information Technology Malla Reddy
Engineering College for Women(UGC-Autonomous)
Maisammaguda, Hyd-500100, Telangana, India.

S.Sathhika

Student

Department of Information Technology Malla Reddy
Engineering College for Women(UGC-Autonomous)
Maisammaguda, Hyd-500100, Telangana, India.

P.L.MShreya

Student

Department of Information Technology Malla Reddy
Engineering College for Women(UGC-Autonomous)
Maisammaguda, Hyd-500100, Telangana, India.

ABSTRACT

Crime has recently emerged as a clear method of putting people and society in peril. A country's population becomes unbalanced as a result of rising crime. Understanding crime trends is essential for analysing and anticipating this kind of criminal activity. This study applies one such analysis of crime patterns by leveraging open-source Kaggle crime data, which is then utilised to forecast the most recent incidents. Estimating which form of crime contributes the most, coupled with the time frame and location where it occurred, is the project's main goal. This work implies a few machine learning methods, namely Naive Bayes

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 also in indicating patterns. The Government has to 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 analysing the crime patterns along with their relations in a specific location. Some of the hotspots analysed 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.

RELATED WORK :

Crime are of different type that occurs at different locations around the various geographical location. Many research scholars have been suggesting a mechanism to analyse the relationship between crime and social variables that includes unemployed individuals, earning amount, level of education and so on. Suhong Kim and Param Joshi [1] proposed two different machine learning models which is used for prediction, K nearest neighbour algorithm (KNN) and decision tree approach. The accuracy obtained ranges between 39 to 44 percent when predicting crime patterns and finding the crime type. Benjamin Fredrick David. H imposed a data mining technique that involves evaluating and inspect large pre-existing datasets in accordance to deliver more information. The extraction of new patterns is cross checked with predefined datasets available. Shraddha S. Kavathekar used association rule mining in predicting crimes. Some Machine learning algorithms including Deep Neural Network (DNN) and Artificial Neural Network (ANN) have been implied. A deep neural network works more accurately using the feature level dataset. Using DNN, entirely connected convolution layers has been used in building the prediction model, mainly for multilabelled data classification. It was

ISBN: 978-X-XXXX-XXXX-X/22/\$31.00 ©2022 IEEE

implemented using Tensorflow that is an API mainly designed for Deep learning technique with the dropout layers. These findings suggest that when there is more count of missing values, there is a need for pre-processing because crimes do not occur in the same manner but focuses on some particular areas. Artificial Neural Network [ANN] is based on the prognosis by trend analysis in solving problems. It comprises of enormous amount of processing constituent that works altogether in building a model. Chandyan and Abraham proposed a random forest classifier in extracting the features for data processing using cloud computing. The extracted features are request number, user identification, expirytime, time of arrival and memory requirement. After feature extraction, the prediction of work load is done by using the trained data that has been perceived from the learning stage that allows to learn the details of the extracted features from user's request. Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimpuria suggests an Apriori algorithm for frequent patterns and the result obtained from K-means is used. Due to increase in crime rate over these recent years, system has to handle an enormous amount of data which requires more time to analyse them manually. Hence, advance machine learning approaches like K means clustering has been used. A literature survey on Spatial and Temporal Hotspot prediction of crime proposed a study to categorize and evaluate the location and time of the crime hotspot detection techniques by performing (SLR) Systematic Literature Review. Fuzhan Nasiri, Zakikhani, Kimiya and Tarek Zayed suggested a failure prediction model that helps in detecting the corrosion in the pipelines of gas transmission. Most of the prediction model depend absolutely on the experimental tests data or involving some of the limited historical data records. This helps in ignoring the corrosion from various geographical circumstances. Nikhli Dubey and Setu K. Chaturvedi imposed pertinent analysis of data mining approaches for the detection of the impending future crime. A Computational mechanism to classify the crime using machine learning techniques proposed a malleable computational implementation tool to analyse the crime rate in a country helps in classifying cybercrimes. Hyeon-Woo Kang and Hang-Bong Kang suggested a fusion method based on Deep Neural Network in predicting the criminal activities from the feature level data with sufficient parameters.

EXISTING SYSTEM

Pre-work is the process of removing duplicate values and features from a dataset that was collected from an open source. Decision trees have been used to extract features from massive amounts of data and to identify patterns in crime. It offers a basic framework for subsequent classification procedures. Deep neural networks are used to extract features from the identified crime patterns. The performance is computed for both trained and test values based on the prediction. The crime prediction aids in foreseeing future instances of any kind of criminal activity and assists the authorities in swiftly resolving them.

DISADVANTAGES :

1. The pre-existing works account for low accuracy since the classifier uses a categorical values which produces a biased outcome for the nominal attributes with greater value.

2. The classification techniques does not suited for regions with inappropriate data and real valued attributes.
3. The value of the classifier must be tuned and hence there is a need of assigning an optimal value

PROPOSED SYSTEM :

To remove redundant and irrelevant data values, the collected data is initially pre-processed using the machine learning techniques filter and wrapper. Additionally, it lowers the dimensionality; as a result, the data is now clean. After then, the data passes through another dividing step. It is divided into a trained data set and a test data set. Both the training and testing datasets are used to train the model. The next step is mapping. To make classification easier, the crime type, year, month, time, date, and location are all mapped to integers. Utilizing Nave Bayes, the independent relationship between the attributes is initially analysed. The retrieved independent features are classified using Bernouille Nave Bayes. It is possible to analyse the occurrence of crime at a specific time by labelling the criminal features. The accuracy rate is used to determine the prediction model. The prediction model was created in Python and utilises Colab, an online compiler for machine learning and data analysis models.

ADVANTAGES :

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

MODULES

SERVICE PROVIDER : In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Predicted Tweet Account Type Details, Find Tweet Account Type Ratio, Download Predicted Data Sets, View Tweet Account Type Ratio Results, View All Remote Users..

VIEW AND AUTHORIZE USERS : In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

REMOTE USER : In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like POST CRIME DATA SETS, PREDICT CRIME TYPE, and VIEW YOUR PROFILE.

CONCLUSION

In this study, two classifiers—Multinomial NB and Gaussian NB—are used to address the challenge of dealing with the nominal distribution and real valued features. Real-time forecasts can be made with the least amount of training time possible. Additionally, it solves the issue of dealing with a continuous target set of variables, which the previous work was unable to accommodate. Therefore, Naive Bayesian Classification could be used to forecast and identify the crimes that occur the most frequently. Additionally, the algorithm's performance is determined using a few common criteria. The metrics average precision, recall, F1 score, and accuracy are the ones that matter most when evaluating an algorithm. By using machine learning methods, the accuracy value might be improved significantly.

FUTURE WORK

Though it overcomes the problem of the existing work, it has some limitations. In the situation of absence of class labels, then the probability of the estimation will be zero. As a future extension of the proposed work, the application of more machine learning classification models proves to increase accuracy in crime prediction and will enhance the overall performance. It helps in providing a better study for the future improvement by taking the income information into consideration for neighborhoods places in order to foresee if any relationship between the income levels of a particular in the neighborhood places and their crime rate.

REFERENCES:

- [1] Suhong Kim, Param Joshi, Parminder Singh Kalsi, Pooya Taheri, "Crime Analysis Through Machine Learning", IEEE Transactions on November 2018.
- [2] 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.
- [3] 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.
- [4] Chandu, Abraham, "Smart resource usage prediction using cloud computing for massive data processing systems" Journal of Information Technology 1, no.02 (2019): 108-118.
- [5] Learning Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimpriya, "Crime Pattern Detection, Analysis & Prediction using Machine", International Research Journal of Engineering and Technology, (IRJET) e-ISSN: 2395-0056, Volume: 07, Issue: 06, June 2020
- [6] Umair Muneer Butt, Sukumar 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.

[7] 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).

[8] Nikhil Dubey and Setu K. Chaturvedi, "A Survey Paper on Crime Prediction Technique Using Data Mining", Corpus ID: 7997627, Published on 2014.

[9]. Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, "Computational System to Classify Cyber Crime Offenses using Machine Learning", Sustainability Journals, Volume 12, Issue 10, Published on May 2020.

[10] Hyeon-Woo Kang and Hang-Bong Kang, "Prediction of crime occurrence from multimodal data using deep learning", Peer-reviewed journal, published on April 2017.