

Felony Type Prediction Using Machine Learning Algorithms

K. Surekha¹, M. Soniya², K. Aashritha³, K. Bhagya Lakshmi⁴, U.V.V. Manikanta⁵, G. Satya Santosh⁶

¹Associate Professor, ²⁻⁶Students B.Tech. Computer Science Engineering, V. S. M. College of Engineering, Ramachandrapuram, A.P, India

Abstract - In this era of recent times, crime has become an evident way of making people and society under trouble. An increasing crime factor leads to an imbalance in the constituency of a country. In order to analyse and have a response ahead this type of criminal activities, it is necessary to understand the crime patterns. This study imposes one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of most recently occurring crimes. The major aspect of this project is to estimate which type of crime contributes the most along with time period and location where it has happened. Some machine learning algorithms such as Logistic Regression is implied in this work in order to classify among various crime patterns and the accuracy achieved was comparatively high when compared to precomposed works.

Key Words: Crime Type, Logistic Regression, Occurrence Prediction, SVM, Accuracy.

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 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 algorithm to find the

matching criminal patterns along with the assist of its category with the giventemporal and spatial data.

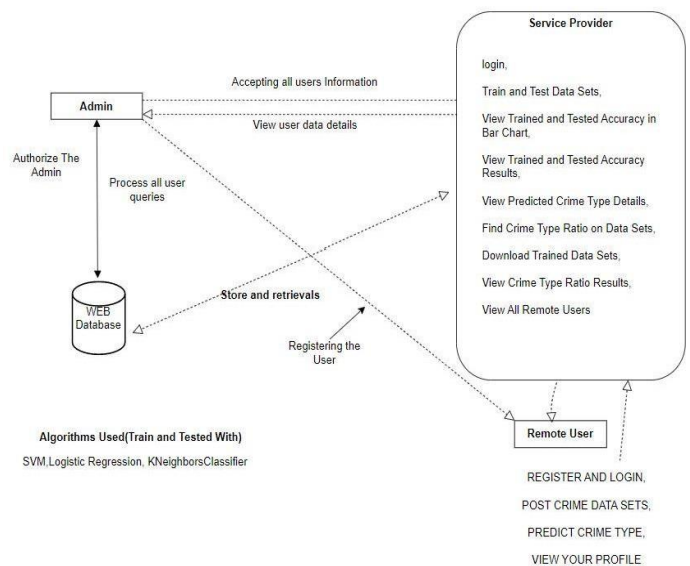


Fig -1: System Architecture

2. IDENTIFY, RESEARCH AND COLLECT IDEA

In [1], Suhong Kim and Param Joshi 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.

In [2], 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.

In [3], Chandy and Abraham proposed a random forest classifier in extracting the features for data processing using cloud computing. The extracted features are requesting number, user identification, expiry time, 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.

In [4], Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimpriya [5] 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.

In [5], Nikhli Dubey and Setu K. Chaturvedi [8] imposed pertinent analysis of data mining approaches for the detection of the impending future crime.

In [6], Hyeon-Woo Kang and Hang-Bong Kang [10] suggested a fusion method based on Deep Neural Network in predicting the criminal activities from the feature level data with sufficient parameters.

3. PROPOSED APPROACH

The data obtained is first pre-processed using machine learning technique filter and wrapper in order 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 into test 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 are analysed initially by using Decision Tree. Logistic Regression 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 for data analysis and machine learning models

DATA CHARACTERISTICS

In addition, When using machine learning algorithms to predict crime type and occurrence, it is important to consider the following data characteristics:

Data quality: The quality of the data is essential for accurate predictions. Data must be complete, accurate, and free from errors or biases. Incomplete or inaccurate data can lead to incorrect predictions.

Data quantity: The quantity of data is also important. The more data available, the more accurate the predictions are

likely to be. Sufficient data is needed to train the machine learning algorithm effectively.

Data relevance: The data used for training the machine learning algorithm must be relevant to the problem being solved. For example, data about traffic violations is not relevant when trying to predict violent crimes.

Data consistency: The data must be consistent over time. Any changes in the data collection process or the environment can affect the accuracy of the predictions.

Data diversity: The data must be diverse enough to capture all possible scenarios. A lack of diversity in the data can lead to biased predictions.

Data pre-processing: The data must be pre-processed before being used for training the machine learning algorithm. This includes cleaning, normalization, and feature extraction.

Data labelling: The data must be labelled accurately to train the machine learning algorithm. Labelling errors can affect the accuracy of the predictions.

Data imbalance: Imbalanced data occurs when one class of data is significantly more represented than another. This can lead to biased predictions and must be addressed during the training process.

Data privacy: Sensitive data such as personal information and criminal records must be protected during the data collection and training process to ensure the privacy and security of individuals.

4. FUTURE WORK

The proposed work in this paper discusses about the Prediction of crime type and occurrence by using machine learning algorithms such as Logistic Regression, SVM, KNN classifier. 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 neighbourhoods places in order to foresee if any relationship between the income levels of a particular in the neighbourhood places and their crime rate.

5. CONCLUSION

In conclusion, the difficulty in dealing with the nominal distribution and real valued attributes is overcome by using classifiers such as Logistic Regression, KNN, SVM. Much

International Conference on Recent Trends in Engineering & Technology- 2023 (ICRTET-3)**Organised by: VSM College of Engineering, Ramachandrapuram**

training time is not required and serves to be the best suited for realtime 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.

ACKNOWLEDGEMENT

We are appreciative to our Division of Computer Science Engineering for their help and giving us a chance to make things simpler. While looking about this point, we found out about different significant and fascinating realities. The tools and the web services provided are of much help.

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] Chandy, 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 Pimparia, "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", Peerreviewed journal, published on April 2017.

BIOGRAPHIES

Team Guide, K. Surekha
Associate Professor
Computer Science and
Engineering in VSM College
of Engineering,
Ramachandrapuram.



Team Leader, M. Soniya is
Computer Science and
Engineering student in VSM
College of
Engineering,
Ramachandrapuram.



Team Member, K.
Aashritha is Computer
Science and
Engineering student in
VSM College of
Engineering,
Ramachandrapuram.



Team Member, K. Bhagya
Lakshmi is Computer
science an
Engineering student in VSM
College of Engineering,
Ramachandrapuram.

International Conference on Recent Trends in Engineering & Technology- 2023 (ICRTET-3)**Organised by: VSM College of Engineering, Ramachandrapuram**



Team Member, U. V. V. Manikanta is Computer Science and Engineering student in VSM College of Engineering, Ramachandrapuram.



Team Member, G. Satya Santosh is Computer Science and Engineering student in VSM College of Engineering, Ramachandrapuram.