INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 05, May 2023, pp : 1262-1266

e-ISSN : 2583-1062

Impact Factor : 5.725

CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING ALGORITHM

Sanjay. S[1], A. Jeevarathinam[2], Sugandan K. M[3]

[1,3]Student, Department Of Computer Science, Sri Krishna Arts And Science College, India.

[2]Assistant Professor Department Of Computer Science, Sri Krishna Arts And Science College, India.

ABSTRACT

Nowadays, in this modern culture, crime has become a visible manner of making people and society in danger. The growing crime factor this causes an imbalance in a country's constituency. Understanding crime patterns is necessary in order to analyse and respond to this type of criminal activity. This work uses crime data from the Kaggle open source database to implement one such crime pattern analysis, which is subsequently used to predict the most recent crimes. The main goal of this research is to identify the type of crime that contributes the most, along with the time period and location where it occurred. In order to classify different criminal patterns, some machine learning methods, such as Random forest classifier are used in this work. while compared to precomposed works, the accuracy achieved was rather high.

Keywords: Crime type, Crime occurrence, Analyse, Crime patterns, Random-forest classifier

1. INTRODUCTION

The crime has emerged as a significant issue that is thought to be intensifying quickly. When an activity is against the law, is extremely offensive, and breaches a rule, it is considered to be a crime. The knowledge of criminology's various facets as well as the ability to identify patterns are both necessary for the crime pattern analysis. Using technology to control some of these illegal enterprises requires a lot of time and effort from the government. As a result, the use of machine learning algorithms and their data is require to anticipate the type and patterns of crime.

It makes use of current crime data to anticipate the type of crime and its occurrence based on location and time. Several studies have been conducted by researchers to aid in the analysis of crime patterns and their relationships in a specific place. Some of the locations studied have made it easier to classify criminal patterns. This enables officials to resolve issues more quickly. This method use a dataset taken from Kaggle open source using multiple criteria, as well as time and place where it occurs during a specific time period. In order to help identify the kind of crimes and crime hotspots that happen at particular times and days, we propose a categorising technique. In the one that is being proposed, machine learning techniques are used to find criminal patterns that match and to help classify them using the provided temporal and spatial data.

2. LITERATURE SURVEY

There are several types of crime that occur in various areas worldwide.Many researchers have proposed a mechanism for analysing the relationship between crime and social variables such as unemployment, earnings,level of education, and so on. Suhong Kim and Param Joshi [1] introduced the K closest neighbour algorithm (KNN) and the decision tree technique as two machine learning models for prediction. The accuracy varies between 39 and 44 percent for predicting crime patterns and identifying the type of crime.

Mr. Benjamin Fredrick David. H [2] A data mining technique was used to examine and analyse big pre-existing datasets in order to deliver more information. Cross-checking against predetermined datasets verifies the extraction of new patterns.

Shraddha S. Kavathekar [3] Association rule mining was used to predict crimes. Two machine learning techniques that have been mentioned are Deep Neural Network (DNN) and Artificial Neural Network (ANN). Using the feature level dataset, a deep neural network performs better. The prediction model was built using DNN and completely connected convolution layers, primarily for multilabelled data classification. It was created with Tenserflow, an API designed particularly for Deep Learning approaches with dropout layers. These findings show that when there are a greater number of missing variables, there is a requirement for pre-processing because crimes do not occur in the same way but instead concentrate in specific places. Chandy and Abraham [4] A random forest classifier has been demonstrated for collecting characteristics for cloud computing data processing. The extracted characteristics include the request number, user identification, expiry time, arrival time, and memory need. Following feature extraction, task load is forecasted using the trained data that was perceived during the learning process, allowing the user's request to be processed in depth. Rohit Patil, Muzamil Kacchi, Pranali Gavali, and Komal Pimparia [5] For typical patterns, an Apriori algorithm is provided, and the K-means result is used. Based on the recent increase in crime rates, the system must manage a vast number of data, which takes longer to manually assess. As a result, advanced machine learning

algorithms such as K means clustering have been used. Nikhli Dubey and Setu K. Chaturvedi [6] Imposed appropriate data mining techniques for detecting upcoming future crime. A computational process that uses machine learning techniques to classify crimes.

3. CURRENT LITERATURE

In the initial pre-processing stage, duplicate values and features are removed from the dataset gathered from open sources. A decision tree has been employed in the detection of crime trends as well as the extraction of features from massive volumes of data. It acts like the foundation for the upcoming classifier. Deep Neural Network is used to extract features from the categorised crime patterns. The prediction is used to calculate the performance for both trained and test values. Crime prediction helps in predicting the occurrence of any type of criminal behaviour in the future and helps authorities put an end to it as soon as possible.

4. LIMITATIONS

1. The previous works use the categorical values of the classifier to explain the low accuracy, thus leads to biased result for the nominal attributes with higher values.

2. The methodologies used for categorization are not ideal for areas with inaccurate data and highly valued qualities.

3. Assigning an ideal value is necessary since the classifier's value must be adjusted.

5. INNOVATIVE MODEL

The data is initially pre-processed using machine learning techniques, and then it is filtered and wrapped to get rid of redundant and unnecessary data values. reduces the dimension even more, cleaning the data. The data is then split into training dataset and test dataset, with both of these datasets being utilised to train the model. The following stage is mapping. For easy categorization, the crime category, year, month, time, date, and location are all converted to integers. Using Random Forest algorithm, the independent relationship between the characteristics is initially analysed. Then Because the criminal features are identified, we may analyse the occurrence of crime at a certain time and area. Finally, the most common offences, as well as location and temporal information, are recognised. The performance of the prediction model can be determined by computing the accuracy rate. The prediction model was built in Python and then evaluated via data analysis and a machine learning method.

6. Advantages

1. The proposed method is perfect for detecting criminal patterns while the majority of the included qualities are time and place dependent.

2. It also solves the difficulty of analysing the qualities' independent effects.

3. The best value takes into account both true and

nominal values as well as the area where there is a lack of

Information. therefore initialising it is not necessary.

4. When compared to other machine learning prediction models, the accuracy was a bit higher.

7. MODULE DETAILS

The Article consist of five modules they are Data pre-processing, Mapping, Model selection, Crime prediction, and Evaluation

1. Data pre-processing

Data gathered from open sources must first be pre-processed to prevent unnecessary violations. Because it contains an enormous amount of crime data over six years, the dataset for Denver city was selected. The machine learning technique filter and wrapper is used to find the missing integral in the supplied attribute values. Both the training of a prediction model and the efficiency of the process require data purification. Datasets are filtered for instances, and unnecessary context is eliminated. The importance of the attributes is determined in part by the filtering techniques. When choosing characteristics, the correlation with the dependent variables is taken into consideration. By building a prediction model on the feature subset, the imposed wrapper approach is utilised to ascertain how useful the feature subset is. The data is separated into test and training characteristics after pre-processing.

2. Mapping

The criminal characteristics such as crime type, date of occurrence, and order of occurrence are initially separated. Then, in order to make labelling easier, it is converted to an integer. Following further analysis, the tagged features are utilised for graph charting. Python was selected as the programming language to do the required job since it works well with machine learning.Using the matplotlib package, a graph showing the prevalence of illegal behaviour is produced. The graph can be used to plot the crime that occurs the most frequently, which aids in making predictions.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 05, May 2023, pp : 1262-1266

3. Model Selection`

In this work, the Random Forest classifier is utilised. When compared to previously composed works, the accuracy obtained was relatively high. In crime prediction, random forest algorithm can be used to analyze various factors that contribute to criminal activity, such as location, time of day, demographics, and historical crime data. By analyzing these factors, the algorithm can predict the likelihood of criminal activity in a particular area or at a particular time.The random forest classifier was implemented on the pre-processed dataset using the Scikit-learn machine learning library. It can handle large and complex of data set is resistant to overfitting, and can handle missing values and outliers. This method makes a prediction by combining multiple decision trees. To create decision trees, the algorithm randomly selects subsets of features and data points from the training data set. The individual decision trees are then integrated to create a final prediction.

4. Crime Prediction

Predicting the type of crime that might occur in a certain location at a specific time is the major objective of our research. With the random forest classifier technique, we were able to accomplish this objective with a respectable level of accuracy. Four related features must be provided to predict the type of crime. The following characteristics are required: the occurrence month, the time, place, and day of the week that the crime occurred. You can enter all characteristics together with their nominal values.

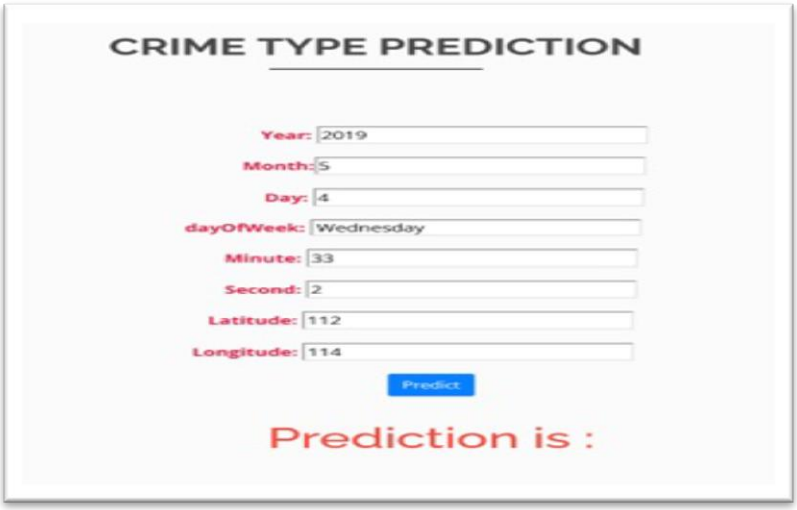

Fig 1. Predicting the type of crime

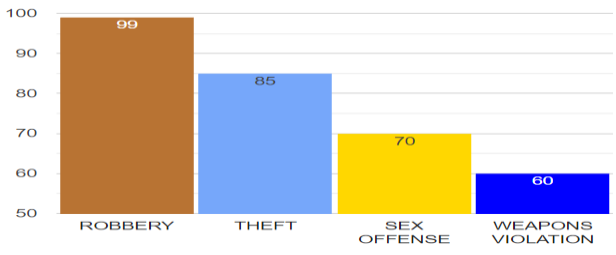

Fig 2. Identifying the most common type of crime

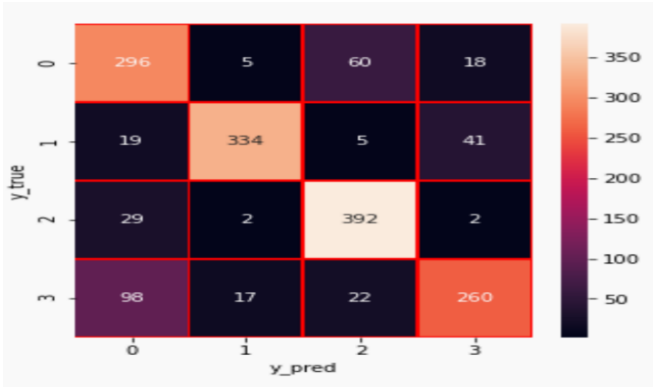

Fig 3. Confusion Matrix

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 05, May 2023, pp : 1262-1266

e-ISSN : 2583-1062

Impact Factor : 5.725

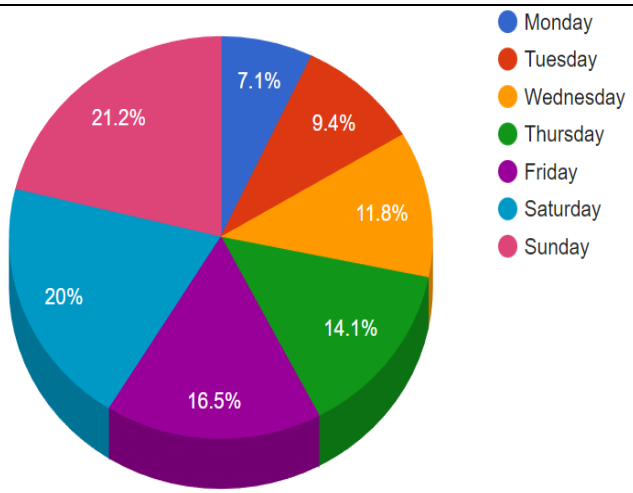

Fig 4. Percentage of crime occurrence

with particular day of week

5. Evaluation

The performance of the proposed prediction is then evaluated in order to get a high degree of accuracy when compared to the previously used model.

Cross validation is used in training to aid in training the data on a variety of training data sets. It will validate the overall divides in the assumed cross validation. To calculate the accuracy number in Python, we must provide data parameters such as the model name, target set, and cv, which aid in showing the split occurrence. Finally, the mean and standard deviation of the average precision are computed. An accuracy of 80% was obtained, which represents a significant improvement over previous prediction models.

Table 1. Performance measure for Random forest classifier

Crime Type	Recall	Precision
Robbery	67%	78%
Sex Offence	93%	84%
Theft	82%	92%
Weapons Violation	81%	65%

8. CONCLUSION

This research evaluated It is now possible to discover relationships and patterns among various sets of data using machine learning technology. The work in this project is mostly concerned with forecasting the type of crime that may occur if we know the place where it occurred. Using the notion of machine learning, we created a model with training data that had been cleaned and transformed. When compared to other algorithms, the accuracy is great. With an accuracy of 80%, the model predicts the type of crime.

Data visualisation helps in data set analysis. As a result, Random Forest Classification can be used to forecast and identify the crimes that occur the most frequently. Also, the algorithm's performance is determined using a few common criteria. When assessing an algorithm, the most important measures are average precision, recall, and accuracy. Using machine learning methods could greatly enhance the accuracy value.

9. FUTURE WORK

Future research in this area should focus on applying more categorization models to improve performance overall and improve crime prediction accuracy. It would be helpful to compare and contrast the accuracy of the various classification models for predicting the kinds of crimes that would occur.

Although socioeconomic factors might affect crime rates, it is also a helpful extension to take into account including income information for communities.

It may be feasible to learn more about the underlying causes of crime and develop targeted measures to stop it in high-risk areas by looking at the association between neighbourhood income level and crime rate. In order to minimise crime and enhance safety, governments and law enforcement organisations may find it useful to use this type of analysis to distribute resources more effectively.

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

[7] prediction using cloud computing for massive data processing systems" Journal of Information Technology 1, no. 02 (2019): 108-118.

[8] 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

[9] Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, "Computational System to Classify Cyber

[10] Crime Offenses using Machine Learning", Sustainability Journals, Volume 12, Issue 10, Published on May 2020.