ESP Journal of Engineering & Technology Advancements (ESP-JETA)
ISSN : 2583-2646
Special Issue

CRIME PATTERN DETECTION USING MACHINE LEARNING ALGORITHMS

B.K. Hemalatha
HOD/Electronics and Communication Engineering.
Mangayarkarasi College of Engineering,
Madurai, Tamil Nadu, India.

K.R. Naganandhini
B.E/Electronics and Communication Engineering,
Mangayarkarasi College of Engineering,
Madurai, Tamil Nadu, India.

T.K. Sabitha
B.E/Electronics and Communication Engineering,
Mangayarkarasi College of Engineering,
Madurai, Tamil Nadu, India.

R. Selvapriya
B.E/Electronics and Communication Engineering.
Mangayarkarasi College of Engineering,
Madurai, Tamil Nadu, India.

T.P. Vinotha
B.E/Electronics and Communication Engineering,
Mangayarkarasi College of Engineering,
Madurai, Tamil Nadu, India.

Abstract - Criminal analysis is a methodological approach for identifying, analyzing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, system can analyze previously unknown, useful information from an unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, system will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. In our system, the crime dataset of India that contains record of serious fraud of property in all states and we will apply k-means clustering which is used for partitioning a dataset into distinct, non-overlapping subgroups or clusters. By implementing the machine learning algorithms, the crime pattern can be predicted. Finally, the experimental results shows that the Accuracy, Precision, Sensitivity and Specificity. Then, we can visualize the Accuracy level in comparison map.

Keywords - Support vector machine, Random Forest, Decision Tree.

I. Introduction

A crime rate has become a topic of major concern certainly to limit the development of good governance and increasing day by day. Crimes are neither systematic nor random otherwise crime cannot be analysis. When crimes like robbery, firebombing etc. have been decreased, crimes like murder, sex abuse, gang rape etc. have been increased. We cannot analyze the victims of crime but can analyze the place where crime occurred or happened. It is difficult to analyze the data to detect crime patterns or predict future crimes by intelligence agencies or local law enforcement agencies. So, there is a need of an effective analyzing tool which can analyze crime data efficiently and quickly to give some useful crime patterns. Predictive policing means, using analytical and predictive techniques to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, we will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. The crime rates accelerate continuously and the crime patterns are constantly changing. As a result, the behaviors in crime pattern are difficult to explain. This paper illustrates how social development may lead to crime prevention. The aim is to provide a comprehensive review of theory and research with respect to the prevention of the crime in the society and to implement different data analysis algorithms which address the connections between crime and its pattern.

Crime is one of the major issues is continuing to grow in intensity and complexity. In the recent years, crime is one of the social problems influencing the nature of life and economic development in a community. Crime can be divided into a few types such as crime against properties (theft, burglary, and robbery) and crime of aggression (homicides, assaults and rape). The availability of information technologies has enabled law enforcement to collect detailed information of crime data. With the increasing numbers of crimes nowadays, crime analysis is needed which comprises measure and procedure that intend to reduce the risk of crime. Crime analysis can be done through both quantitative and qualitative methods. Qualitative approaches in predicting crime such as scenario writing or environmental scanning are valuable in identifying the future of criminal activity. Meanwhile, quantitative method is used to predict the crime rates in future specifically. Moreover, crime analysis is a practical approach to analyze and identify the pattern of crimes. Crime analysis is part of crime prevention which has the tasks of discovering and detection of crimes and their relation with criminals.

II. Existing Technology:

Crimes are increasing with a high frequency rate in this new era of world and hence it's a devastating issue that everyone has been experiencing. For finding a pattern that can be used for prediction is necessary. The objective of this paper is to understand the concept of data mining and machine learning which can be used for finding criminal patterns and behaviours. The KNN Categorization and various other algorithms will be examined for crime data prediction and prevention one with progressing accuracy will be used for implementing. The essence purpose of this project is to provide a just idea of how machine learning can be used by the law enforcement agencies to detect, predict and prevent solving crimes at much quicker rate. Disadvantage - The accuracy is low when compared with proposed solution. It does not work well with large dataset and imbalanced datasets. It has several disadvantages including, challenges in handling missing values and high storage requirements.

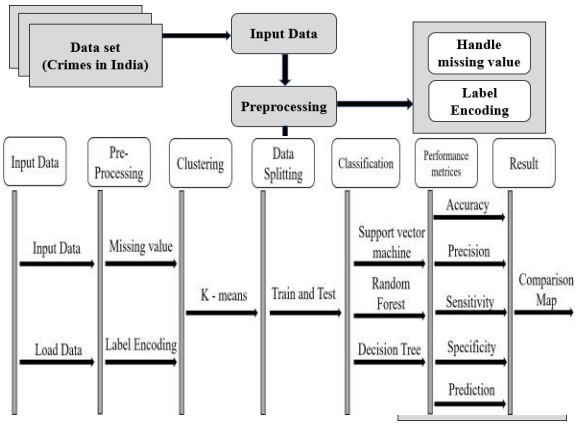
III. Proposed System

The problem of crime pattern detection involves analyzing data related to criminal activities such as fraud of property, cheating etc., to identify recurring patterns. Traditional methods of crime analysis may be time-consuming and prone to human error. The main challenge is to develop a robust algorithm that can accurately identify patterns in the data and distinguish between random occurrences and genuine trends. This will involve preprocessing the data, to handle missing values, outliers, and inconsistencies, as well as selecting appropriate machine learning models and techniques for pattern detection. In the recent past, crime analyses are required to reveal the complexities in the crime dataset. In this system, the crime in India dataset was taken as input.

The input data was taken from the dataset repository. Then, we have to implement the data preprocessing step. In this step, we have to handle the missing values for avoid wrong prediction, to encode the label for input data. Then we have to implement the machine learning algorithms such as Support Vector Machine, Random Forest and Decision tree to find the accuracy and to predict the pattern. In each algorithm, we get a highest accuracy level than the existing work. Finally, the experimental results shows that the accuracy, precision. Then, we can visualize on heat map and comparison map.

IV. Methodology

System architecture



Sequence Diagram

A. Data Selection Modules

The data selection process involves several crucial steps to ensure that the data used for analysis is relevant, reliable, and appropriate for the intended purpose. Determine what specific types of crimes or patterns you are interested in identifying. Identify potential data sources that contain relevant information about crimes. Evaluate the quality of the data from each source. Consider factors such as completeness, accuracy, consistency, and timeliness. Data quality issues can significantly impact the reliability of the analysis. Collect the necessary data from the identified sources.

This may involve obtaining permission to access certain datasets and ensuring compliance with data privacy regulations. Integrate data from multiple sources into a single dataset for analysis. This may require standardizing data formats, resolving inconsistencies, and merging datasets based on common identifiers.

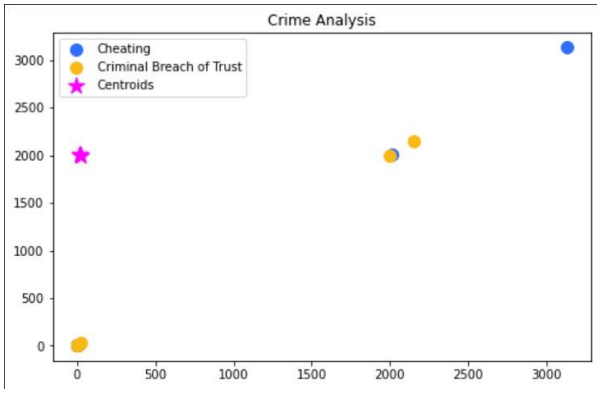
B. Preprocessing Modules

Data preprocessing is essential for preparing data in machine learning, involving operations like removing irrelevant data, transforming features, and handling missing values. The primary goals are to enhance data quality, improve model performance, and facilitate efficient analysis. Unwanted data, such as redundant features, is removed, while missing values are addressed through imputation or removal. Duplicate records are also identified and eliminated to maintain data integrity. Feature selection ensures that only relevant variables contribute to model accuracy, reducing complexity and the risk of overfitting. Finally, preprocessed data is formatted for analysis, enabling seamless integration into machine learning algorithms.

C. Clustering Modules

Clustering is a powerful technique for uncovering natural groupings within crime data. Using the K-means algorithm, we partition the data into distinct clusters based on similarity. Visualizing these clusters reveals hidden patterns and trends. K-means is favored for its simplicity and efficiency, iteratively assigning points to clusters and recalculating centroids. This method uncovers structures that might represent different offenses, locations, or time frames. Understanding these clusters aids predictive modeling, helping forecast future crime occurrences. By interpreting the clustered data, analysts inform evidence-based strategies for crime prevention, fostering safer communities.

```
------------------------------------------------
================= K means  =============
------------------------------------------------
```



D. Data Splitting Modules

Splitting data for crime pattern detection is vital in machine learning. It involves dividing the dataset into training, validation, and testing subsets. The training set is used to train the model, while the validation set helps in fine-tuning and preventing overfitting. The testing set evaluates the model's performance on unseen data. Preprocessing the dataset precedes splitting, ensuring data readiness. Features like time, location, and demographics are key. Various algorithms, like decision trees and neural networks, are employed based on the crime pattern problem. Once trained, the model's accuracy, precision, and recall are evaluated. This process ensures reliable crime prediction and detection, aiding law enforcement efforts.

```
------------------------------------------------
================= Data Splitting  =============
------------------------------------------------

Total number of rows in dataset: 448

Total number of rows in training data: 385

Total number of rows in testing data: 63
```

E. Classification Modules

Machine learning, an analytical paradigm automating model building, burgeons as a cornerstone of artificial intelligence. Its premise rests on systems learning from data, discerning patterns, and making decisions with minimal human intervention. In crime pattern detection, the selection of classification algorithms is pivotal. Support Vector Machine (SVM), adept with labeled datasets, excels in classifying crimes based on attributes, offering swift predictions even with imbalanced data. Random Forest, renowned for handling complexity and large datasets, ensures accurate predictions by balancing class distributions automatically. Meanwhile, Decision Trees, prized for their interpretability, provide transparent insights into the decision-making process, ideal for real-time crime detection. Each algorithm brings distinct advantages, empowering crime analysts with versatile tools to combat and predict criminal activities effectively. As machine learning continues to evolve, its fusion with crime analysis promises increasingly sophisticated approaches to enhance public safety and law enforcement efforts.

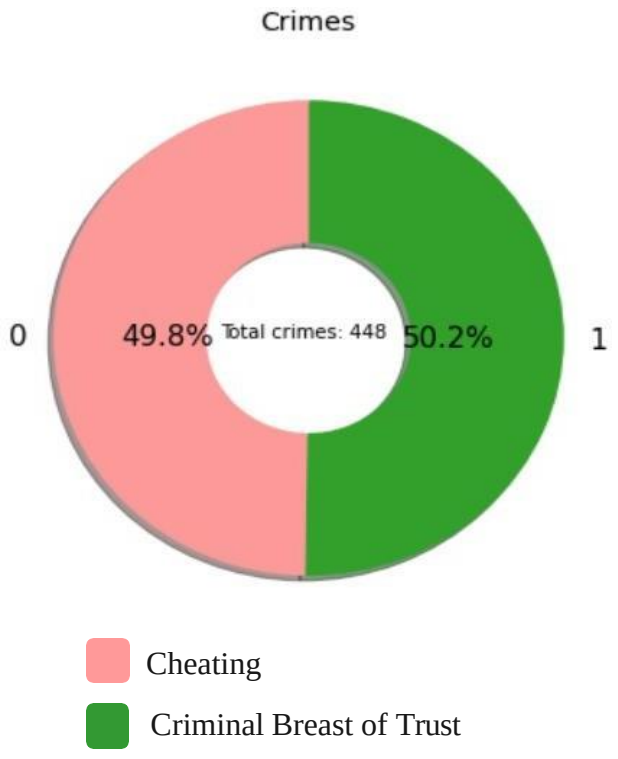
F. Performance metrices Modules

Performance metrices are essential for evaluating the effectiveness of machine learning algorithms in crime pattern detection. Choose appropriate performance metrics based on the specific objectives of the crime pattern detection task. Here, we choose accuracy. While accuracy is a commonly used metric, it might not always be the most suitable, depending on the specific objectives of the task. After selecting appropriate metrics based on the objectives, the trained model can be used for making predictions on new instances of crime data. Given the relevant features of a crime event, the model categorizes it into predefined categories, aiding in crime pattern identification and analysis.

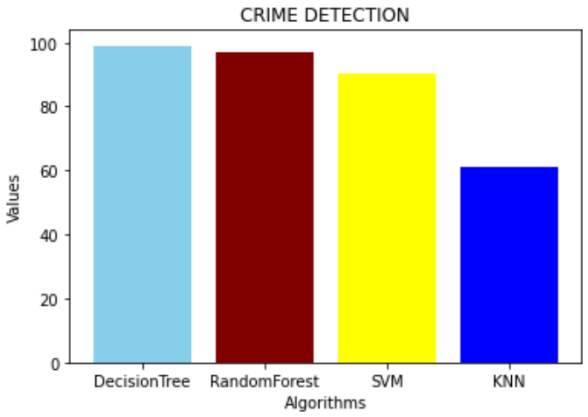
G. Visualization

The accuracy level for each algorithm can be visualized by using heat map and comparison map. Visualization plays a crucial role in understanding and communicating the accuracy levels of different algorithms. Heat maps and comparison maps are effective tools for visualizing accuracy levels. Heat maps provide a visual representation of accuracy across different regions or categories, allowing for easy identification of patterns and areas of improvement. Comparison maps enable side-by-side comparison of accuracy levels between multiple algorithms, aiding in the selection of the most suitable model for the task at hand.

EDA Diagram



Comparison Diagram



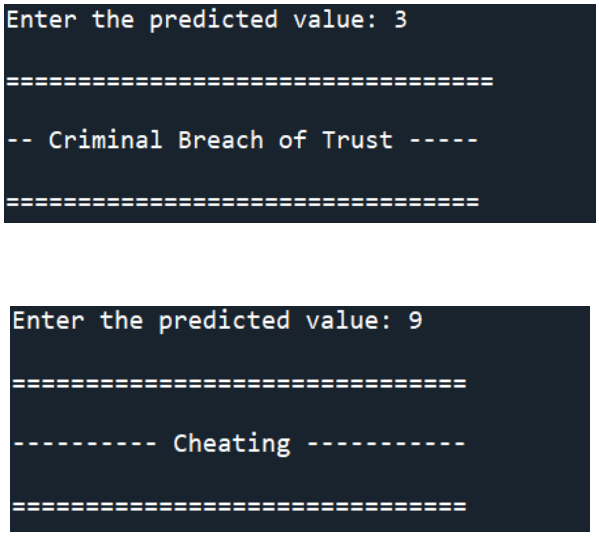
V. Results:

```
-----------------------------------------------
============== Decision Tree  ===============
-----------------------------------------------


1. Confusion Matrix : [[32  0]
 [ 0 31]]

2.Accuracy    : 99.02127659574468 %

3.Precision   : 81.79365079365078 %
-----------------------------------------------
============== Random forest  ===============
-----------------------------------------------


1. Confusion Matrix : [[30  0]
 [ 0 29]]

2.Accuracy    : 97.04545454545455 %

3.Precision   : 79.84745762711864 %

4.Sensitivity : 100.0 %

5.specificity : 79.0 %
```

```
 [ 0 29]]

3.Accuracy    : 90.06024096385542 %

3.Precision   : 75.2962962962963 %

4.Sensitivity : 100.0 %

5.specificity : 79.0 %
```

Decision tree, Random Forest and SVM

(Accuracy, Precision, Sensitivity, Specificity)



Prediction

VI. Conclusion:

With the advancement in technologies that are coming recently in data science and especially in machine learning, it becomes easy and efficient to discover patterns and information which might get useful for future prediction in crime analysis and behavior segmentation. Clustering is the process of grouping similarities in a dataset so that it can get useful for analysis, discovering patterns and prediction.

VII. Future scope:

As a future work, it would be interesting to evaluate the performance of some unsupervised algorithms. Furthermore, we applied various deep and machine learning algorithms independently from each other. In the future, we should like to combine different machine learning and deep learning algorithms as a multi-layered model to improve the detection performance.

VIII. Reference:

1. Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225-230). IEEE.
2. Suhong Kim ; Param Joshi ; Parminder Singh Kalsi ; Pooya Taheri. Crime Analysis Through Machine Learning, 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).
3. Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT
4. Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on (pp. 41-44). IEEE.
5. Suhong Kim, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri Fraser. Crime Analysis Through Machine Learning, November 2018, International College, Simon Fraser University.
6. Alkesh Bharati, Dr Sarvanaguru RA.K, Crime Prediction and Analysis Using Machine Learning, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 09 | Sep 2018.
7. Arpita Nagpal, Aman Jatain, Deepti Gaur. Review based on Data Clustering Algorithms, Proceedings of 2013 IEEE International Conference on Information and Communication Technologies (ICT 2013).
8. Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi. Analysis and Prediction of Crimes by Clustering and Classification, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, vol 4, No.8, 2015.
9. Lawrence McClendon and Natarajan Meghanathan*. USING MACHINE LEARNING ALGORITHMS TO International (pp. 1-5). IEEE.

ANALYZE CRIME DATA, Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015.

10. Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp. 406-412). IEEE.

11. N. Tyagi, A. Rana, "Fuel your growth with integration: Hybrid cloud computing", in International Journal of Applied Engineering Research, Vol. 10, Isuue 13, pp 32761-32762 (2015).

12. A. Singh, A. Rana, J. Ranjan, "Proposed analytical customer centric model for an automobile industry", in International Journal of Data Mining, Modelling and Management, Vol. 7, Isuue 4, pp 314-330 (2015).

13. A. Singh, A. Rana, J. Ranjan, "Data mining techniques and its effect in customer relationship management", in International Journal of Data Analysis Techniques and Strategies, Vol. 7, Isuue 4, pp 406-427 (2015).