iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37323

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Crime Pattern Analysis, Visualization and Pridiction Using Data Mining

H S Tanvi Srikanth[1], Mohammed Ammar Shariff[2], Nakul B P[3], Yashwanth N[4], Prof. Vijay C P[5]
[1, 2, 3, 4]*Dept. OF Information Science & Engineering, Vidyavardhaka College of Engineering, Mysuru, India*
[5]*Assistant Proffesor, Dept. OF Information Science & Engineering, Vidyavardhaka College of Engineering, Mysuru, India*

Abstract: ***Crime against women these days has become problem of every nation around the globe many countries are trying to curb this problem. Preventive are taken to reduce the increasing number of cases of crime against women. A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analyzing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in police department. Study of crime data can help us analyze crime pattern, inter-related clues& important hidden relations between the crimes. That is why data mining can be great aid to analyze, visualize and predict crime using crime data set. Classification and correlation of data set makes it easy to understand similarities & dissimilarities amongst the data objects. We group data objects using clustering technique. Dataset is classified on the basis of some predefined condition. Here grouping is done according to various types of crimes against women taking place in different states and cities of India. Crime mapping will help the administration to plan strategies for prevention of crime, further using data mining technique data can be predicted and visualized in various form in order to provide better understanding of crime patterns***
Keywords: Crime prediction, Decision trees, Linear Regression, k-means)

I. INTRODUCTION

India is a vast country with diversified societies. Position of women has been of great importance since ancient times in Indian culture. Unfortunately, current scenario depicts a different story. According to National Crime Records Bureau, crime against women has significantly increased in recent years. It has become the most prior to the administration to enforce law & order to reduce this increasing rate of the crime against women. This is where criminology comes into picture. Criminology is scientific study of crime and criminal behaviour in order to detect crime characteristics. Use of data mining techniques can produce important results from crime dataset. The very step in study of crime is crime analysis. Crime analysis is exploring, inter relating and detecting relationships between various crimes and characteristics of crimes. Police department maintains crime data at the record. This data contains huge amount of data set with complex relationships which needs use of data mining techniques in order to be transformed into useful information. The knowledge extracted from the dataset can be a great tool & support to the police department to prevent crimes. An ideal crime analysis tool should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action. However, in the present scenario, the following major challenges are encountered.

A. Increase in the size of crime information that has to be stored and analysed. Problem of identifying techniques that can accurately and efficiently analyse these growing volumes of crime data Different methods and structures used for recording crime data.
B. The data available is inconsistent and are incomplete thus making the task of formal analysis a far more difficult.
C. Investigation of the crime takes longer duration due to complexity of issues all the above challenges motivated this research work to focus on providing solutions that can enhance the process of crime analysis for identifying and reducing crime in India.

The main aim of this project consist of developing analytical data mining methods that can systematically address the complex problem related to various form of crime. Thus, the main focus is to develop a tool that assists the police in crime analysis.

1) Detecting crime patterns and perform crime analysis
2) Provide information to formulate strategies for crime prevention and reduction
3) Identify and analyse common crime patterns to reduce further occurrences of similar incidence

The present work proposes the use of an amalgamation of data mining techniques that are linked with a common aim of developing such a crime analysis tool. For this purpose, the following specific objectives were formulated.

a) To develop a data cleaning algorithm that cleans the crime dataset, by removing unwanted data to explore and enhance clustering algorithms to identify crime patterns from historical data
b) To explore and enhance classification algorithms to predict future crime behaviour based on previous crime trends
c) To develop anomalies detection algorithms to identify change in crime patterns.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

Akoglu, Leman, and Christos Faloutsos [1], Since crime is a growing concern in every part of the world it is very essential to find techniques to reduce it and also enable the police officials to easily catch the culprits. There are many approaches in solving crimes faster and a lot of researches are going on to find the best technique in data mining. The authors of this paper developed a new tool to track the culprits. Two algorithms Data Association and Back Propagation NN Classifier are used to analyse the data stored in the database. In order to extract criminal relations from an incidents summary and to create a group of suspects two approaches are used; With the help of BPN Classifier and Data Association algorithm the network is partitioned into subgroups and the interaction pattern is studied. The results prove that BPN Classifier is very accurate in identifying the crime patterns and also for future predictions. Bengio, Yoshua, Nicolas Boulanger- Lewandowski, and Razvan Pascanu [2], ACDCI (Crime Detection and Criminal Identification) technique was used to fasten the process of detecting the crimes in our Indian cities. In this technique the criminals were identified based on features like suspects name, sex, origin, facial features, crime reason, location, weapon used, etc. It had six main modules data extraction, pre-processing, clustering, map representation, and classification and WEKA tool. K-means algorithm was used for crime detection and it generated two clusters of crime. The KNN classification was used for criminal identification. The combination of k-means and KNN helped in improving the filtration for large databases. Bergstra, James, and Yoshua Bengio [3], The authors focused on the day-to-day factors rather than the causes for crime occurrences like the culprit's background or political enmity. The proposed system can predict the regions with high crime occurrences and also visualize those regions. The system will help the investigating officials to resolve crimes faster. The steps followed in this approach are data collection, classification, pattern prediction and visualization. Bayes theorem is used for classification and by using this algorithm the news articles were trained and the model was built. Apriority algorithm helps in finding the frequent patterns of a particular region. The system developed predicts crime regions in India on a particular day. Boureau, Y-Lan, Jean Ponce, and Yann LeCun [4], The paper concentrated on analysing the approaches between Computer Science and Police department as one of the main application of data mining. Pattern detection technique has been implemented and suggestion for future prediction is also included. K-means algorithm is used for clustering and this will help in identifying the patterns of crime and hence, will help in solving crimes faster. In order to increase the accuracy of prediction semi-supervised technique is used. The crimes are represented using Geo-spatial spots. Based on the selection of time range, type of crime and geographical region the results are shown graphically. Chang, Yi-Chun [5], The authors used algorithms like Naïve Bayesian, K-nearest Neighbour and Neural Networks (Multilayer-Perceptron) and proved that it is better than Decision tree and Support Vector Ma-chine. Two different feature selection methods are tested on the dataset. Comparison of algorithms are carried out on the basis of Area under Curve (AUC). The Chi-square feature selection technique is used for improving the performance of data mining re-salts. KNN gives better results by using Chi-square feature selection technique. The dataset chosen is categorised into two different types

III. EXISTING SYSTEM

These are enhanced query based layered approach for detecting SQL attacks and other web attacks, which is built from the ground up. Keeping in mind the various intricacies involved in such kind of attacks. The proposed system uses some knowledge base and query generation using the history of previous attacks and some java script feature which is determined by the user access level. Using the knowledge base detection of SQL attacks can be easily performed, it also maintains a list of some keywords, which make it easier to detect a large number of attacks at a faster pace the steps followed by in the proposed detection approach to identify the SQL injections above algorithm validates the input. That enhanced query based layered approach, for detecting SQL attacks and other web attacks. that enhanced query based layered approach, for detecting SQL attacks and other web attacks, which make it string using the initial knowledge base which stores all the frequent sql attacks of each category and is managed by the probabilistic approach if the new input string pattern already stored in the base, then it is declared as an sql attacks and a warning message will be generated automatically.

A. Drawbacks

1) Data quality can vary.
2) Limits to question that can be asked and thus can be answered by further analysis.
3) Comparisons between countries may not be valid because of the different ways data is collected and crime will be detected.
4) Data may be missing in part for some years.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

IV. PROPOSED SYSTEM

The proposed application helps in reducing crime against women using linear regression algorithm. This project gives a prediction crime rate in various cities. The user can predict the crime rate before booking the tickets. The user can book the tickets to the location where crime rate is less the manager update the locations, the user books the ticket the manager tracks the availability of tickets and confirms the user ticket booking and the police can view the tickets confirmed users. In this system admin adds managers, view user bookings.

A. Advantages

Predicting crime before it takes place.
Understanding crime pattern.

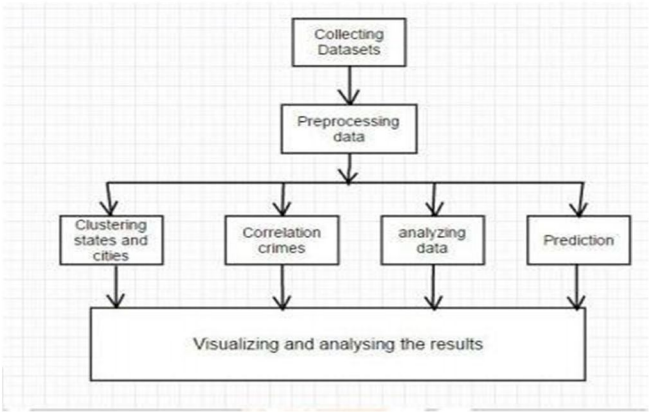
V. METHODOLOGY

The proposed application helps in reducing crime against women using data mining techniques in various cities. It is the technical challenges in designing and implementing the planner. First, the transformation operations are composable. The order of applying transformation operations also matters for performance and cost optimizations. The searching space for an optimal transformation sequence is huge. Second, the optimization is an online process and should be lightweight. This approach should find a good balance between the quality of the transformation sequence and the runtime overhead of the planner. Due to the huge space, a thorough exploration of the optimization space is impractical. Third, the planner should be able to handle different tradeoffs on the monetary cost and performance goals. The proposed concept is implemented using k-mean algorithm and implemented in 4 modules.

A. Algorithm Used Are

1) *Linear Regression:* It is simple form of regression. Linear regression attempts to model the relationship between the two variables by fitting a linear equation to observe the data. this is widely used in statistics. For this purpose linear functions are used for which the unknown parameter i.e., weight of the independent variables, are estimated from the training data.this can be used to predict the values One of the most common estimating method is least mean square. Linear regression algorithms for predicting include simple regression multiple regression and pace regression, which is suitable for data of high dimensionality and only accepts binary nominal attributes.
2) *Decision Tree:* Decision tree is used for both the prediction and classification. for the classification purpose a function can be learned this is intervals defined by splits on the individuals attributes value.
3) *K-Means Algorithm:* K –means is the simplest and most commonly used portioning algorithm among the clustering algorithm in scientific and industrial software.Acceptance of k means is mainly due to its being simple .This algorithm is also suitable for clustering of a large datasets since it has much less computational complexity grows linearly by increasing of the data points. Advantages of the k-means algorithm are relatively simple to implement, Scales to large dataset, Guarantees convergence, easily adapts to new examples. Disadvantages of the k-means algorithm are Choosing manually, Being dependent on initial values, clustering data of varying sizes and density.

VI. IMPLEMENTATION



©IJRASET: All Rights are Reserved

1) *Data Collection:* The data collection is first methodology in crime analysis. Data's are collected from various different websites, news sites and blogs. The collected data is stored into database for further process. This is unstructured data and it is object oriented programming which is easy to use and flexible.
2) *Classification:* In this step use Naive Bayes Algorithm which is supervised learning method. Naive Bayes classifier is a probabilistic classifier which when given an input gives a probability distribution of set of all classes rather than providing a single output. One of the main advantages of the Naïve bayes Classifier is simple, and coverage quicker than logistic regression [2].Compare to other algorithm like SVM (Support Vectormachine) which takes lots of memory. Using naïve Bays algorithm is create a model by training crime data related to vandalism, murder, robbery, burglary, sex abuse, gang rape, etc. Naive Bayes is that works well for small amount of training to calculate the classification parameter. Estimating probability sometimes while checking a probability P(A) * P(B/D) *P(C/D) * P(E/D) where P(C/D)=0[2].
3) *Pattern Identification:* A third step is the pattern identification where we have identify trends and patterns in crime. For finding crime pattern that occurs frequently we are using apriori algorithm. Apriori can be used to determine association rule which highlight general trends in the database. By using pattern identification it will helps to the police officials in an effective manner and avoid the crime occurrences in particular place by providing security, CCTV, fixing alarms etc.
4) *Crime Prediction:* The second Approach is predicting the crime type that might occur in a specific location within particular time. To predict an expected crime type is provide four related features of the crime. The features are: occurrence month, the occurrence day of the week, the occurrences time and the crime location. Prediction is stating probability of an event in future period time. A Classification approach is used crime prediction in data mining classify areas into hotspots and cold spots and to predictive an area will be a hotspot for residential burglary. Variety of classification techniques are used for predicting the crime:-

a) K-Nearest Neighbor (k-NN)
b) Decision trees (J48)
c) Support Vector Machine (SVM)
d) Neural Networks
e) Naïve Bayes and ensemble learning

Linear Regression methods are also used for predicting the crime prediction. Based on the crime probability. The formula for a regression line is
Y=aX + b where, Y is the predicted score, b is the slope of the line, and A is the Y intercept. b = r sx\sy And the intercept (A) can be calculated as A=MY –bMX.

5) *Visualization:* The crime prone area can be graphically represented using a heat amp which indicates level of activity, dark colour indicates low activity and brighter colour indicates the high activity.

VII. CONCLUSION

As a future extension of our work, we plan to apply more classification models to increase crime prediction accuracy and to enhance the overall performance.
It is also a helpful extension for over study to consider the income information for neighborhoods in order to see if there are relationships between neighborhoods income level and their crime findings. Furthermore, we want to study other crimes datasets from ne cities along with their demographics datasets.

VIII.FUTURE ENHANCEMETS

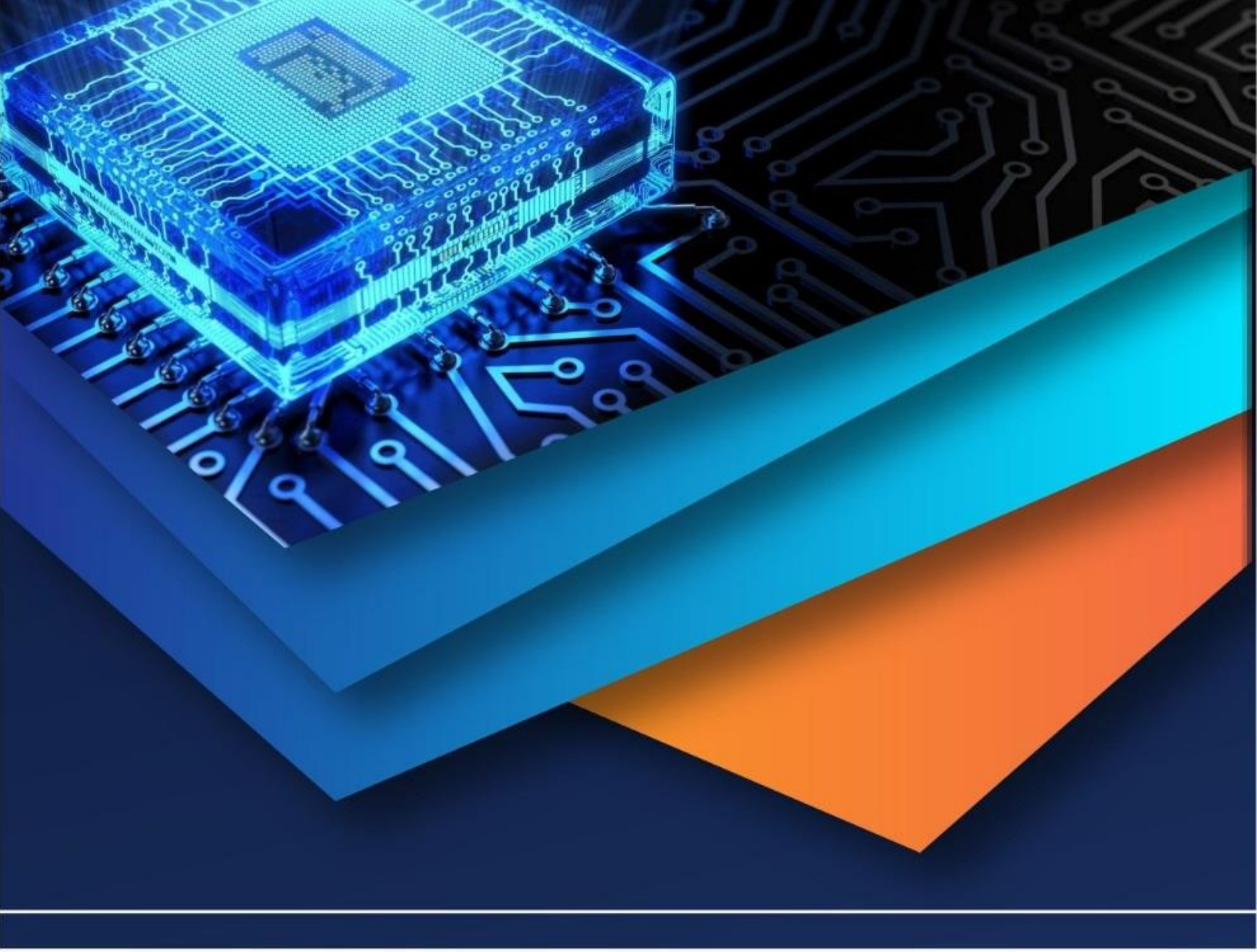
Nothing is perfect in this world. Therefore, we are also no exception. Although, we have tried our best to present the information effectively, yet, there can be further enhancement in the Application. We have taken care of all the critical aspects, which need to take care during the development of the Project. Like the things, this project also has some limitations and can be further enhanced.
The current system is predicting the crime and avoid the user from travelling to such places where crime rate is more in future in this protection is also provide where the police can track the user location.

©IJRASET: All Rights are Reserved

REFERENCES

[1] Akoglu, Leman, and Christos Faloutsos.(2015) "Anomaly, event, and fraud detection in large network datasets." Proceedings of the sixth ACM international
[2] Bengio, Yoshua, Nicolas Boulanger- Lewandowski, and Razvan Pascanu.(2016) "Advances in optimizing recurrent networks." Acoustics, Speech and Signal Processing (ICASSP), IEEE International journal on. IEEE.
[3] Bergstra, James, and Yoshua Bengio.(2015) "Random search for optimization." Journal of Machine Learning Research13-281-305.
[4] Boureau, Y-Lan, Jean Ponce, and Yann LeCun. "A theoretical analysis of feature pooling in visual recognition." Proceedings of the 27th international journal on machine learning.
[5] Chang, Yi-Chun (2017) "Mining the Networks of Telecommunication Fraud Groups using Social Network Analysis.".
[6] Ehab Hamdy, Ammar Adl, Aboul Ella Hassanien, Osman Hegazy and Tai-Hoon Kim,(2015) “Criminal Act Detection and Identification Model”, Proceedings of 7th International journal Advanced Communication and Networking, pp. 79-83.
[7] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training feedforward neural networks." Aistats. Vol. 9.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)