IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

Review on Crime Data Analysis and Prediction

Wasim Khan[1] and Dr. Hemant Kumar Gupta[2]
Research Scholar, Oriental University, Indore[1]
Professor, SAGE University, Indore[2]

Abstract: *Crime is one of the most important social problems in the country, affecting public safety, children development, and adult socioeconomic status. Understanding what factors cause higher crime rate is critical for policy makers in their efforts to reduce crime and increase citizens' life quality. Crime analysis and prediction is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems. In this Paper, we had done the review of previous research papers on Crime Data Analysis & Prediction as well as algorithms used in them.*

Keywords: Crime Data, Crime Data Analysis, Crime Prediction, Crime Data Visualization, Crime Prediction methods.

I. INTRODUCTION

Crime are a social nuisance and it has a direct effect on a society. Governments spend lots of money through law enforcement agencies to try and stop crimes from taking place. Today, many law enforcement bodies have large volumes of data related to crimes, which need to be processed to turn into useful information. To reduce and prevent crimes it is important to identify the reasons behind crimes, predict crimes, and prescribe solutions. Due to large volumes of data and the number of algorithms needed to be applied on crime data, it is unrealistic to do a manual analysis. Therefore, it is necessary to have a platform which is capable of applying any algorithm required to do a descriptive, predictive, and prescriptive analysis on large volume of crime data.

II. RELATED WORK

S. Hossain, A. Abtahee, I. Kashem, M. M.Hoque, and I. H. Sarker (2020) proposed system that predicts crimes by analyzing data-set that contains records of previously committed crimes and their patterns. The system stands on two main algorithms - i) decision tree, and ii) k-nearest neighbour. Random Forest algorithm and Ad boost are used to increase the accuracy of the prediction. Finally, oversampling is used for better accuracy.[1]

B. Yang ,L. Liu,H. Yu , M. Lan, Z. Wang,H. Zhou (2020) a spatio-temporal Cokriging algorithm to integrate historical crime data and urban transitional zones for more accurate crime prediction. Time-series historical crime data were used as the primary variable, while urban transitional zones identified from the VIIRS nightlight imagery were used as the secondary co-variable. The algorithm has been applied to predict weekly-based street crime and hotspots in Cincinnati, Ohio. Statistical tests and Predictive Accuracy Index (PAI) and Predictive Efficiency Index (PEI) tests were used to validate predictions in comparison with those of the control group without using the co-variable .[2]

P. Khobragade, P. Saraf, P. maidamwar, P. Thakre (2020) proposed a method to collect the datafrom the crime scene efficiently which includes log data, transactional data, physical drive data, and network data, later this collected data analyzed to find out the theft node in the network. FTK 4.0 digital forensic tool used to reduce plenty of time for data processing and later report will be produce that will be accepted tin the court of law. Authors also focuses the data collection method with in the network and reach to the faulty node and later this faulty node analyzed with all collected data for forensic analysis.[3]

S. A. Chun, V. A. Paturu, S. Yuan,R. Pathak ,V. Atluri, N. R. Adam(2019) have focused on how neuralnetworks can be advantageous in classification of crime prediction. The specific kind of neural network that has been used in the project is a deep fully connected neural network. Fully connected neural networks are suitable for problems where domain knowledge is limited and many to many relations between features are important.[4]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

M. FENG, J. ZHENG, J. REN ,A. HUSSAIN , X. LI , Y. XI , AND QIAOYUAN LIU(2019) applied BDA to criminal data where exploratory data analysis is conducted for visualization and trends prediction. Several the state-of-the art data mining and deep learning techniques are used. . The result prediced show that the Prophet model and Keras stateful LSTM perform better than neural network models, where the optimal size of the training data is found to be three years.[5]

H. K. Reddy, B. Saini , G. Mahajan (2018)applied various visualizing techniques and machine learning algorithms are adopted for predicting the crime distribution over an area. In the first step, the raw datasets were processed and visualized based on the need. Afterwards, machine learning algorithms were used to extract the knowledge out of these large datasets and discover the hidden relationships among the data which is further used to report and discover the crime patterns that is valuable for crime analysts to analyse these crime networks by the means of various interactive visualizations for crime prediction and hence is supportive in prevention of crimes.[6]

A. Bharati , Dr S. RA.K (Sep2018) Analyze dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. Authors used the technique of machine learning and data science for crime prediction of crime data set. Crime data consists of crime information like location description, type of crime, date, time, latitude, longitude. The K-Nearest Neighbor (KNN) classification and various other algorithms will be tested for crime prediction and one with better accuracy will be used for training.[7]

Dev Naomi.G, Karthigaa.M ,Keerthana.B ,Janani A (March 2017) comments that large amount of data were found to be tedious for storage and analysis. Now a day's various tools and techniques are coming to an existence to solve this problem. One of the application where huge amount of data is massively increasing is Crime which makes a huge issue for government to provide deliberate decision by following the laws and order. [8]

G. Matto and J. Mwangoka (2017)used text mining to extract crime patterns from sources of crime data outside police databases. With the help of developed patterns mining model we extracted several crimes reported in the newspapers, they mapped the distribution of the mined crimes country-wide, and with the use of FP-growth, generated association rules between the mined crimes.[9]

T. T. Nguyen, A. Hatua, and A. H. Sung (May 2017)described a crime predicting method which forecasts the types of crimes that will occur based on location and time. In the proposed method the crime forecasting is done crime data set. The method comprises the following steps: data acquisition and preprocessing, linking data with demographic data from various public sources, and prediction using machine learning algorithms.[10]

A. Babakura, N. Sulaiman, and M. Yusuf (2017) applied classification to crime dataset to predict the 'crime category' for diverse states of the crime data set. The paper compares two different classification algorithms namely - Naïve Bayesian and Back Propagation (BP) for predicting 'Crime Category' .[11]

M. Sevri, H. Karacan, and M. Ali Akcayol (May 2017)aims to reveal the relations between the attributes of independent criminal records. The association rules created by the Apriori algorithm have been used to extract the relationships between features of criminal records. The experimental results show that the association rules created by the Apriori algorithm are useful for criminal analysis. [12]

S. J. Linning, M. A. Andresen, and P. J. Brantingham (March 2016) suggest that cities that experience greater variations in weather throughout the year have more distinct increases of property offences in the summer months and that different climate variables affect certain crime types, thus advocating for disaggregate analysis in the future.[13]

Luca Venturini, Elena Baralis (November 2016) attempt an exploratory analysis of spatio temporal patterns of crime in San Francisco.They apply spectral analysis to the temporal evolution of all categories of crime, finding that many have a weekly or monthly period icity, along with other components. They show that spatial distribution has weekly patterns as well.[14]

T. Chauhan, R. Aluvalu (Jan. 2016) Primarily collected data will be distributed over geographic location and based on that clusters will be created. In second phase the created clusters are analyzed using Big Data Analytics. Finally that analyzed clusters are given to the Artificial Neural Network which will results in production of prediction pattern. That prediction pattern can be used by security authorities for allocating resources that helps in reducing crime. [15]

R. Kiani, S. Mahdavi, A. Keshavarzi (2015) classifed clustered crimes based on occurrence frequency during different years. Data mining is used extensively in terms of analysis, investigation and discovery of patterns for

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

occurrence of different crimes. They applied a theoretical model based on data mining techniques such as clustering and classification to real crime dataset.[16]

T. Almanie, R. Mirza and Elizabeth Lor(July 2015)focuses on finding spatial and temporal criminal hotspots. It analyses two different real-world crimes datasets and provides a comparison between the two datasets through a statistical analysis supported by several graphs. Then, it clarifies how we conducted Apriori algorithm to produce interesting frequent patterns for criminal hotspots. In addition, shows how we used Decision Tree classifier and Naïve Bayesian classifier in order to predict potential crime types. To further analyse crimes' datasets, the paper introduces an analysis study by combining our findings of Denver crimes' dataset with its demographics information in order to capture the factors that might affect the safety of neighborhoods.[17]

S. Chakravorty, S. Daripa (January,2015)have demonstrated effectiveness of data mining techniques like clustering on publicly available structured data in getting insights. Additionally an approach to retrieve data through web scraping from news media and necessary natural language techniques to extract meaningful information have been outlined, information not available through traditional structured source of data.[18]

ShijuS. , Devan M.S and Surya Gangadharan.S(August 2014)predicted regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes.[19]

J. Agarwal, R. Nagpal, R. Sehgal (December 2013) performed crime analysis by k-means clustering on crime dataset using rapid miner tool.[20]

Anna L. Buczak, Christopher M. Gifford (July 25, 2010)predicted the application of fuzzy association rule mining for community crime pattern discovery. Discovered rules are presented and discussed at regional and national levels. Rules found to hold in all states, be consistent across all regions, and subsets of regions are also discussed. [21]

Shyam Varan Nath (January 2007) analyzed at use of clustering algorithm for a data mining approach to help detect the crimes patterns and speed up the process of solving crime & looked at k-means clustering with some enhancements to aid in the process of identification of crime patterns.[22]

S. No	Authors	Objective	Dataset	Algorithms/ Technology/Keypoints	Accuracy
1.	Sohrab Hossain, Ahmed Abtahee,Imran Kashem , Mohammed Moshiul Hoque , and Iqbal H. Sarker (2020)	1)To comprehend the patterns of criminal activity to prevent them. 2)To use supervised learning technique to predict crimes with better accuracy.	San francisco crime data	Decision tree, k-nearest neighbor, Random Forest algorithm and Adaboost	resulted accuracy is 68.03%
2.	Bo Yang ,Lin Liu,Hongjie Yu , Minxuan Lan, Zengli Wang,Hanlin Zhou (2020)	1)To help allocate police resources for crime reduction and prevention.	Cincinnati Police Department.	ST-Cokriging, Spatio-temporal covariance modeling,	Increased the correlation coefficient by 5.4% for weekdays and by 12.3% for weekends in statistical tests,
3.	Prashant Khobragade, Pranay Saraf, Priya maidamwar, Preeti Thakre (2020)	1)To use Forensic Tool Kit 4.0 which provides remote data investigation and visualization analysis. 2)To analyze process information, service	KDD Cup dataset	SVM, KMeans Clustering, Network Analysis,File System Analysis	99.12 in case of cluster 1 and 98.24 in case of cluster 2.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

		information, driver information, network device, network information.			
4.	Soon Ae Chun, Venkata Avinash Paturu, Shengcheng Yuan,Rohit Pathak ,Vijay Atluri, Nabil R. Adam(2019)	1)To investigates the feasibility of using machine learning techniques, specifically neural networks, to make prediction on criminal behavior. 2)To challenge, data augmentation and weighted loss function is being developed to extract information from the minority classes.	booking history of criminals from 1997 to 2017	Machine Learning, Deep Learning, Neural Networks, Trajectory Analysis	99.7% accuracy to predict whether one will commit a crime or not given all the history, and 94% accuracy in predicting the level of crimes
5.	MINGCHEN FENG, JIANGBIN ZHENG, JINCHANG REN ,AMIR HUSSAIN , XIUXIU LI , YUE XI 1 , AND QIAOYUAN LIU (2019)	1) To propose a visual representation which is capable of handling large datasets and enables users to explore, compare,and analyze trends and patterns of crime incidents 2)To propose a combination and comparison of different machine learning, deep learning and time series modeling algorithms to predict trends with the optimal parameters, time periods and models.	crime data , three US cities, i.e. San-Francisco, Chicago and Philadelphia	Neural network, Time series forecasting	The results also showed that Prophet model and LSTM model performed better than traditional neural network models
6.	Hitesh Kumar Reddy, Bhavna Saini , Ginika Mahajan (2018)	1) To provide a framework for optimum analysis and prediction of crime incidents 2)To use various visualization techniques to analyze the data in a better way.	U.K. Crime data,	Classification algoritms, K-Nearest Neighbour, Naïve Bayes, Visualization techniques	There is 37.5% chance of Anti-social behaviour case and 10.7% chance of Burglary
7.	Alkesh Bharati , Dr Sarvanaguru RA.K (Sep2018)	To train a model for prediction which will be validated using the dataset and will be done using better algorithms.	National Crime Records Bureau data	KNN Classification, Logistic Regression, Support Vector Machine (SVM), Bayesian	KNN algorithms with accuracy of 0.787

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

				methods,predictive modelling.	
8.	Dev Naomi.G, Karthigaa.M ,Keerthana.B ,Janani A (March 2017)	1) Performance trade-offs of Hadoop deployment models on both physical and virtual clusters for analysing the crime data 2)Investigation of Hadoop applications provides power consumption.		Clustering, Mapreduce, Apriori algorithm Tools-Hbase,hive, hadoop	Through Grunt shell algorithm, the performance speed is increased up to 20%
9.	George Matto* and Joseph Mwangoka (2017)	1)To study, the text mining to extract crime patterns from sources of crime data.	Tanzania Police Force crime data records	FP-growth algorithm	Results showed that out of 30 regions of Tanzania, crime occurrences in 16 regions
10.	T. T. Nguyen, A. Hatua, and A. H. Sung (May 2017)	1)To predict the type of crime which can occur based on the given location and time.	Portland Police Bureau (PPB) and the public government source	Machine classifier models, gradient boosting, Support Vector Machine SVM, neural networks	With SCG, we obtained the best result of 74.02% classification accuracy . With RP, we could obtain the best result of 74.24% classification accuracy
11.	A. Babakura, N. Sulaiman, and M. Yusuf (2017)	1)To propose an improved method of classification algorithms for crime prediction.	U.K. Crime data	Naïve Bayesian and Back Propagation (BP) classification algorithms	-
12.	Mehmet Sevri, Hacer Karacan, and M. Ali Akcayol (May 2017)	1)To study the relationships between the attributes of different criminal records. 2)To establish relationships between new and old incidents	NIBRS Real crime cases in the USA in 2013 recorded by the FBI	Apriori algorithm	Minimum support value is set as 0.05 and minimum confidence value is set as 0.6. Association rule mining is performed on nearly 5 million crime cases and as a result, 300 frequent itemsets and 368 strong relations are defined.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

13.	S. J. Linning, M. A. Andresen, and P. J. Brantingham (March 2016)	1)To observe crime during certain time intervals can be distinctly observed in case of cities where the seasons are more distinct. 2)To observe types of crimes that will be more frequent in certain seasons because of their nature	-	Routine activities theory; Ordinary Least Squares (OLS) Regression ; Binomial Regression	Cities that experience greater variations in weather throughout the year have more distinct increases of property offences in the summer months and that different climate variable.
14.	Luca Venturini, Elena Baralis (November 2016)	1)To analize spatio temporal patterns of crime in San Francisc. 2)To apply spectral analysis to the temporal evolution of all categories of crime	Crime data of San Francisco	Lomb-Scargle periodogram	Results can improve our understanding of the dynamics of crime and be exploited to design predictive models for policing.
15.	Tirthraj Chauhan, Rajanikanth Aluvalu (Jan. 2016)	1)Using Big data Analytics for analyzing crime related data to develop crime predictive model. 2)To clusters analyze the crime Data.	State Crime Records Bureau data	Tools used - weka tool, rapid minor tool, R tool, KNIME, ORANGE and Tanagra etc.	Accuracy of Artificial Neural Network is normally very high as compared to other systems like Fuzzy Logic Series or Bayesian Network
16.	Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi (2015)	1)Extraction of crime patterns 2)Prediction of crimes based on spatial distribution of existing data, 3)Crime recognition	crimes recorded by police in England and Wales	clustering; classification; genetic algorithm; weighting; rapidminer; Clustering by K-means Algorithm	Author has done the analysis and not performed any experimental result
17.	Tahani Almanie, Rsha Mirza and Elizabeth Lor (July 2015)	1)To find spatial and temporal criminal hotspots using a set of real-world datasets of crimes. 2)To predict what type of crime might occur next in a specific location within a particular time	Crimes data in two cities of the US	Apriori Algorithm; Naïve Bayesian Classifier; Decision Tree Classifier;	Achieved 51% of prediction accuracy in Denver and 54% prediction accuracy in Los Angeles

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

18.	Subhayu Chakravorty, Souparno Daripa (January,2015)	1) To demonstrate effectiveness of data mining techniques like clustering on available structured data	Data from National Crime Records Bureau	Big Data, NER, Topic Modeling, Web-scraping, Clustering	The focus of the paper has been kept on Murder (IPC 302) and Culpable Homicide (IPC 304, 307 and 308), which can be extended for other major crimes.
19.	ShijuSathyadevan, Devan M.S and Surya Gangadharan.S (August 2014)	1) to identify the trends and patterns in crime 2)To determine association rules highlighting general trends in the database	Web sites, news sites, blogs, social media, RSS feeds	Apriori algorithm; naïve Bayes algorithm	Naive Bayes algorithm gave 90%accuracy.
20.	Jyoti Agarwal, Renuka Nagpal, Rajni Sehgal (December 2013)	1. Extraction of crime patterns by analysis of available crime and criminal data 2. Prediction of crime based on spatial distribution of existing data and anticipation of crime rate using different data mining techniques 3. Detection of crime	Crime dataset on offences recorded in England and Wales	k-means algorithm Tools-R, Tanagra ,WEKA ,KNIME ,ORANGE ,Rapid miner	Crime analysis by considering crime homicide and plotting it with respect to year and got into conclusion that homicide is decreasing from 1990 to 2011
21.	Anna L. Buczak, Christopher M. Gifford (July 25, 2010)	1) To study the application of fuzzy association rule mining for community crime pattern discovery	URL: archive.ics.uci.edu/ml/datasets/Communities+and+Crime.	fuzzy association rules, rule pruning, community-based crime	The use of relative support achieves a 95.2% reduction in the final number of rules.
22.	Shyam Varan Nath (January 2007)	1) To perform crime identification issues utilizing Data mining techniques. 2)To implement machine learning framework works with the geo-spatial plot of crime	-	Clustering, k-means, semi-supervised learning	Identified the significant attributes; using expert based semi-supervised learning method and developed the scheme for weighting the significant attributes.

Table 1. Summary of previous work

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

III. ALGORITHMS

3.1. Clustering Algorithms

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters. Let's understand this with an example. Suppose, you are the head of a rental store and wish to understand preferences of your costumers to scale up your business. Is it possible for you to look at details of each costumer and devise a unique business strategy for each one of them? Definitely not. But, what you can do is to cluster all of your costumers into say 10 groups based on their purchasing habits and use a separate strategy for costumers in each of these 10 groups. And this is what we call clustering.[22][8][6][3]

(a) K-means Clustering Algorithm: his algorithm is mainly used to partition the clusters based on their mean. As a first step number of objects are grouped and specified as k clusters. K numbers of objects are initially selected as the cluster enters. Then again these objects are assigned based on cluster centre. Then cluster means are updated again. This algorithm is used as a base for most of the other clustering algorithms.[22][3][16][20]

(b) AK-mod Algorithm: -This is a clustering algorithm consisting of two phases. In the first phase attributes are weighted. Weights of the attributes are calculated using the Information Gain Ratio (IGR) for each attribute. The attribute with the greatest value is taken as the decisive attribute. In the second phase (clustering) first the number of clusters k and initial mode of each cluster are found. Then distance for every mode and its closest mode are calculated. After that each cluster mode is updated. This process keeps going till all modes are not updated again [22].

(c) Expectation-Maximization Algorithm:- This is an extension of k-means clustering algorithm. It is used to calculate parameter estimates for each cluster. Weights of attributes are measured in probability distribution and each object is to be clustered based on the weights. To measure parameter estimates two steps are followed.

i. Expectation Step: In this step, for each object of clusters the probability of cluster membership of object x(i) is calculated.

ii. Maximization Step: Re-estimate/refine model parameters using estimation from step one.[22]

3.2. Classification Algorithms

Classification is considered as a supervised prediction technique. It has been used in many domains like weather forecasting, health care, medical, financial, etc. Two different classification algorithms are considered. They are namely Decision Tree and Naive Bayesian. Naive Bayesian is considered as an effective algorithm to solve classification tasks. Decision tree is a commonly used predictive model and it also follows supervised learning approach. As the name suggests it forms a tree like structure and each node represents a test on attribute value. Leaves represent classes that predict model for classification. Branches represent conjunctions of features. This algorithm applies a top down approach. Gain in entropy is used to guide the algorithm for creation of nodes. There are some pros and cons of both algorithms. Naive Bayesian requires a shorter training time and it has a fast evaluation. It is more suitable for real world problems. However, for complex classification problems decision tree is more suited. It produces reasonable and interpretable classification trees. This paper suggests that Decision Tree has a higher accuracy and precision over Naive Bayesian

(a) Decision Tree: Decision tree learning is a method commonly used in data mining and machine learning tasks, classification and regression. Decision trees break down a dataset into smaller subsets while at the same time an association tree is incrementally developed. Final tree has decision nodes and leaf nodes. Decision nodes have two or more branches while leaf node represents classification or decision. Decision trees can handle both categorical and numerical data. The core algorithm for building decision tree is called ID3 which employs a top-down, greedy search through the space of possible branches with no backtracking. Each partition is chosen greedily by selecting the best split from a set of possible splits, in order to maximize the information gain at a tree node. Information gain is the difference between the parent node impurity and the weighted sum of the two child node impurities. Impurity can be measured using Entropy.

C 1 p − f i log f i equation calculates the entropy. Here, C is the number of unique labels, f i is the frequency of label i at a node .

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

Advantages of using decision tree algorithm are

- It is simple to understand and interpret.
- This technique requires very little data preparation.
- Decision trees can handle both numerical and categorical data.
- Decision trees perform well with large datasets.
- It can be used to analyze large datasets within reasonable time using standard computing resources.[17][1]

(b) Random Forest Classification: Random forests are one of the most successful machine learning models for classification and regression. Random forests algorithm does not over fit. It is achieved by combining many decisions trees. One can run as many trees as needed. Random forests algorithm is considered fast. Basicalgorithm of Random forests trains a set of decision trees separately, so the training can be done in parallel. The algorithm injects randomness into the training process to make each decision tree different from each other. By combining predictions from each tree reduces the variance of predictions, improving the performance on test data. Following are some of the features of Random forests classification .

- Random forests handle categorical features and can even extend to multi class classification.
- It is able to capture non-linearity and feature interactions.
- This algorithm runs efficiently on large data bases.
- It can handle thousands of variables without variable deletion.
- Random Forest algorithm is effective for estimating missing data and maintains accuracy even when large parts of data are missing .[1]

(c) Multi-Layer Perceptrons (MLPs) Multi-Layer Perceptrons (MLPs) are classified as a type of Artificial Neural Network. The computation is performed using set of many simple unitswith weighted connections between them. Multi-layer perceptron consists of some layers. They are namely,

i. Input layer: This is the bottom most layer that takes input from a dataset. It is the exposed part of the network.
ii. Hidden layers: One or more layers which are not directly exposed to the input.
iii. Output layer: takes the output from the final hidden layer. It outputs a value or vector of values that correspond to the format required for the problem

(d) K-Nearest Neighbors (KNN):-K-Nearest Neighbors (KNN) is one of the simplest algorithms used in Machine Learning for regression and classification problem. KNN algorithms use data and classify new data points based on similarity measures (e.g. distance function).
Classification is done by a majority vote to its neighbors. The data is assigned to the class which has the nearest neighbors. As you increase the number of nearest neighbors, the value of k, accuracy might increase.[1][6][7]

(e) Support vector machines (SVMs) :-Support vector machines (SVMs) are particular linear classifiers which are based on the margin maximization principle. They perform structural risk minimization, which improves the complexity of the classifier with the aim of achieving excellent generalization performance. The SVM accomplishes the classification task by constructing, in a higher dimensional space, the hyperplane that optimally separates the data into two categories.[3][7][10]

(f) Naïve Bayes:-Naïve Bayes is a simple learning algorithm that utilizes Bayes rule together with a strong assumption that the attributes are conditionally independent, given the class. While this independence assumption is often violated in practice, naïve Bayes nonetheless often delivers competitive classification accuracy. Coupled with its computational efficiency and many other desirable features, this leads to naïve Bayes being widely applied in practice.
Naïve Bayes provides a mechanism for using the information in sample data to estimate the posterior probability P(y | **x**) of each class y, given an object **x**. Once we have such estimates, we can use them for classification or other decision support applications.[6][7][11][17][19]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

(g) Neural network:-A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so the network generates the best possible result without needing to redesign the output criteria. The concept of neural networks, which has its roots in artificial intelligence, is swiftly gaining popularity in the development of trading systems.[4][5][10]

3.3 Prediction Algorithms

a. **Apriori algorithm: Apriori algorithm** is given by R. Agrawal and R. Srikant in 1994 for finding frequent itemsets in a dataset for boolean association rule. Name of the algorithm is Apriori because it uses prior knowledge of frequent itemset properties. We apply an iterative approach or level-wise search where k-frequent itemsets are used to find k+1 itemsets.

To improve the efficiency of level-wise generation of frequent itemsets, an important property is used called *Apriori property* which helps by reducing the search space.

Apriori Property

All non-empty subset of frequent itemset must be frequent. The key concept of Apriori algorithm is its anti-monotonicity of support measure. Apriori assumes thatAll subsets of a frequent itemset must be frequent(Apriori propertry). If an itemset is infrequent, all its supersets will be infrequent. [8][12][17][19]

Forecasting:-Forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends. A commonplace example might be estimation of some variable of interest at some specified future date. Prediction is a similar, but more general term. Both might refer to formal statistical methods employing time series, cross- sectional or longitudinal data, or alternatively to less formal judgmental methods. Usage can differ between areas of application: for example, in hydrology the terms "forecast" and "forecasting" are sometimes reserved for estimates of values at certain specific future times, while the term "prediction" is used for more general estimates, such as the number of times floods will occur over a long period.

Risk and uncertainty are central to forecasting and prediction; it is generally considered good practice to indicate the degree of uncertainty attaching to forecasts. In any case, the data must be up to date in order for the forecast to be as accurate as possible. In some cases the data used to predict the variable of interest is itself forecast.[5]

IV.METHODOLOGY

There are five step for data visualization of crime data.

1. Data Collection: In this step we are collecting the data from various resources like new site, blogs, social media etc. .These collected data is stored into data base for future use and it is unstructured data. In this we can use Object Oriented Programming which is easy to use and flexible.

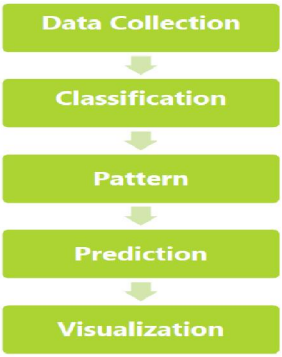

Figure 1 Basic Step of Data Visualization

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

2. Classification: In this step we use different Algorithms which classifies a news article into a crime type to which it fits the best & works well for small amount of training to calculate the classification parameters .The concept of Named Entity Recognition(NER) is used to find and classify elements in text into predefined categories such as person names, organizations, locations etc.. by using this concept we can gather more information details about related crimes.

3. Pattern Identification: In this step we have to identify trends and patterns in crime for this step we are using Apriori Algorithm, it is used to determine association rules which highlight general trends in the database. In this step we identify the crime pattern for a particular place and corresponds to each location we take the attributes and predict some pattern to that particular place. when a new case comes it follows the same pattern and says that this area has a chance for crime occurrence .By using this pattern identification it will helps to the police officials in an effective manner and avoid the crime occurrence in particular place by providing security, CCTV,fixing alarms etc.

4. Prediction: In this step we are using the decision tree concept. It is simple to understand and interpret and it works well with large datasets. It is similar to graph in each internal node represents test on an attribute and each branch represents outcome of test.
The tree has three types of nodes:
(i)Root Node:- It has incoming edges and zero or more outgoing edges.
(ii)Internal Node:- It has only one incoming edge and two or more outgoing edges.
(iii)Leaf Node or End Node:-It has exactly one incoming edge and no outgoing edges.

5.Visualization: Data visualization is the presentation of data in a pictorial or graphical format. It enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns. With interactive visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed.

V.CONCLUSION

This work is a theoretical study for several methods and methodologies in identification of crime and criminals which includes crimepatterns,geo location, prisoner methods , communication based methods Data collection, classification, pattern identification, prediction, Visualisation are the Methodologies used for this survey. There are several techniques & algorithms from this survey for identifying the criminals in the society and also providing the better future to live in.

REFERENCES

[1]. Sohrab Hossain , Ahmed Abtahee , Imran Kashem , Mohammed Moshiul Hoque , and Iqbal H. Sarker ,"Crime Prediction Using Spatio-Temporal Data",arXiv :2003.09322v1 [cs.Lg] 11 March 2020

[2]. Bo Yang ,Lin Liu,Hongjie Yu , Minxuan Lan, Zengli Wang,Hanlin Zhou ,"A spatio-temporal method for crime prediction using historical crime data and transitional zones identified from nightlight imagery",INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE(2020).

[3]. Soon Ae Chun, Venkata Avinash Paturu, Shengcheng Yuan, Rohit Pathak, Vijay Atluri and Nabil R. Adam." Crime Prediction Model using Deep Neural Networks",In Proceedings of dg.o 2019: 20th Annual International Conference on Digital Government Research (dg.o 2019), June 18-20, 2019.

[4]. MINGCHEN FENG, JIANGBIN ZHENG, JINCHANG REN ,AMIR HUSSAIN , XIUXIU LI , YUE XI 1 , AND QIAOYUAN LIU , "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data",Article in IEEE Access · July 2019.

[5]. Hitesh Kumar Reddy,Bhavna Saini , Ginika Mahajan,"Crime Prediction & Monitoring Framework Based on Spatial Analysis ",International Conference on Computational Intelligence and Data Science (ICCIDS 2018),Procedia Computer Science 132 (2018) 696–705.

[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.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

[7]. Dev Naomi.G ,Karthigaa.M ,Keerthana.B ,Janani A,"Big Data Prediction on Crime Detection",GRD Journals | Global Research and Development Journal for Engineering | National Conference on Computational Intelligence Systems (NCCIS'17) | March 2017.

[8]. George Matto, Joseph Mwangoka, "Detecting crime patterns from Swahili newspapers using text mining",International Journal of Knowledge Engineering and Data Mining Volume 4,Issue 2,Pages 145-156, Inderscience Publishers (IEL),2017.

[9]. Trung T. Nguyen, Amartya Hatua, and Andrew H. Sung,"Building a Learning Machine Classifier with Inadequate Data for Crime Prediction",Journal of Advances in Information Technology Vol. 8, No. 2, May 2017.

[10]. A. Babakura, N. Sulaiman, and M. Yusuf, "Improved method of classification algorithms for crime prediction,"10.1109 ISBAST.2014.7013130,January 2015.

[11]. Mehmet Sevri, Hacer Karacan, and M. Ali Akcayol,"Crime Analysis Based on Association Rules Using Apriori Algorithm"International Journal of Information and Electronics Engineering, Vol. 7, No. 3, May 2017.

[12]. Shannon J. Linning , Martin A. Andresen , Paul J. Brantingham ,"Crime Seasonality: Examining the Temporal Fluctuations of Property Crime in Cities With Varying Climates",International Journal of Offender Therapy and Comparative Criminology1–26,March 2016.

[13]. Luca Venturini, Elena Baralis,"A spectral analysis of crimes in San Francisco", Politecnico di Torino Dipartimento di Automatica e Informatica Corso Duca degli Abruzzi 24, Torino, Italy.

[14]. Tirthraj Chauhan, Rajanikanth Aluvalu, "Using Big Data Analytics for developing Crime Predictive Model",RK University's First International Conference on Research & Entrepreneurship (ICRE 2016).

[15]. 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.

[16]. Tahani Almanie, Rsha Mirza and Elizabeth Lor ,"CRIME PREDICTION BASED ON CRIME TYPES AND USING SPATIAL AND TEMPORAL CRIMINAL HOTSPOTS,International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.4, July 2015 .

[17]. Subhayu Chakravorty, Souparno Daripa,"Data mining techniques for analyzing murderrelated structured and unstructured data",Subhayu Chakravorty et al, American Journal of Advanced Computing, Vol. II (2), 47-54(January,2015)

[18]. Shiju Sathyadevan, Devan M.S,Surya Gangadharan. S,"Crime Analysis and Prediction Using Data Mining",First International Conference on Networks & Soft Computing (ICNSC), 2014.

[19]. Jyoti Agarwal, Renuka Nagpal, Rajni Sehgal ,"Crime Analysis using K-Means Clustering",International Journal of Computer Applications (0975 – 8887) Volume 83 – No4, December 2013.

[20]. Anna L. Buczak, Christopher M. Gifford, "Fuzzy association rule mining for community crime pattern discovery" , ISI-KDD '10: ACM SIGKDD Workshop on Intelligence and Security InformaticsJ uly 2010 Article No.: 2 Pages 1–10

[21]. Shyam Varan Nath ,"Crime Pattern Detection Using Data Mining",Conference Paper · January 2007 DOI: 10.1109/WI-IATW.2006.55 · Source: IEEE Xplore.