Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 172 (2020) 122–127

Procedia Computer Science

www.elsevier.com/locate/procedia

The 9th World Engineering Education Forum (WEEF - 2019)

Design and Analysis of Machine Learning Algorithms for the reduction of crime rates in India

Shraddha Ramdas Bandekar[1], C. Vijayalakshmi[2*]

[1] Mathematics division, School of Advanced Sciences, Vellore Institute of Technology, Chennai
[2]Department of Statistics & Applied Mathematics,Central University of Tamil Nadu,Thiruvarur
shraddha.ramdas2019@vitstudent.ac.in[1],viusesha2010@gmail.com[2]

Abstract

A country's economic growth is adversely affected with the ever-increasing crimes every day. It is one of the most severe issues in our society and reducing the crime rates have become an extremely important task. Hence, it is very important to identify different factors, occurrence relations of crimes and thus determining optimized way to reduce crimes rates. For this a database must be maintained which keeps record of different crimes with details related to place, time and nature so on for future reference. This research work focuses on how machine learning algorithms can be designed and analyzed to reduce crime rates in India. By the means of machine learning techniques, determining the pattern relations among huge set of data has become easier. This research mainly depends on providing a prediction on crime type that might occur based on the location where it has already taken place. Machine learning has been used to develop a model by the use of training data set that have gone through the process of data cleaning and transformation. Analysis of data set along with its characteristics can be implemented with the aid of data visualization. The various factors are being identified and captured. Risk factors are being identified and predictive measures are designed which help in keeping society safe. Various clustering algorithms, optimization algorithms and statistical analysis has been done in this work.

Keywords: National Crime Records Bureau; K Means Clustering; Bayesian Neural Network; Levenberg Algorithm; Classifier Regression; Support Vector Machine; K-Nearest Neighbor; Optimization.

1. Introduction

From time immemorial crime has been a concern of social problem. It has been a part of society even after various enforcement of laws. A huge number of crimes happen every second in different places, in different patters and in different times and the number is increasing each growing day. Crimes can be categorized into different types be it murder, theft, rape, kidnap and burglary so on. Crime is violation of law for which the accused is convicted.

10.1016/j.procs.2020.05.018

Crime is uncertain and cannot be predicted. Crime prediction is significant to determine increase or decrease in crime rate from preceding years. According to Yu et al. [12] crime analysis is done using the already existing past data to predict the time and place where the crime could take place. Standard crime prediction techniques include centrography, crime journey, routine activity theory and circle theory. Crime prediction techniques are useful in visualizing criminal network, reducing risks and increasing productivity of crime analysts. A good prediction technique provides faster crime data set evolution, helps in predicting correct place of crime and criminal as well as aids in keeping track of resources pertaining to analysis of crime. Crime analysis can be done by a fitting data mining approach, machine learning algorithms and statistical tools. Clustering is one of the methods in data mining and in this paper K-Means clustering, density-based clustering, coweb clustering and filtered clustering techniques are used to interpret data and analyze ways to reduce crime rates. The other techniques for crime prediction and analysis discussed in this paper are Bayesian neural networks, Levenberg Marquardt and scaled algorithm. Statistical analysis using ANOVA is also done in this work.

Data mining and machine learning are inter-disciplinary fields involving computers and mathematics wherein the programming is done for the system to carry out the operation. Both are highly important in detection and prevention of crime. Crime analysis involves extraction of crime patterns, prediction and crime detection. De Bruin et al [15] developed a crime styles structure using measure of distance to categorize as per the profile and involved clustering. Using rapid miner tool Agarwal et al. [11] performed k-means clustering on crime data set and identified crime trends for future precautions related to crime. Visualization techniques and a series of algorithms were used by Aravindan Mahendiran et al [14] to find the hidden human perceptions of crime in order to help the law enforcement agencies for better crime prediction. Khadim B. Swadi Al- Janabi [13] used decision tree algorithm and k- means clustering for developing a framework on criminal data analysis with major use of WEKA software. Babakura et al. [9] did a comparative study of Naïve Bayesian and Back Propagation algorithms and developed an improved approach of these classification algorithms for prediction of crime. They used measures like accuracy and precision for comparing these algorithms. Premalatha and Vijayalakshmi [1,8] used support vector machine approach and regression for controlling complexity by applying absolute value combination technique, and in the former with epsilon insensitive region of feature selection.

Classification is a predictive technique which has been applied nominal labelling of classes and is used in crime analysis, weather forecasting and finance so on. Iqbal et al. [10] used 10 fold cross validation for comparing decision tree and Naïve Bayesian algorithm in WEKA in crime prediction. Sivaranjani et al. [6] estimated performance of DBSCAN , agglomerative hierarchical and k-means clustering for prediction of crime and forecasting in the state of Tamil Nadu and visualized these algorithms in google map. They used K-Nearest Neighbour(KNN) classification for predicting crimes using data from National Crime Records Bureau. Prabakaran and Mitra [5] did a detailed study on crime analysis using genetic algorithm, rule induction and decision tree algorithm. In [7], three different algorithms namely K-Nearest Neighbour, neural networks and Parzen were built to investigate crime prediction in city of San Francisco. Krithika et al. [3] used supervised machine learning technique and performed crime rate analysis including data validation, visualization, cleaning of data and single, bi and multi variate analysis. The analysis conducted by them paved way to sensitivity analysis of parameters with accuracy calculation. In [4], Alves et al. did statistical learning and urban metric for crime rates analysis in Brazil. They used complex system, random forest and regression for the crime prediction. In [2] crime analysis was done in Vancouver using data of last 15 year. In this they used the technique of data processing which involved predictive model of machine learning, KNN and boosted decision tree and hence obtained an accuracy ranging between 39% to 44% in crime analysis. Optimization is required in every field and in crime and criminal analysis, the aims generally include minimizing crime rates, criminal hotspot optimization, financial crime optimization as well as determining criminal pattern. In this paper data is collected from National Crime Records Bureau and is analysed to minimize the crime rate level. Various clustering, classifier regression, best fit attribute selection and other statistical techniques, deep learning methods under machine learning are used in this paper. The results are based on Bayesian Neural Networks, Levenberg algorithm, scaled algorithm and statistical analysis with an intention to reduce the crime rate. The functional diagram is depicted below.

2. Models in proposed system

2.1. Predictive Modeling

Predictive modelling involves building of models which help in prediction. This includes decision trees and classification, support vector machines, KNN classification, regression, K-Means clustering, density based clustering and random forest. Decision tree is based on models of utility and outcomes for display of algorithm. KNN is an observed ML algorithm and stores instances relating to training dataset in multi-dimensional space. Support vector machine sets a hyper plane and classify data above and below providing higher level of predictivity followed by clustering. Random forest groups distinct classifications, regression and other operations giving array of decision trees at training time. K-Means clustering involves diving say N number of observations into k number of clusters in which each of the observations is a part of cluster with closest mean. Density based clustering is a ML algorithm which is unsupervised, and clusters are dense spots in the data space. Regression is used for predictions wherein one of the variables is found when other set values are given.

2.2. Functionality of proposed work

3. Procedure in crime analysis

The following steps are implemented to carry out the analysis:

1. Data collection - The data is collected from Public Domain Data- National Crime Records Bureau.
2. Data pre-processing- Raw data is transformed to required format
3. Feature selection- Murder attribute is considered as dominating crime type in this work.
4. Prediction and Visualization – Models considered are predicted and visualized
5. Interpretation – Crime rate analysis is obtained.

4. Results and Discussion

Models:	K- Means Clustering	Density Based Clustering	Input Mapped Classifier	Expectation Maximization Algorithm
	Clustered Instances 0 20 (95%) 1 1 (5%)	Clustered Instances 0 20 (95%) 1 1 (5%) Log likelihood: -74.26837	Correlation coefficient 0.5584 Mean absolute error 26.5054 Root mean squared error 54.4709 Relative absolute error 100 % Root relative squared error 100 % Total Number of Instances 21	STATISTICAL_ANALYSIS(Probability distributions) Time taken to build model (full training data) : 0.1 seconds Model and evaluation on training set Clustered Instances 0 21(100%) Log likelihood: -94.96133

Interpretations	The Clusters depends on the selection of initial centroid which is at random and convergence is achieved	The noise is identified while clustering with respect to the arbitrary size and shape.	Significantly enhanced because of mere co location of related activity.	For fitting mixed distribution which will converge to local optimum with log likelihood of the data.

4.1 Statistical Analysis

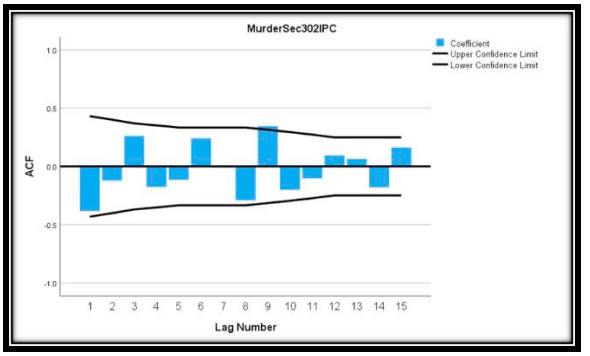


Figure1: Lag No. vs autocorrelation

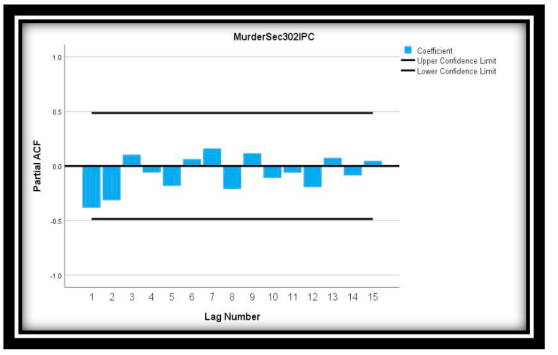


Figure2: Lag No. vs partial autocorrelation

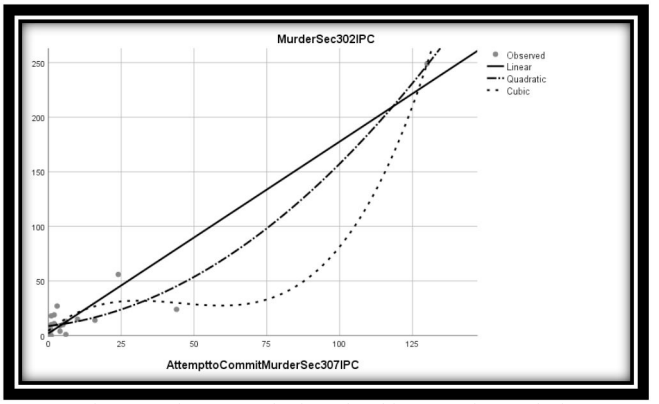
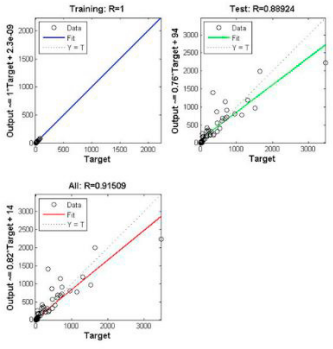
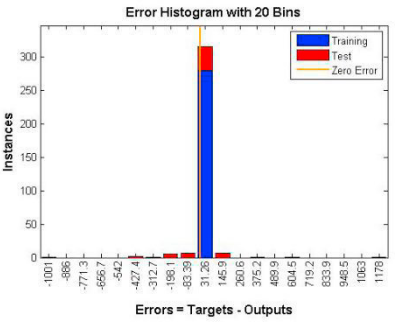
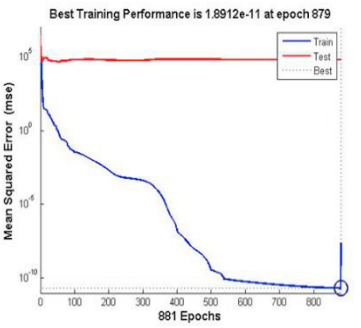


Figure 3: Linear, Quadratic and Cubic regression analysis

4.2 Bayesian Neural Networks

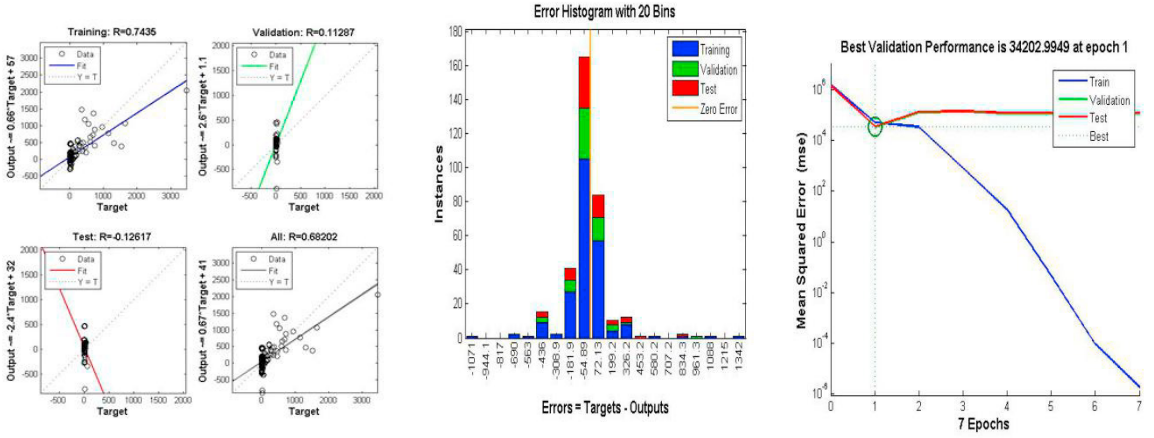






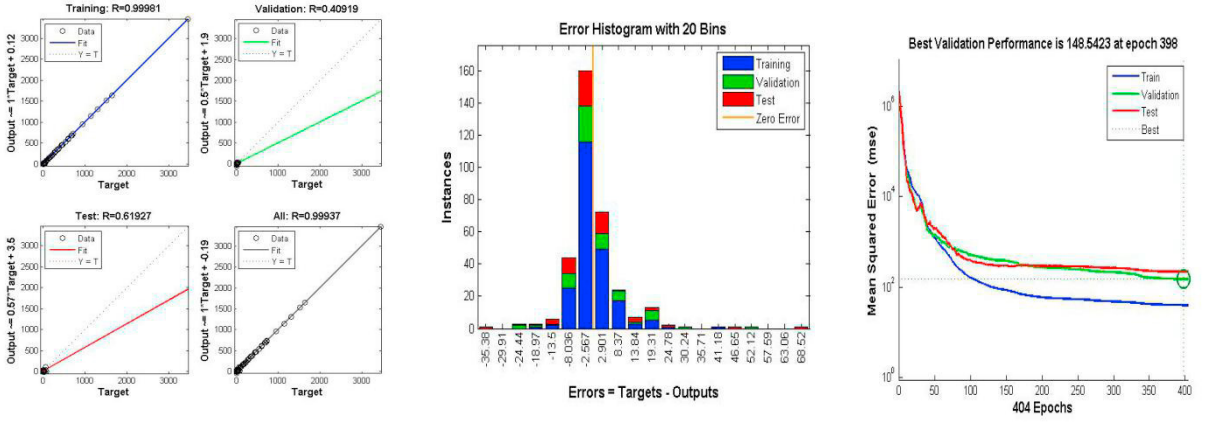
Based on the above graphical representations with respect to Bayesian Neural Networks consistent ways for inference and prediction of the quality property over fitting is obtained further facilitating the estimation of the confidence interval.

4.3 Levenberg Algorithm



The above graphs obtained by applying Levenberg algorithm infers the stability of the steepest descent method in which convergence is obtained and multiple free parameters are handled which leads to optimized value.

4.4 Scaled Algorithm



From the diagrammatic representation of scaled algorithm, it is inferred that data is normalized within a range and in data pre-processing scaling is supplied by the independent variable to perform future scaling.

5. Conclusion

In the growing research advancement, detection of crime using machine learning and data mining aims at reducing the crime rate levels. This research work concentrates on the distinct crime types, their occurrences in different places and times. Major analysis is done keeping murder as the dominating crime attribute in comparison with the rest types. Using the Bayesian, Levenberg and Scaled algorithm on train and test data we determined at which

iteration best valid performance was attained, and it was observed that for the data considered, scaled algorithm gave the best result in comparison with the other two. Statistical analysis was done based on correlation, ANOVA and graphs were obtained of which the results are discussed in previous section. The analysis obtained was the crime rate can be reduced to 78 percent implying accuracy of .78. The future scope is to extend this work to obtain an optimization model and applying it to huge data and obtaining results based on comparative analysis of different ML algorithms and other like Genetic Algorithm, deep learning algorithms.

References

[1] Premalatha, M. & Vijayalakshmi, C.. (2019). SVM approach for non-parametric method in classification and regression learning process on feature selection with ε - insensitive region. Malaya Journal of Matematik. S. 276-279. 10.26637/MJM0S01/0051.

[2] Suhong Kim, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri,(2019), "Crime Analysis Through Machine Learning",DOI: 10.1109/IEMCON.2018.8614828 Conference: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

[3] Kirthika V, Krithika Padmanabhan A , Lavanya M & Lalitha S D,(2019), Prediction of Crime Rate Analysis Using Supervised Classification Machine Learning Approach, International Research Journal of Engineering and Technology (IRJET), Vol. 6, pp. 6771-6775.

[4] L. G. A. Alves, H. V. Ribeiro, and F. A. Rodrigues,(2018), "Crime prediction through urban metrics and statistical learning", Physica A, Vol. 505, pp. 435-443.

[5] S. Prabakaran and S. Mitra, (2018), "Survey of analysis of crime detection techniques using data mining and machine learning", Nat. Conf. on Math. Techn. and its Appl. (NCMTA 2018), IOP J. of Physics: Conf. Series, Vol. 1000.

[6] Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016). Crime prediction and forecasting in Tamilnadu using clustering approaches. 2016 International Conference on Emerging Technological Trends (ICETT). doi:10.1109/icett.2016.7873764 .

[7] M. V. Barnadas, Machine learning applied to crime prediction, Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, Sep. 2016.

[8] Premalatha, M. & Vijayalakshmi, C.. (2015). SVM approach for classification and regression with absolute value combination method for controlling complexity. International Journal of Pure and Applied Mathematics. 101. 811-820.

[9] Babakura, N. Sulaiman, and M. Yusuf, (2014), "Improved method of classification algorithms for crime prediction",International Symposium on Biometrics and Security Technologies (ISBAST) IEEE .

[10] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy & N. Khanahmadliravi,(2013, March) "An experimental study of classification algorithms for crime prediction", Indian J. of Sci. and Technol., Vol. 6, No. 3, pp. 4219-4225.

[11] J. Agarwal, R. Nagpal & R. Sehgal,(2013,December),"Crime analysis using k-means clustering", International Journal of Computer Applications, Vol. 83 – No.4.

[12] Chung-Hsien-Yu, Max W.Ward, Melissa Morabito & Wei Ding. "Crime Forecasting Using Data Mining Techniques" 11th International Conference on Data Mining pp. 779-786, IEEE 2011.

[13] Kadhim B. Swadi Al-Janabi,(2011,May), "A Proposed Framework for Analyzing Crime Data Set using Decision Tree and Simple K-means Mining Algorithms",Vol. 1- No. 3, pp. 8-24.

[14] Aravindan Mahendiran, Michael Shuffett, Sathappan Muthiah, Rimy Malla & Gaoqiang Zhang,(2011), "Forecasting Crime Incidents using Cluster Analysis and Bayesian Belief Networks".

[15] Jeroen S. de Bruin, Tim K. Cocx, Walter A. Kosters, Jeroen F. J. Laros & Joost N. Kok(2006),"Data mining approaches to criminal career analysis" ,In Proceedings of the Sixth International Conference on Data Mining (ICDM06) ,pp. 171-177

.