JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CRIME RATE PREDICTION AND ANALYSIS SYSTEM USING MACHINE LEARNING ALGORITHMS

[1]K.Mahima Chowdary, [2] M. Vaishnavi, [3]A.Saketh Rao, [4]B.Hemath, [5]Dr.M.Padmaja

[1,2,3,4] Student, [5] Assistant Professor

Department of CSE, Gandhi School of Technology, Visakhapatnam-530045, Andhra Pradesh, India

Abstract: The ensemble learning approach is a cooperative dynamic technique that creates new models by using the expectations of learned classifiers. Tentatively and hypothetically, early exploration has shown that gathering classifiers are more dependable than single-part classifiers. Despite the fact that a ton of gathering techniques have been portrayed, it's still difficult to come by the best arrangement for a specific dataset. In India, various forecast-based speculations have been introduced to resolve the issue of ML crime expectation. It becomes challenging to decide the unique idea of crimes. The target of bad behavior assumption is to restrict crime rates and dissuade crime. Assemble-stacking-based crime prediction method (SBCPM), in light of SVM calculations, is introduced in this paper as a compelling genuine technique for coordinating learning-based techniques in MATLAB to choose satisfactory crime expectations. The SVM technique is utilized to make area explicit arrangements as opposed to the Random Forest, SMO Naive Bayes bagging, and another machine learning model called J48. Therefore, an entertainer model for the most part neglects to really work. The troupe model has the most noteworthy relationship coefficient and the least normal and outright blunders when contrasted with different models in certain circumstances. The proposed strategy had the option to characterize the testing information with 99.5% accuracy. It has been demonstrated that the model has greater predictive power than previous studies that focused solely on violent crime datasets. Additionally, the results demonstrated that criminological concepts and any empirical data on crime can be reconciled. The recommended strategy was likewise demonstrated to be advantageous for expecting possible violations.

Keywords: Random Tree Algorithm, K-Nearest Neighbor (KNN), Bayesian model, Support Vector Machine (SVM), and Neural Networks.

1. Introduction:

A couple of crime examiners and specialists have actually hoped to gauge and explore approaches to reducing bad behavior using different showing and genuine strategies. There might be a requirement for huge examinations that can assist policymakers and concerned divisions with resolving issues and worries in the field of techniques for anticipating and controlling crime as the pace of wrongdoing keeps on rising. Human aptitude can't monitor criminal records whenever taken care of physically. As an outcome of this, an original way to deal with distinguishing information connected with violations that will support investigation is required. The ongoing examination concerning wrongdoing expectation centers around two critical perspectives: crime area of interest estimate [3] and crime risk field forecast [1], Techniques for handling information are utilized to improve this work. Police divisions have had the option to incorporate significant datasets of various crimes, including assault, fire-related crime, and murder, because of headways in information science and the expanded accessibility of PCs. Lately, a great deal of violations has been accounted for around the world. A vicious demonstration where the culprit takes steps to utilize force against the casualty is a serious crime. It remembers the two violations for which viciousness is utilized for impulse and crimes in which a rough demonstration is a reason, similar to assault or murder. Rough violations can go from murder to badgering and

might possibly be begun with weapons, contingent upon the purview. Instances of savage violations incorporate homicide, burglary, assault, endeavored murder, capturing, robbery, riots, share killing, endowment outrage, and others. In India, 22 regions in the territory of Madhya Pradesh and 23 areas in the province of Rajasthan have extremely high paces of savage crime, while 34 and 25 states have high paces of violations against ladies, separately. In terms of the number of deaths, 17 districts in Maharashtra, followed by Andhra Pradesh and Rajasthan, are said to be among the poorest counties. having the second-highest concentrations of violent and criminal districts).

2. Propose:

The majority of crimes are committed against women and involve violence. In a similar vein, out of a total of 100 districts, 13 districts in Rajasthan, 12 districts in Bihar, and 11 districts in Tamil Nadu were found to have committed a crime of opposing public order [4]. In general, machine learning algorithms are used to make predictions about how much violence will be involved in future crimes. With restricted calculations and a little dataset, this study has been progressing for quite a while. Through observational ML examination and different commitments portrayed in this segment, this study affirms advancement. Despite the fact that ML models are broadly utilized for crime expectation, there are numerous regions where the new man-made brainpower methodology hasn't been completely investigated and has significant downsides. The most well-known calculations that have exhibited satisfactory exactness in ML classifiers are the Random Tree Algorithm, K-Nearest Neighbor (KNN), Bayesian model, Support Vector Machine (SVM), and Neural Network [5]. The magnificence of this examination exertion is featured by the idea of a crime expectation approach that consolidates different calculations into a crime forecast gathering model utilizing stowing and stacking outfit methods. A gathering model is a technique for making a prescient model by consolidating various models to resolve a solitary issue to work on prescient proficiency. Various applications have shown the worth of gathering learning techniques [6]. In outfit learning-based models, expectation gatherings are utilized to coordinate the result limit of more than one classifier to come by the eventual outcome. Single-learning classifiers' efficiency can be compared to that of ensembles.

3. Motivation:

The consequences of the information examination are introduced in this part. The data for the dataset comes from different sources, remembering criminal records and information for violations like homicide, burglary, and abducting. The organized dataset is preprocessed, and the suggested model is surveyed using a 5-fold cross-endorsement test. The strategy for testing prescient models is made by partitioning the underlying informational index into a model preparation test and a test set for assessment.

4. Proposed system:

Assemble-stacking-based crime prediction method (SBCPM), in light of SVM calculations, is introduced in this paper as a viable genuine technique for coordinating learning-based techniques in MATLAB to choose adequate wrongdoing expectations. The SVM technique is utilized to make area explicit setups rather than the Random Forest, SMO Naive Bayes bagging, and another ML model called J48. Therefore, an entertainer model by and large neglects to successfully work. The outfit model has the most elevated relationship coefficient and the least normal and outright mistakes when contrasted with different models in certain circumstances. The recommended technique had the option to group the testing information with 99.5% precision. It has been exhibited that the model has a more noteworthy prescient power than past investigations that zeroed in exclusively on fierce wrongdoing datasets. Also, the outcomes showed the way that criminological ideas and any exact information on crime can be accommodated.

Additionally, the results demonstrated that criminological concepts and any empirical data on crime can be reconciled. The recommended strategy was likewise demonstrated to be advantageous for expecting possible violations.

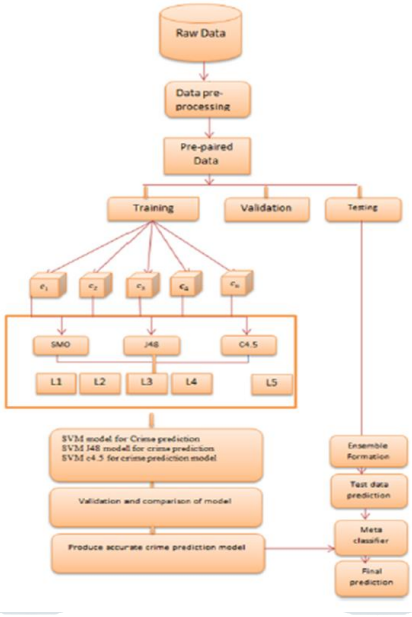


Fig1 .Proposed system architecture

We began data preparation for the raw data that needed to be prepared for the subsequent phase in this suggested system by taking raw data from the dataset. Before being divided into three major components, the preprocessed data will now be prepared. Preparing, approval, and testing are the three key parts. The training is now further broken down into levels, and the levels are broken down even further into classifiers, and testing takes place simultaneously. In the end, we get the results we want.

4.1. Data source and data preprocessing:

Gathered a dataset from the Kaggle website which is very important for the project implementation detail analyzing of the dataset is done. The dataset contains detailed important points of the crimes in India. The dataset has every states and union territories crime rate information. This dataset contains complete details on various aspects of crimes that have happened in India since 2001. The dataset is large to work so we basically classified the 57 excel sheets into 4 main crimes and also according to state wise.

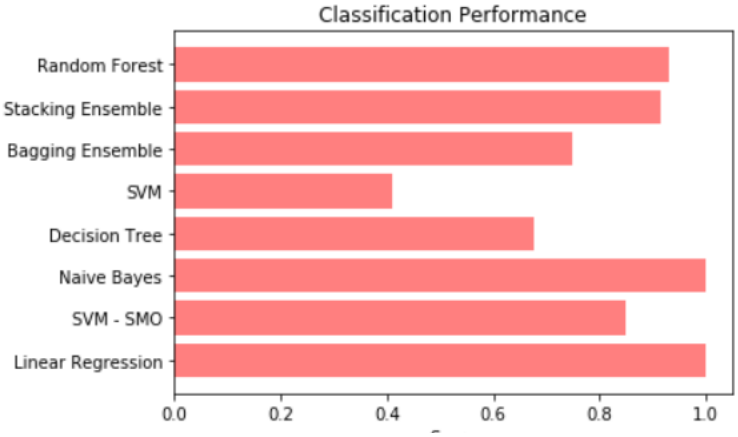
- Crime against women
- Crime against children
- IPC crimes
- SLL crimes

4.2. Modules:

Load dataset: The dataset was loaded by this module; Pre-processing of Data: The data to be processed were read by this module; also, Split information: The data were divided into train and test sections by this module.Making Models: This module was used to start making models.Make RF, LR, SVM, NB, DT, BR, and SR. Classifiers: The algorithms' accuracy is calculated with this module.Comparison of precision: The comparison of all algorithms can be seen with this module. Register Users: The user registers for the application using this module.User login: The user logs in using this module.Upload Crime Information: The user uploads input for prediction using this module.

4.3. Model accuracy graph:

In this project, we have chosen eight different algorithms to find the best model with the best accuracy. In this, we have taken every dataset and applied every algorithm, here we have done two main things using algorithms. They predicted the data and found out the accuracy of every algorithm.



5. Crime rate analysis in Indian cities architecture:

Prediction and visualization are the two important parts of this project. Let us discuss briefly them.

5.1. Predictions:

The adjusted threshold value for the predictions' accuracy is 65%, as the study of algorithms suggests that not all the standard algorithms are suitable for these predictions when considered directly. The accuracy of the predictions of certain states for a particular crime type shows drastic values which are improved by identifying the trend-changing year. This compensates for the issue of insufficient or unsuitable data being used for predictions.The data of newly established states has also been considered which may not have sufficient data for predictions. This is handled using data wrangling.This system would achieve more accuracy and functionalities as the data supplied to the system would increase.Certain cases of prediction may not seem to be ideal due to the availability of data which depends on the timely and systematic reporting of data. The lack of ideal reported data indicates that there could be:

Lack of awareness among the victims

Fear of the loss of reputation of the victims in the society

Disorganized reporting of cases

5.2. Visualizations:

The parameters considered in visualization are the crime rates per one lakh people and the literacy rates for every state.The graph of certain states shows radical progress over the years, which may be misinterpreted as the increase in the number of crimes in that state but the actual reason for this change is the increase in literacy rates of that state giving rise to the increase in awareness for reporting the crime cases.The visuals provide a broader view of the comparison of crime rates depending on the population as well as literacy of all the states together, or a fixed number of states as per the user's choice. Similarly, the progress of a single state over the years can also be observed.

6. Outcome and results:

Utilizing outfit and comparison machine learning (ML) strategies like SMO, SVM bagging, SVM-Random forest, SVM-stacking J48, and Gullible Predisposition, ML models were developed and carried out in this review. Each particular component was present in a violent crime training dataset. (e.g., robbery, murder, rape, etc.) After successfully training and evaluating six models, we came to an important conclusion.

We came to the major conclusions outlined below: 1. The SBCPM model outperforms other models in terms of prediction accuracy. This is simpler to find by taking a gander at how savage wrongdoing measurements will quite often rehash the same thing. 2. The SBCPM employs predictive label transmission with one-versus-five classifiers. In this novel labeled classification method, SB CPM machine learning classifiers at various tree levels can effectively collaborate to deny the relationship between labels and more precisely forecast labels and instance labels. The 15-year criminal ML-Tree method is contrasted with the ensemble random forest bagging classifiers J48, Naive Bayes, and SMO. 3. The suggested SBCPM model's efficacy is demonstrated by the experiments' outcomes. During the training phase, the novel SBCPM's core working time expands at a rate of 99.5%. In wrongdoings, the accurate process of selecting an acceptable method might be anticipated in such a way that the estimation is unexpected. It is also taught how to prepare for further investigations into violations and exceptions by incorporating the SBCPM approach into the selection process. The actual issues and hypotheses of the current study were formulated during the theory's construction.

7. Future work:

From this point on, the model as a whole will be transformed into an open-source library and connected to the lawbreaker website, allowing it to operate to its fullest potential. Work inside a structure that permits the limit for class crime percentages not entirely set in stone. It is possible to determine the highest crime rate rather than a limited number of crimes. To assess the viability of each of the lawbreaker characters assessed in the proposition, a little gathering is utilized. With a calculation on a sufficiently large scale of local or cloud-based crime and

large datasets, our study can be realistically improved, new options can be expanded, and the prediction of multi-label pricing can be paid for.

8. References:

[1] M. Cahill and G. Mulligan, ``Using geographically weighted regression to explore local crime patterns," Social Sci. Comput. Rev., vol. 25, no. 2, pp. 174193, May 2007, doi: 10.1177/0894439307298925.

[2] J. M. Caplan, L. W. Kennedy, and J. Miller, ``Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting," Justice Quart., vol. 28, no. 2, pp. 360381, Apr. 2011, doi: 10.1080/07418825.2010.486037.

[3] A. Almehmadi, Z. Joudaki, and R. Jalali, ``Language usage on Twitter predicts crime rates," in Proc. 10th Int. Conf. Secur. Inf. Netw., Oct. 2017, pp. 307310, doi: 10.1145/3136825.3136854.

[4] V. K. Borooah and N. Ireland, ``Deprivation, violence, and conict: An analysis of Naxalite activity in the districts of India," Int. J. Conf. Violence, vol. 2, no. 2, pp. 317333, 2008, doi: 10.4119/UNIBI/ijcv.42.

[5] A. Babakura, M. N. Sulaiman, and M. A. Yusuf, ``Improved method of classication algorithms for crime prediction," in Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST), Aug. 2014, pp. 250255, doi: 10.1109/ISBAST.2014.7013130.

[6] P. Pªawiak, M. Abdar, and U. R. Acharya, ``Application of new deep genetic cascade ensemble of SVM classiers to predict the Australian credit scoring," Appl. Soft Comput., vol. 84, Nov. 2019, Art. no. 105740, doi: 10.1016/j.asoc.2019.105740.

[7] Z. Li, T. Zhang, Z. Yuan, Z. Wu, and Z. Du, ``Spatio-temporal pattern analysis and prediction for urban crime," in Proc. 6th Int. Conf. Adv. Cloud Big Data (CBD), Aug. 2018, pp. 177182, doi: 10.1109/CBD.2018.00040.

[8] A. Almaw and K. Kadam, ``Survey paper on crime prediction using ensemble approach," Int. J. Pure Appl. Math., vol. 118, no. 8, pp. 133139, 2018. [Online]. Available: https://internal-pdf://107.93.182.66/18. pdf%0Ahttp://www.ijpam.eu

[9] T. B. Hyde, H. Dentz, S. A. Wang, H. E. Burchett, S. Mounier-Jack, and C. F. Mantel, ``The impact of new vaccine introduction on immunization and health systems: A review of the published literature," Vaccine, vol. 30, no. 45, pp. 63476358, 2015, doi: 10.1016/j.vaccine.2012.08.029

[10] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma, and N. Yadav, ``Crime pattern detection, analysis & prediction," in Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Apr. 2017, pp. 225230, doi: 10.1109/ICECA.2017.8203676.

[11] N. H. M. Shamsuddin, N. A. Ali, and R. Alwee, ``An overview on crime prediction methods," in Proc. 6th ICT Int. Student Project Conf. (ICT- ISPC), May 2017, pp. 15, doi: 10.1109/ICT-ISPC.2017.8075335.

[12] S. S. Roy, V. M. Viswanatham, P. V. Krishna, N. Saraf, A. Gupta, and R. Mishra, ``Applicability of rough set technique for data investigation and optimization of intrusion detection system," in Proc. Int. Conf. Heteroge- neousNetw. Quality, Rel., Secur. Robustness, in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 115, 2013, pp. 479484, doi: 10.1007/978-3-642-37949- 9_42.

[13] L. Acion, D. Kelmansky, M. D. Van Laan, E. Sahker, D. S. Jones, and S. Arndt, ``Use of a machine learning framework to predict substance use disorder treatment success," PLoS ONE, vol. 12, no. 4, pp. 114, 2017, doi: 10.1371/journal.pone.0175383.

[14] V. Ingilevich and S. Ivanov, ``Crime rate prediction in the urban environment using social factors," Procedia Comput. Sci., vol. 136, pp. 472478, Jan. 2018, doi: 10.1016/j.procs.2018.08.261.

[15] G. R. Sinha, K. S. Raju, R. K. Patra, D.W. Aye, and D. T. Khin, ``Research studies on human cognitive ability," Int. J. Intell. Defence Support Syst., vol. 5, no. 4, pp. 298304, 2018.