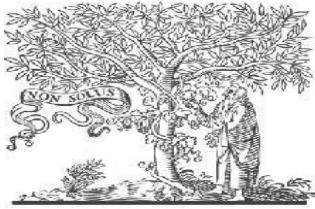




COPY RIGHT



2021 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 12th June 2022.

Link : <https://ijiemr.org/downloads/Volume-11/Issue-06>

Title: Twitter CRIME PREDICTION USING RANDOM FOREST ALGORITHM

volume 11, Issue 06, Pages: 1438-1447

Paper Authors: **Mrs.N.V.MahaLakshmi**, **S.Tejaswini**, **Ch.Jayapriya**, **Ch.DivyaSudha**, **Ch.TagoreBabu**



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

CRIME PREDICTION USING RANDOM FOREST ALGORITHM

¹Mrs.N.V.MahaLakshmi M.Tech, (Ph.D.),

²S.Tejaswini,³Ch.Jayapriya,⁴Ch.DivyaSudha,⁵Ch.TagoreBabu

¹Associate Professor, CSE, PSCMR CET, Vijayawada, Andhra Pradesh

^{2,3,4,5} Student, CSE, PSCMR CET, Vijayawada, Andhra Pradesh

ABSTRACT:

Crime poses a serious threat to humanity. Many different forms of crimes occur in everyday life. These types of crimes occur throughout the world. There are many different forms of crimes, including robbery, murder, assault, rape, battery, false imprisonment, kidnapping, and so on. All of these illicit behaviors are on the rise in modern society. To some extent, reducing crime aids in the resolution of this issue. It can be carried out by foreseeing a crime before it occurs. The two basic ways to prevent crime before it occurs are crime prediction and crime identification. Several thinkers have devised crime prediction models to aid in crime prevention. To calibrate the prediction models, the majority of them used Machine Learning Algorithms and Historical Crime Data. Machine learning is a branch of artificial intelligence that focuses on problem-solving. Machine Learning is a branch of artificial intelligence that works with statistical approaches that allow computers to learn from their previous experiences. Data gathering, categorization, pattern recognition, prediction, and visualization are all common steps in machine learning-based crime analysis. The main goal of this project is to develop a prediction model based on a crime dataset in India that can accurately forecast crime locations using machine learning algorithms.

KEYWORDS:

Crime Prediction, Crime Rate, Number of Crimes, Random Forest Algorithm, Data Analysis, Crime Forecasting, Machine Learning

1. INTRODUCTION

Development in human standards of living along with Crime rate is also increasing. As Technology advances, new kinds of Crimes are rising. Recently in human history, we see Cyber Crime and in recent days virtual reality Crimes. As a result, Crime may take some lives. It is better to stop something bad from happening, than it is to deal with it after it has happened. As it is difficult for security forces to predict the crime before it's going to happen. This would become a helping hand for the Security forces that

we hope for. This project helps them by predicting the crime. By Analyzing the Previous Conditions and Crime records. That result of analysis matching with recent conditions makes it predict the forthcoming crime. This helps the Security Forces by marking the anticipated location of Crime in real-time.

2. RELATED WORK

In different applications, crime prediction is varied. The subsequent square measure number of the studies:

Suhong Kim (et al.) [1] provided crime analysis victimization machine learning and also the town of Vancouver's openknowledge catalog for the last fifteen years. 2 datasets square measure utilized in this study: crime and neighborhood.The VPD (Vancouver Police Department) collects the crime knowledge, that is updated each Sunday morning. Datasetoncriminalactivity

contains committed crime, time, and place, whereas town boundaries square measure control within the neighborhooddataset. This data is updated weekly. The applied mathematics analysis provides trend patterns that may be wont toverify the best and minimum crime rates over a definite amount. The geographic study of crime episodes is stated aschoroplethmapping.Python was wont to produce the map as a result of it includes many helpful libraries for visualizinggeographic knowledge, as well as Py Sal, Geo Pandas, Folium, and comely. The KNN (K-Nearest Neighbor) and DT(Decision Tree) algorithms were used during this study. each algorithms were created victimization 2 completelydifferent approaches: one with binary variables and also the alternative with distinctive Ids. A preliminary frameworkfor knowledge analysis is obtainable during this model. though each KNN and Boosted call Tree have poor accuracy,theiraccuracy isthirty-ninep.candforty fourp.c,severally.By fine-tuning,eachthealgorithmsandalsotheknowledgeforindi vidualapplications, theaccuracyofcrimepredictionis increased.

Yingrespiratory

organMaya

Lin(etc. al) [2].plannedGrid-based

crimeprediction **victimization**

the errors using the Time Shift **style** for validation. TheCrime Prediction modeling was established on **eighty-four forms of** geographical **info** by applying the Google

placesAPI(ApplicationProgrammingInterface)bydefa ultsupportsettings.Thegrid-based **house style** is **enforced** insuch **some way** that **initially** it defines the borders of **the town** on the map, then divides **the realm** into 5-by-5 to 100-by-100 grids **to provide** a grid map of a **town** and eliminate the crime-free grids **so as to scalebackthecomplexness of knowledge.**

Structuraloptio

and **the alternative** is to **victimize** Google place API. This model usedDNN **calibration because**the main **formula tomatch** crimehotspots, **prognostication** performanceagainst **theopposite** algorithms **together with** KNN (K-Nearest Neighbor), SVM (Support Vector Machine), and RF (RandomForest). **comparedwithaccuracyDNNprovestobe themosteffective** andsecondcomesRFandSVMandKNNsquaremeasur **ethought-abouttobethesmallest amountperformancewithin thereadofaccuracy.**

ChMahendra(etc.al) [3], developed **rate** Prediction to predict the **rate supported theft victimization** theMulti **regression** (MLR).The **construct** ofMulti **regression** **isemployed** forpredictingthegraphbetween **thecategories** ofCrimes(IndependentVariable) **andalsothe** Year(DependentVariable)throughfrequency.Thisproje ct **conjointly** provestheadvantagetopredict **theshare** ofthecrime in future by **victimization** theprevious

knowledge. The dataset is extracted from official sites with machine learning algorithms **victimization** python as core. This model performs **2 ways that** for **rate** Prediction: they predict the crime and analyze the **rate** to happen in future by **creating** hotspots **supported** time, **kind** or **alternative** factors. **during this** the dataset is **chosen** from Open **knowledge supply** of a **selected town** of a **period** of **twelve** years. **supported** the dataset, **totally different** algorithms like **logistical** regression, **KNN**, **Multi regression square measure** went to notice hotspots of criminal activities. The **logistical** regression **relies** on binary variables, **Multi regression** uses **informative** variables. **when** the comparison of **those** results of **varied** algorithms, **Multi regression** is documented **because** the effective approach by **manufacturing** the token Error rate **whereas** **coaching** the model with the **assistance** of various parameters. The methodology used for this project is **assortment of knowledge**, Feature selection **victimization info** gained, Train the model, Testing **and eventually** the results of **rate is ready** through **programme victimization** Django for framework and **Aws** for storing **the info** of Crime.

Neil **Shah of Iran** (etc.al)[4], **planned** Crime **prognostication.** A Machine Learning **associated** **degree** **pc** Vision approach to crime prediction and **bar** **because** the **range** and **styles** of criminal activities increase at an **atrocious** rate, forcing agencies to develop **economical ways to require** preventive measures. **the most** objective of this project is to deliver results **expeditiously** **victimization** **KNN** (k-nearest Neighbour) and **Boosted call** Tree algorithms were **enforced to investigate** the crime **connected** dataset of **urban center** and **us** with accuracy of between **thirty ninth** and **four hundred and forty yards**

severally. Total of 5,60,000 crime datasets were analyzed between 2003 and 2018. The accuracy was low for predicting the model, however the accuracy maybe **exaggerated** by **calibration** each the algorithms. The methodology **supported** crime analysis involves **assortment of knowledge, knowledge** classification, Identification of pattern, and **visual image.** **the matter of prognostication** crime is split into 3 parts: **deciding whether or not** crime happens, **incidence** of crime and **presumably** crime. **logistical** regression, **KNN**, tree **ways square measure** went to train the dataset to assist in preventing the crime by **helping** the **law enforcement officials** to carry the burden. The model **recommended together** with machine learning and **pc** vision algorithms and techniques like **surveillance.** **there's the employment** of **ray** for **trailing** phones. **and also** the prediction technologies **referred to** as Mahalanobis and a Dynamic Time **deformation square measure** used for implementation of crime **knowledge** as these deliver **the chance** of predicting crime and apprehending **the particular perpetrator.**

XU Zhang (etc.al) [5], introduced Comparison of Machine Learning Algorithms for predicting Crime Hotspots by taking the advantage of **each** Historical Crime **knowledge** and Covariates **related to sociology** theories for future crime prediction. There **square measure** **fourteen** **differing types** of **sociology** theories **within which** it focuses on **Psychodynamic, behavioural and psychological feature.** The Predicting crime **kind** **relies** on the property **publically** places that occupies the **possession** of others **like theft, thievery** and Snatching. This **larceny** Crime is of the **city during** a Coastal Megacity in Southeast China. Six **totally different** machine learning algorithms **square measure** **thought-about** for analyzing the crime data; **they're** **KNN** (K-nearest neighbor), **RF** (Random Forest), **SVM** (Support Vector Machine), **NB** (Naïve Bayes), **CNN** and **LSTM** algorithms. The



those of the **opposite** models. It **will** finely extract the pattern and regularity from historical crime **knowledge**. The LSTM could be a reasonably deep neural network **supported** RNN (Recurrent Neural Network). Results **supported** the historical crime **knowledge** alone **recommend** that the LSTM model outperformed KNN, random forest, support vector machine, naïve **mathematician**, and convolutional neural networks. The **analysis** on crime prediction **presently** focuses on **2** major aspects, crime risk **space** prediction and crime hotspot prediction. The crime risk **space** prediction **relies** on the relevant influencing factors of criminal activities, correlation between criminal activities and physical **surroundings** These risk **square measure**s of crime are mapped to **work out** crime hotspot prediction.

3. Methodology:

In this project, we mainly use machine learning methods for crime prediction, and that we use Python and its libraries to try to do the coding. From the start we load the historical dataset to predict the forth coming crime. Then classification is performed using the random forest algorithm, generate a link. It provides web application th

which we offer the input requirements. As per the user requirements it shows the shape of bar charts and predicts the accuracy and number of crime cases specifically state.

To develop the proposed system, we need the subsequent steps:

1. Loading the Dataset

5. Result evaluation

1. Loading the Dataset:

Firstly, we load the historical crime dataset for predicting the crime. The dataset consists of crime type, number of crimes. Which is further implemented by performing machine learning methods.

2. Data Preprocessing:

In this step all the null values are removed. The most goal of information preprocessing is to boost quality of data so it can perform analyses more effectively. We will eliminate undesired irregularities, Data cleaning. It includes eliminating redundant data, missing values, data transformation and improve some attributes that are important for the applications we're performing on crime prevention preprocessing.

3. Classification:

The Classification model could be a Supervised Machine Learning technique that uses training information to measure the category of observational data. A program learns from a dataset or observation the n classifies fresh

observations into a variety of classes or groups in classification. Machine learning algorithms are frequently accustomed to recognize data and type it into categories of information. This one is understood as classifier, and it enables us to separate large amounts of information into values. We are using Random Forest algorithm for predicting the crime.

4. Prediction:

In this step, after performing the

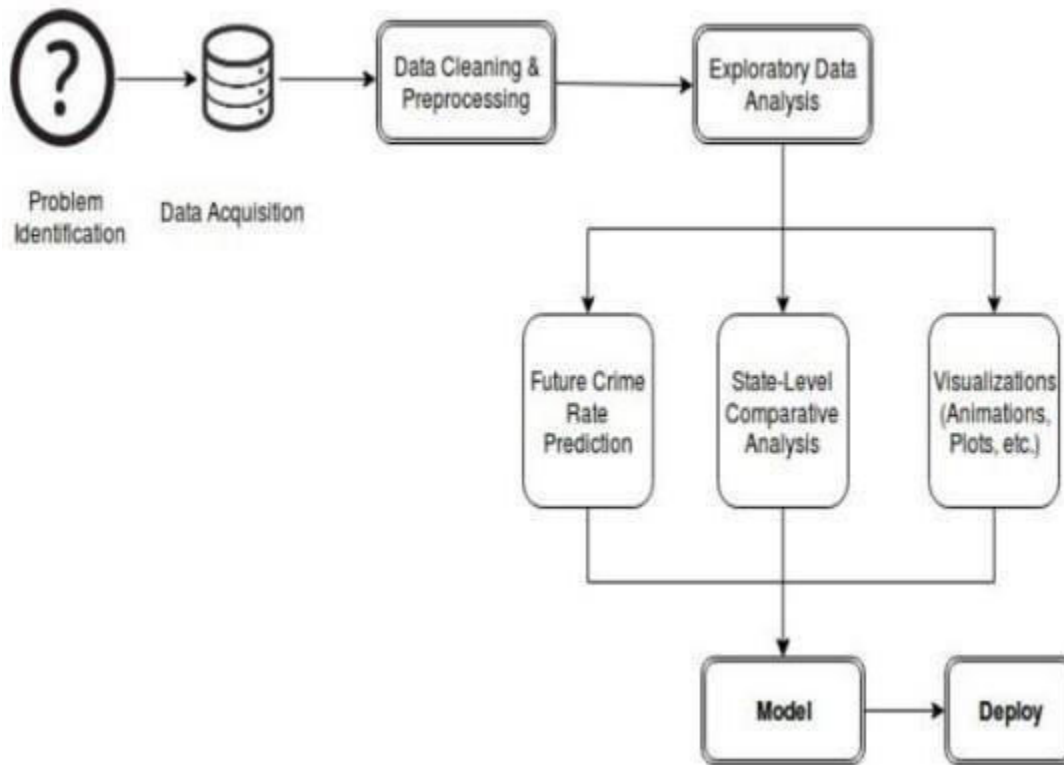


Fig3.5: System Architecture for Crime Prediction

Above shown diagram is the system architecture that clearly depicts every module of our system. The problem is invented and appropriate data is collected from India's National Crime Records Bureau. Collected data then undergoes preprocessing, cleaning, wrangling,

merging and removing null values. Analysis of this data is done

a. Result Evaluation:

The last and final step for completion of given command. In this phase has some requirements are need to be provided by the user. By consideration of the requirements of the user such as year, state, type of crime the output is generated by displays crime prediction with the particular requirements provided

according to various aspects of crime patterns such as clustering states and districts based on the crime intensity, predictions of crime ratio in future, district level comparative analysis of crime etc. All the predictions are visualized on a dynamic website.

5. Experimental Result:

The proposed system mainly focuses on the increase in crime rate day by day. These problems are assessed in a manner so that we come to a solution and have introduced this “**CRIME PREDICTION**” using Random Forest Algorithm.

When we run the source code it generates the link as highlighted in the below figure. Now this link is copied and pasted in the browser to provide a web application as a website.

User Interface:

By directing into the website through the link generated by the code the user is available to provide user requirements by clicking on the button “**Prediction**” which leads to the user interface as shown in the below figure with different crimes.

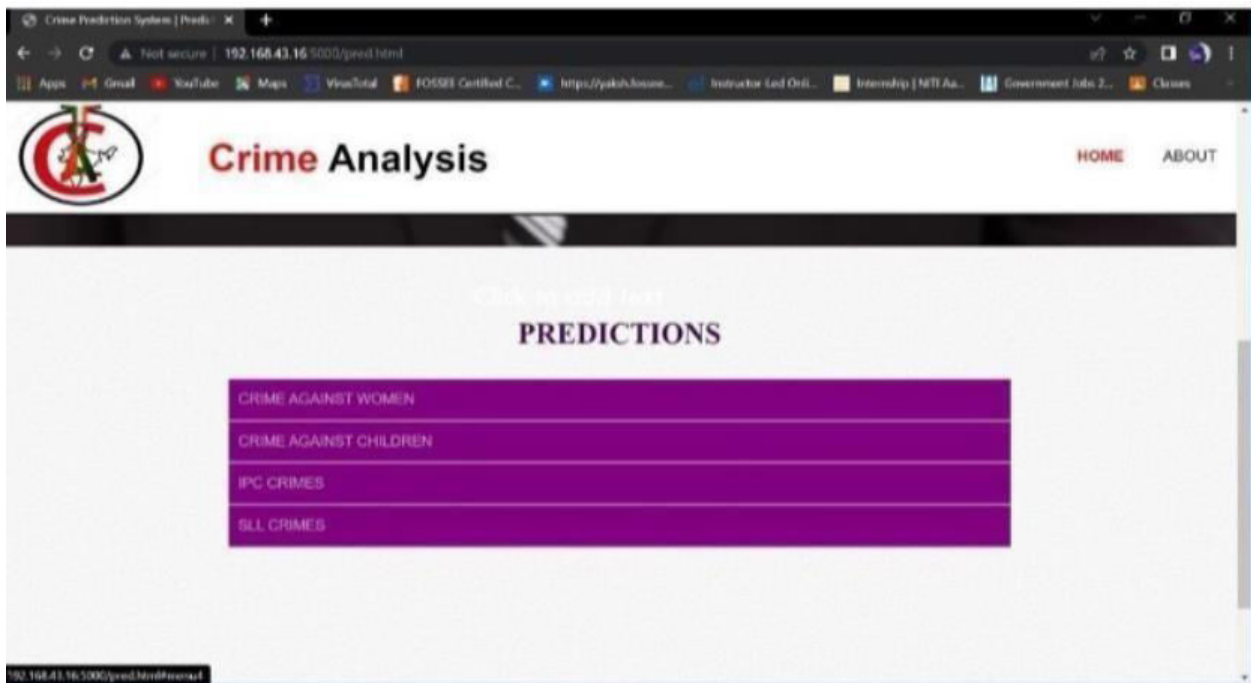


Figure 4.2: User Interface

By selecting one of the crimes we get the following requirements that the user has to choose the following requirements from the dropdown provided in user interface such as Year, Type of Crime, State.

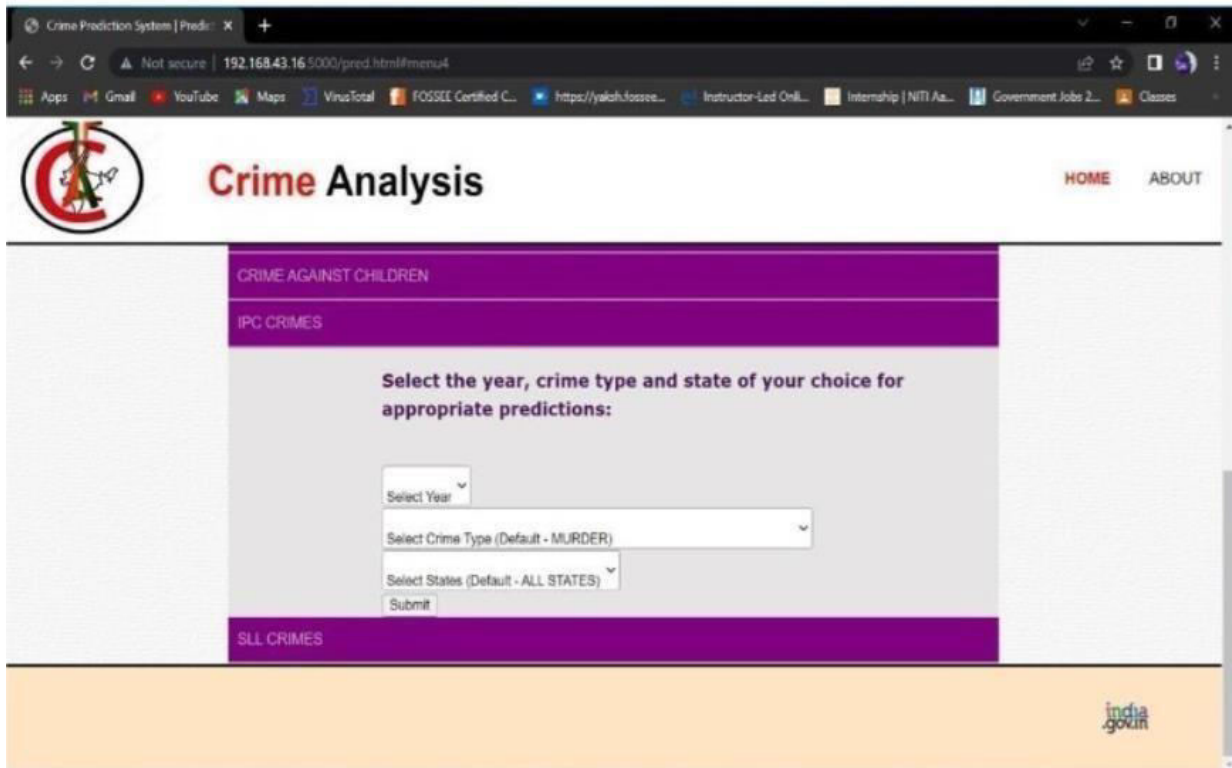


Figure 4.2.1 Input requirements

Output:

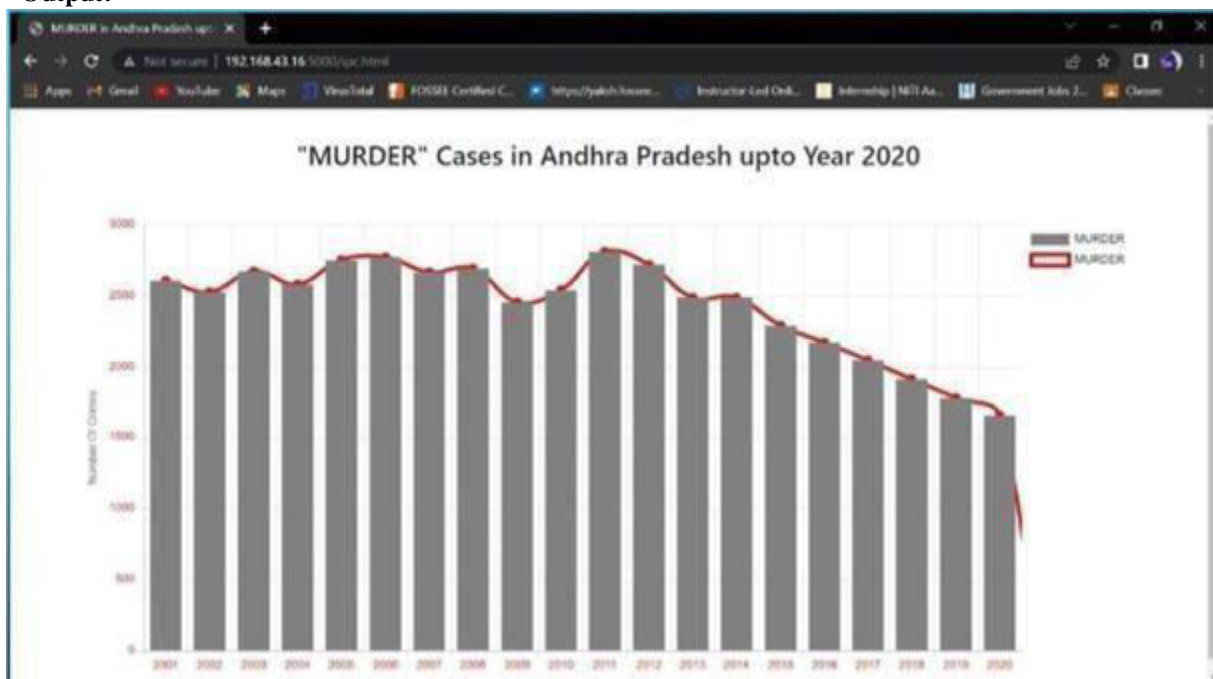


Figure 4.3 Output

As per user requirements, it provides the data in the form of bar charts by analyzing the previous data considered for predicting the crime, Number of crimes in at particular year, and state by predicting along with the code and particular requirements given by the user.

6. Conclusion

The result of this project is to present a system that can analyze, correlate and predict the crimes from huge data available. Results are going to be within the type of correlation between various crimes and site of crime i.e. state/city. Crime may also be correlated to the idea of the old group, location of crime & type of crime. Prediction of the crime is presented using various techniques and algorithms. Another point noted is that the crime rate is increasing and crime prevention has become an upheaval task. The legal force departments around the world are required to stay ahead within the eternal race between lawbreakers and law enforcers. So, we are presenting the system which maintains, predicts, and visualizes the crime records. This method is meant specifically for crime analysis to perform functions that aren't available in other existing software. Thus, it is understood that, although several solutions to unravel the problem have been proposed, it is often seen that an ideal solution to every city, state, and country continues to be exclusive.

7. References

1. S. Kim, P. Joshi, P. S. Kalsi and P. Taheri, "Crime Analysis Through Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile

Communication Conference

(IEMCON), 2018, pp. 415-420,

doi:10.1109/IEMCON.2018.861482

8.

2. Lin, Ying-Lung, Meng-Feng Yen, and Liang-Chih Yu. 2018. "Grid-Based Crime Prediction Using Geographical Features" ISPRS International Journal of Geo-Information 7, no. 8: 298. <https://doi.org/10.3390/ijgi7080298>.
3. Mallikarjuna Reddy, A., Venkata Krishna, V. and Sumalatha, L. Face recognition approaches: A survey. International Journal of Engineering and Technology (UAE), 4.6,6(7)(2018) 117-121. doi: 10.14419/ijet.v7i4.6.20446.
4. M. R. Ayaluri, K. Sudheer Reddy, S. R. Konda, and S. R. Chidirala, "Efficient steganalysis using convolutional auto encoder network to ensure original image quality," PeerJ Computer Science, vol. 7, p. e356, 2021.
5. Mallikarjuna Reddy, A., Rupa Kinnera, G., Chandrasekhara Reddy, T., Vishnu Murthy, G., et al., (2019), "Generating cancelable fingerprint template using triangular structures",

- Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, pp. 1951-1955(5), doi:
<https://doi.org/10.1166/jctn.2019.7830>.
6. Shah, Neil, Nandish Bhagat, and Manan Shah. "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention." *Visual Computing for Industry, Biomedicine, and Art* 4, no. 1 (2021): 1-14
7. B. Pruthvi Raj Goud, G. L. Anand Babu, G. Sekhar Reddy, A. Mallikarjuna Reddy "Multiple object detection interface using Hough and haar-classifier" *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-8 Issue-9, July 2019.
8. X. Zhang, L. Liu, L. Xiao and J. Ji, "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots," *in IEEE Access*, vol. 8, pp. 181302-181310, 2020, doi: 10.1109/ACCESS.2020.3028420.
9. H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," *IEEE Computer*, vol. 37, no. 4, pp. 50-56, Apr. 2004.
10. C. Ramakrishna, G. Kiran Kumar, A. Mallikarjuna Reddy, and P. Ravi, "A Survey on various IoT Attacks and its Countermeasures," *Int. J. Eng. Res. Comput. Sci. Eng.*, vol. 5, no. 4, pp. 2394-2320, 2018, [Online]. Available: <http://ijercse.com/specissue/april-2018/27.pdf>.
11. Swarajya Lakshmi V Papineni, Snigdha Yarlagadda, Harita Akkineni, A. Mallikarjuna Reddy. *Big Data Analytics Applying the Fusion Approach of Multicriteria Decision Making with Deep Learning Algorithms International Journal of Engineering Trends and Technology*, 69(1), 24-28, doi: 10.14445/22315381/IJETT-V69I1P204.
12. E. Cesario, C. Catlett, and D. Talia, "Forecasting crimes using autoregressive models," *IEEE 14th Intl. Conf. on Dependable, Auton. And Secure Comput., Auckland, New Zealand*, Aug. 2016.
13. N. Baloian, E. Bassaletti, M. Fernández, O. Figueroa, P. Fuentes. R. Manasevich, M. Orchard. S. Peñafiel, J. A. Pino, and M. Vergara, "Crime prediction using patterns and

- context," IEEE 21st Intl. Conf. on Comput. Supported Cooperative Work in Des. (CSCWD), Wellington, New Zealand, Apr.2017.
14. G. Zhou, J. Lin and W. Zheng, "A web-based geographical information system for crime mapping and decision support," IEEE Intl. Conf. on Comput. Problem-Solving (ICCP), Leshan, China, Oct.2012.
15. Yu CH, Ward MW, Morabito M, Ding W (2011) Crime forecasting using data mining techniques. In: Paper presented at the 2011 IEEE 11th international conference on data mining workshops. IEEE, Vancouver 11-11 December 2011.<https://doi.org/10.1109/ICDMW.2011.56>
16. A Mallikarjuna Reddy, VakulabharanamVenkata Krishna, LingamguntaSumalatha and AvukuObulesh, Age Classification Using Motif and Statistical Feature Derived On Gradient Facial Images, Recent Advances in Computer Science and Communications (2020) 13:965.
<https://doi.org/10.2174/2213275912666190417151247>.