



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Prediction of Crime Rate Analysis Using Supervised Classification Machine Learning Approach

Mrs.V.SHANMUGAVALLI¹, GAYATHRI.S², HARSHNI.J.R³

Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women (Autonomous), Tiruchengode, Tamilnadu, India¹

Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women (Autonomous), Tiruchengode, Tamilnadu, India^{2,3}

ABSTRACT: In current years, crime factors out that the crimes in India have considered a spike. The document provides that the instances of murder, rapes, and kidnapping have viewed a rise. Most of nations in the world have viewed a fantastic make bigger in the crime rate. There is no specific purpose for any bother for crook activities. Sometimes society, cultural factors, exclusive household systems, political influences and regulation enforcement are accountable for the crook things to do of an individual. So, the crime price is developing in India. Crime can be discovered in a variety of categories. To forestall this hassle in police sectors have to predict crime fee the use of computer studying techniques. The intention is to check out computer getting to know primarily based methods for crime price by using prediction consequences in satisfactory accuracy and discover in this work the applicability of statistics approach in the efforts of crime prediction with specific significance to the statistics set. The evaluation of dataset with the aid of supervised laptop mastering technique(SMLT) to seize various information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, lacking cost redress and analyze the statistics validation, facts cleaning/preparing and information visualization will be finished on the complete given dataset. Our evaluation offers a complete information to sensitivity evaluation of mannequin parameters with regard to overall performance in prediction of crime fee with the aid of accuracy calculation from evaluating supervise classification computing device studying algorithms. Additionally, to evaluate and talk about the overall performance of a number laptop getting to know algorithms from the given police branch dataset with assessment classification report, become aware of the confusion matrix and to categorizing records from precedence and the end result indicates that the effectiveness of the proposed desktop getting to know algorithm approach can be in contrast with high-quality accuracy with precision, Recall and F1 Score.

KEYWORDS: Dataset, Crime rate analysis, Machine Learning-Classification method, Python, Prediction of Accuracy result.

I.INTRODUCTION

OBJECTIVES

This evaluation goals to have a look at which facets are most useful in predicting the crime charge and to see the commonplace tendencies that may additionally assist us in mannequin determination and hyper parameter selection. The aim is to classify whether or not the crime fee is excessive or low with to useful for figuring out the fraud areas in actual time world. To acquire used computing device getting to know classification strategies to suit a feature that can predict the discrete type of new enter facts about crimes.

- The repository is a gaining knowledge of workout to:
- Apply the quintessential standards of laptop gaining knowledge of from an on hand dataset and Evaluate and interpret my effects and justify my interpretation based totally on located dataset.
- Create notebooks that serve as computational files and report my concept procedure and look into purposes of information for mortgage to analyses the facts set.

- Evaluate and analyses statistical and visualized results, which discover the popular patterns for all regiments.

II.LITERATURE REVIEW

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discuss published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and may be just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis.

Review of Literature Survey

Title: A Relationship between Fines and Violent Crimes

Author: Samuel Smith, Simranjyot Singh Gill and KedarGangopadhyay Year: 2018

The results of our experiments confirm our alignment with previous studies which debunked the common that unemployment and violent crimes are strongly correlated and then tested whether there was any positive linear relationship between fines and violent crimes. Due to the complex ways in which boundaries are drawn and crimes are defined, we saw no relationship at the local level. At the state level, however, the linear relationship became apparent and statistically significant. The results of our fit were confirmed by overlaps between the top fine states and top violent crime states. It also discussed equitable stop and search treatment with respect to subsets of the population. The causes of violent crime are a highly nuanced topic. It showed that a relationship between areas marked by high fines and high rates of violent crime exists, and there are potential consequences of excess fining in certain areas, it analyzes and discusses the dependence of city and county revenue generated from fines (primarily traffic violations) and their potential effects on the incidence of violent crimes on an aggregated state level.

Following the riots, several press articles pointed to Ferguson's elevated levels of municipal court fines (again, usually for traffic violations) and how they reduced the local population's faith in the police and overall city government. It tested whether the practice of collecting significant municipal revenue from low-level offenses had an impact on violent crimes not only in Missouri, but in other states as well.

Title: Forecasting of Annual Crime Rate in India:A case Study

Author: Manish Kumar, Athulya S, Mary Minu MB

Year: 2018

India's population is estimated to be around one billion. The high population density, combined with other factors such as lack of jobs, poverty, and illiteracy will result in a higher violence rate. The crime and violence rate vary from state to state. States like Uttar Pradesh, Bihar etc records high crime rates according to 2017 statistics. Like other counties increase in crime rate is a major concern in India also. From the reports of National Crime Record Bureau (NCRB), states that most of crime incidents recorded are in urban area. In India, crime rate (case reported per lakh population) has increased from 166.7 to 215.5 in years from 1953 to 2013. By analyzing the data, crime rates got highly fluctuated in the years 1970- 2005. The statistics indicate that crime rate in India is steadily increasing for the past 8-9 years. Source of data is from the National Crime Record Bureau of India. As apart of modeling, data is divided into training data for the years 1953 to 2008 and test data for the years 2009 to 2013. By examining the model, it's clear that the forecast values are within the 95% confidence interval of the test data and accuracy measurements are also significant. Hence the time series model suitable for crime forecasting. This paper concluded that time series model can be applied for crime forecasting. The result obtained from both the models conclude that they are significant for forecasting all test data which are lying between a 95% confidence interval and accuracy measurements for training data shows that they are numerically significant. In future, we are trying to analyze crime against women, children so that we can predict how much police strength is convenient to decrease the crime rate.

Title: Intelligent Crime Anomaly Detection in Smart Cities using Deep Learning

Author: SharmilaChackravarthy, Steven Schmitt, Li Yang Year: 2018

In the past a strong reliance has been put on standard video surveillance in order to achieve this goal. This often creates a backlog of video data that must be monitored by a supervising official. For large urban areas, this creates an increasingly large workload for supervising officials which leads to an increase in error rate. Solutions have been implemented to help reduce the workload. Currently, auto regressive models have been used to better forecast criminal acts, but also have a list of shortcomings. It proposed a solution of using neural networks in combination with a Hybrid Deep Learning algorithm to analyze video stream data. Our system will be able to quickly identify and assess criminal activity which will in turn reduce workloads on the supervising officials. When implemented across smart city infrastructure it will allow for a efficient and adaptable crime detection system. Our system can be applied to various video surveillance systems to act as an alert system, which would reduce the overall workload on security officials. Automation and smart, adaptive security systems are a way to increase detection rates in hopes of curbing crime rates in large hard to monitor areas.

EXISTING SYSTEM

Latest technical developments in sophisticated tools of data analytics and visualization are helping the society in different ways to analyze the data of social relevance. One of such socially relevant activities is crime details of different demographic places. The analysis of the crime data will help decision making agencies to take precautionary steps to control the crime rate over demographic places. Advancements in the field of information technology, publicly available information and services, somehow help criminals to achieve their misdeeds and involve them in much serious crimes than earlier. As a result crime rate is increasing with a very high rate in developed and under-developed nations. Based on the previous year crime details in Indian states, It present statistical models through Weighted Moving Average, Functional Coefficient Regression and Arithmetic-Geometric Progression based prediction of the crime in coming years. Difference between actual records and our predicted both years gives the accuracy of the proposed approaches between the range 85% and 90%. In future, this work can be modified by using Machine Learning (ML) models for forecasting crime, as the data points will sufficiently increase to apply ML models. This can also increase the accuracy of the predictions. Further, statistical modeling's methods can also be clubbed with ML models and then calculate weighted accuracy for a district, this can make the solution more robust.

DRAWBACKS

- The accuracy results are not more than 90%. Its cannot work on top features and find- out the Recall, Precision, Confusion matrix and compare it with our old result.
- It's cannot work on using the popular machine learning algorithm to find out the features importance..

III.PROPOSED SYSTEM

Machine learning is a computer system's method of learning by way of examples. There are many machine learning algorithms available to users that can be implemented on datasets. However, there are two major types of learning algorithms: supervised learning and unsupervised learning algorithms. Supervised learning algorithms work by inferring information or "the right answer" from labeled training data. The algorithms are given a particular attribute or set of attributes to predict. Data preprocessing process includes methods to remove any null values or infinite values which may affect the accuracy of the system. The main steps include Formatting, cleaning and sampling. Cleaning process is used for removal or fixing of some missing data there may be data that are incomplete.

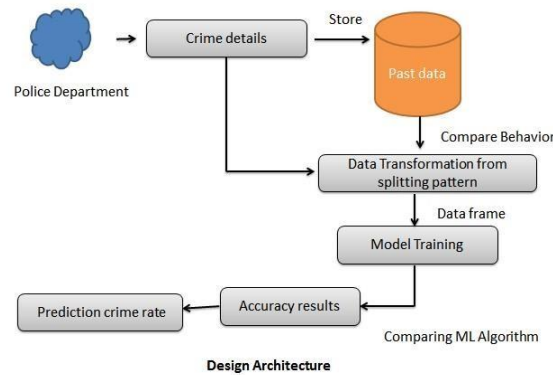
Crimes Prediction ways:

- To utilize the resources identify the hotspots of crimes and allocate vigilante resources such as policeman, police cars, weapons etc. reschedule patrols according to the vulnerability of a place.
- Through that avoid crimes Ensure better civilization through avoiding happening crimes such as murder, rapes, thefts, drug, smugglings etc.

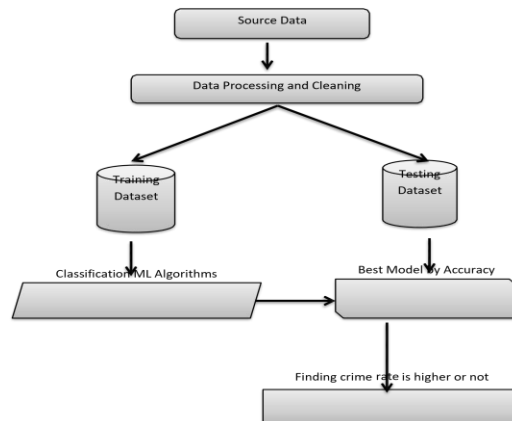
IV.SYSTEM DESIGN

Design is significant engineering illustration of some thing that is to be built. Software format is a technique layout is the best way to precisely translate necessities in to a completed software program product. Design creates a illustration or model, offers element about software program facts structure, architecture, interfaces and elements that are essential to enforce a system.

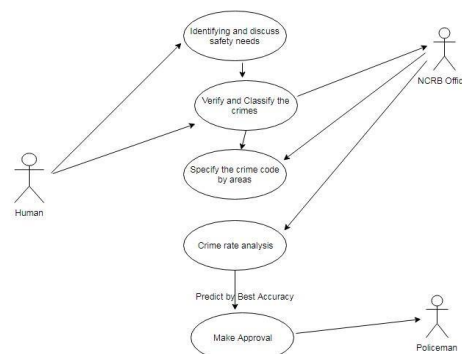
SYSTEM ARCHITECTURE



WORK FLOW DIAGRAM



USE CASE DIAGRAM



Use case diagrams are considered for high level requirement analysis of a system. So when the requirements of a system are analyzed the functionalities are captured in use cases. So, it can say that uses cases are nothing but the system functionalities written in an organized manner. Now the second things which are relevant to the use cases are the actors(human/police).

V.ALGORITHM AND TECHNIQUES

K-NEAREST NEIGHBOR (KNN)

K-Nearest Neighbor is a supervised machine learning algorithm which stores all instances correspond to training data points in n-dimensional space. When an unknown discrete data is received, it analyzes the closest k number of instances saved (nearest neighbors) and returns the most common class as the prediction and for real-valued data it returns the mean of k nearest neighbors. In the distance-weighted nearest neighbor algorithm, it weights the contribution of each of the k neighbors according to their distance using the following query giving greater weight to the closest neighbors.

Usually KNN is robust to noisy data since it is averaging the k-nearest neighbors. The k- nearest-neighbors algorithm is a classification algorithm, and it is supervised: it takes a bunch of labeled points and uses them to learn how to label other points. To label a new point, it looks at the labeled points closest to that new point (those are its nearest neighbors), and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point (the “k” is the number of neighbors it checks).Makes predictions about the validation set using the entire training set. KNN makes a prediction about a new instance by searching through the entire set to find the k “closest” instances. “Closeness” is determined using a proximity measurement (Euclidean) across all features.

RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

The following are the basic steps involved in performing the random forest algorithm:

- Pick N random records from the dataset.
- Build a decision tree based on these N records.
- Choose the number of trees you want in your algorithm and repeat steps 1 and 2.

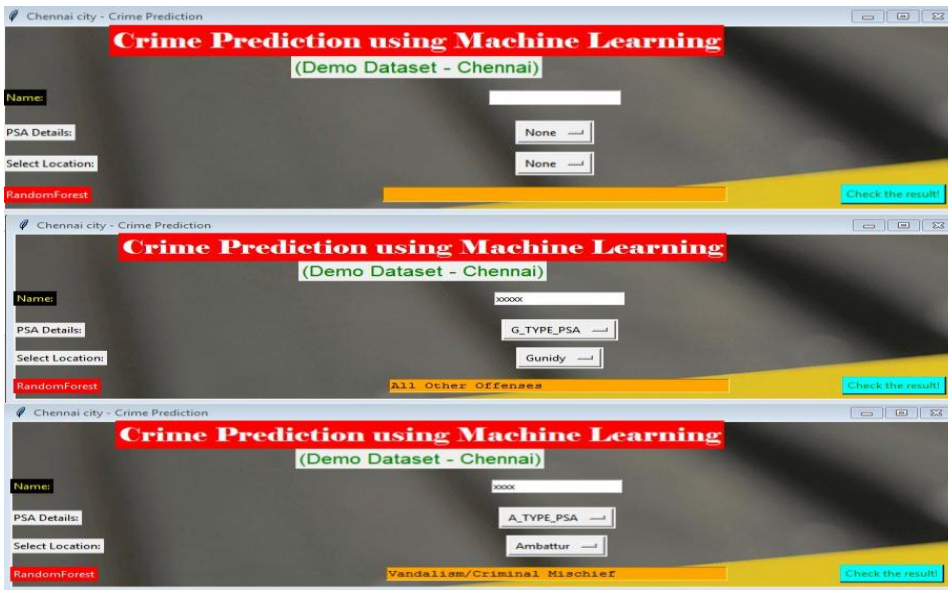
In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output). The final value can be calculated by taking the average of all the values predicted by all the trees in forest. Or, in case of a classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.

VI.RESULT

Comparison result for best Accuracy

Algorithm	Precision	Recall	F1-Score	Accuracy(100%)
DT	1	1	1	100
SVC	0.76	0.60	0.45	59.97
LR	0.81	0.80	0.79	79.71
KNN	0.87	0.87	0.86	86.73
RF	1	1	1	100

This table illustrates the comparison results for the best accuracy with respect to the different algorithms. The cent percent efficiency results are achieved by the DT and RF algorithms. While the above average results is showed by LR and KNN algorithms with 79.71 and 86.73 respectively. The lowest value is scored by SVC with 59.97 accuracy. The numerical values are accurate it is not round off to the nearest integer value.



VII.CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score is will be find out. This brings some of the following insights about crime rate. It has become easy to find out relation and patterns among various data's. It mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred in real time world. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation. Data visualization generated many graphs and found interesting statistics that helped in understanding Indian crimes datasets that can help in capturing the factors that can help in keeping society safe.

VIII.FUTURE ENHANCEMENT AND APPLICATION

FUTUR WORK

Police department wants to automate the detecting the crime from eligibility process (real time) based on the crime rate of areas.

To automate this process by show the prediction result in web application or desktop application.

To optimize the work to implement in Artificial Intelligence environment.

APPLICATION

To prove, how effective and accurate machine learning algorithms can be at predicting violent crimes, there are other applications in the territory of law enforcement such as determining criminal, creating criminal profiles, and learning crime trends. Utilizing these applications can be a long and tedious process for law enforcement officials who have to sift through large volumes of data. However, the precision in which one could infer and create new knowledge on how to slow down crime is well worth the safety and security of people.

REFERENCES

- [1] A Relationship between Fines and Violent Crimes Samuel Smith, Simranjyot Singh Gill and KedarGangopadhyay 2018
- [2] Forecasting of Annual Crime Rate in India: A case Study Manish Kumar, Athulya S, Mary Minu MB 2018
- [3] Intelligent Crime Anomaly Detection in Smart Cities using Deep Learning SharmilaChackravathy, Steven Schmitt, Li Yang 2018
- [4] J. Kerr, "Vancouver police go high tech to predict and prevent crimebefore it happens," Vancouver Courier, July 23, 2017.[Online]Available:<https://www.vancourier.com/news/vancouver-police-go-high- tech- topredict-and-prevent-crime-before-it-happens-1.21295288>. [Accessed:09- Aug- 2018]
- [5] J. Han, Data mining: concepts and techniques, Morgan Kaufmann, 2012.
- [6] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N.Khanahmadliravi, "An experimental study of classification algorithms forcrime prediction," Indian J. of Sci. and Technol., vol. 6, no. 3, pp. 4219-4225, Mar. 2013.
- [7] M. Al Boni and M. S. Gerber, "Area-specific crime prediction models,"15th IEEE Intl. Conf. on Mach. Learn. and Appl., Anaheim, CA, USA,Dec. 2016.
- [8] Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, "Mixed spatial-temporal characteristics based crime hot spots prediction," IEEE 20th Intl. Conf. on Comput. Supported Cooperative Work in Des. (CSCWD), Nanchang, China, May 2016.
- [9] N. Mahmud, K. Ibn Zinnah, Y. Ar Rahman, and N. Ahmed,"CRIMECAST: a crime prediction and strategy direction service," IEEE 19th Intl. Conf. on Comput. and Inform. Technol., Dhaka, Bangladesh,Dec. 2016.
- [10] Zakir J, Seymour T, et al., BIG DATA ANALYTICS. Issues inInformation Systems, vol. 16, no. 2, pp. 81-90, 2015.
- [11] Wang Y, Kung L A, et al. An integrated big data analytics-enabled transformation model: Application to health care. Information &Management, vol. 55, no. 1, pp. 64-79, Jan. 2018.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details