IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Empirical Analysis For Crime Prediction And Forecasting Using Machine Learning And Deep Learning Techniques

[1] Rachakonda Geethika, [2]Goda Bhargavi,[3]Meenagari Samatha Sri,[4]Preethi Jeevan

[1,2,3] Graduate Student-CSE, Sreenidhi Institute of Science and Technology, Ghatkesar, Hyderabad
[4] Asst.Professor, Department of CSE, Sreenidhi Institute of Science and Technology, Ghatkesar, Hyderabad

***Abstract*—Crime forecasting and prediction using machine learning has** become an important field of study in recent years. By analyzing historical crime data, machine learning algorithms can identify patterns and trends that help predict the likelihood of future crimes occurring in specific areas or times.

Every nation needs to prioritize its efforts in the areas of criminal investigation and prevention. The number of criminal cases in India and other countries is increasing at a rapid place, which has resulted in an increasing number of active cases. Keeping up with the growing quantity of criminal cases is becoming more and more challenging. Preventing criminal behavior in a given area requires an understanding of its trends. Having a comprehensive understanding of the patterns of criminal conduct prevalent in a certain area might enhance the efficacy of crime-solving organizations. Despite the significant and ongoing issue of crime against women, women's safety and protection are primary concerns.This work uses ensemble methods to anticipate crime patterns that could emerge in a given area.

***Index Terms*—crime forecasting, Preventive measures , Criminal behaviour , Interventions.**

INTRODUCTION

Project Overview

Crime prevention is an essential endeavor since it is one of our society's most widespread issues. significant and concerning issues. "Crime analysis and prediction" refers to the systematic examination and interpretation of numerous connections, patterns, and trends in crime and disorder. Crime is a common occurrence everywhere in the globe, particularly in cities. Consequently, fundamental human rights are violated and the social order of today's society is upended. Law enforcement officers sometimes find it difficult to manually identify the locations with a high crime rate over a certain time period. because criminal urges are dynamic in nature and because the available crime data is being used incorrectly. A growing population causes the patterns of criminal behavior to change and become more intricate.

Due to job insecurity and financial losses brought on by the COVID-19 shutdown, there has been psychological suffering. The number of crimes has increased, including shoplifting and break-ins. Although these offenders are listed in databases kept by governments, cases must be decided based on particular evidence, which requires independent investigation and assessment. The large amount of crime data accessible in India makes it difficult to evaluate and make choices on these issues.

After recognizing that crime decision-making is a serious problem, the goal of this research is to develop a remedy. The car begins to move on its own. As it takes various judgments to accomplish its objective, an autonomous driving car goes through the identification, judgment, and control stages multiple times. recognizing the pattern of criminal activity. Place should be the organizing principle for crime. analysis of criminal activity.

Criminal activity poses a significant threat to humanity, with a wide range of crimes occurring at predictable intervals. Authorities struggle to predict and identify perpetrators due to the abundance of data on crime. To simplify the criminal case settlement process, reinforcement learning algorithms and data science are explored. The study aims to use a dataset to develop predictions about criminal behavior, using a machine learning algorithm in Python. The model will be trained using a classification system similar to the K-Nearest Neighbor technique. Data visualization will be used to examine potential criminal activity, potentially reducing crime rates in Chicago.

Objective

The process of studying spatial distribution, analyzing temporal dynamics, detecting patterns and trends, evaluating the predictive performance of machine learning models, identifying significant features, adjusting model parameters, and assessing predictive resilience is all outlined in the text. The significance of temporal dynamics, cross-validation, and geographical analysis in guaranteeing the precision and dependability of crime prediction models are also covered. The necessity of ongoing development for crime prediction models is emphasized.

The study intends to improve public safety by creating efficient crime prediction models for proactive law enforcement tactics, incorporate outside variables like socioeconomic indicators and meteorological conditions into crime prediction accuracy, and offer law enforcement and policymakers useful insights.

LITERATURE SURVEY

Existing System

The study of data that is pertinent to social concerns is being made easier for society by the development of sophisticated data analytics and visualization tools in modern technology. Among these socially significant activities are crime statistics for different demographic groups. Decision-making organizations will benefit from the analysis of crime data as they take proactive steps to lower crime in populated regions.

Publicly accessible information and services, along with advancements in information technology, somehow assist criminals in carrying out their misdeeds and entangling them in more serious crimes than before. As a result, both industrialised and developing countries are experiencing extraordinarily rapid increases in crime. It presents statistical models using weighted moving average and functional statistics based on the previous year's crime data for Indian states.

Related Work

Numerous studies have been conducted to address the issue of lowering crime, and numerous crime-prediction algorithms have been put forth.The kind of data used and the qualities chosen for prediction both affect how accurate the forecast is.With an accuracy of almost 70%, mobile network activity was utilized in[5] to gather human behavioral data that was used to forecast the crime hotspot in London. This data was used to determine whether or not a particular location of the city will see a spike in crime. IN[6], The Naive Bayes algorithm and decision trees were used to predict and classify crime using data gathered from multiple websites and newsletters. It was discovered that the former performed better. In [7].

After a thorough analysis of several crime prediction techniques, including Support Vector Machines (SVM) and Artificial Neural Networks (ANN), it was determined that no single technique could address the issues presented by distinct crime datasets. IN[8], different methods for supervised and unsupervised learning[9], research was conducted on the crime records to examine the relationships between crime and crime patterns in an effort to increase the predicted accuracy of crime through knowledge discovery. As stated in [10].

Several approaches to prediction, including data mining, deep learning, sentiment analysis, and crime casting, were examined. It was shown that each method had advantages and disadvantages. Every technique produces superior results in a certain situation. Classification methods were employed to predict crime, and clustering approaches were used to detect it [11].

.Accuracy is the criterion used to assess the performance of the K-Means clustering. When evaluating the effectiveness of various clustering algorithms, DBSCAN produced the most accurate results, and the KNN classification technique is employed to predict crime.

Therefore, this technology aids in the accurate and enhanced analysis of crimes by law enforcement authorities.Within [12] WEKA data mining software was used to compare two
categorization algorithms: decision trees and Naïve Bayes. USCensus 1990 provided the datasets used in this investigation. After accounting for a number of variables, including the driver, the vehicle, the road conditions, etc., the pattern of road accidents in Ethiopia was examined in [13].On a dataset with about 18,000 datapoints, many classification techniques were applied, including K-Nearest Neighbor, Decision Tree, and Naive Bayes. For each of the three approaches, the prediction accuracy ranged from 79% to 81%.

Problem Statement

Determining how to extrapolate past criminal activity data into the likelihood of future occurrences occurring at particular points in place and time presents a challenge for crime analysts. The goal of descriptive analysis is to find spatial and temporal patterns in crime data. Usually, predictive analytics approaches are employed to predict the kind of crime that might occur at any location at any given moment. The An analyst would prefer a graphic map that showed the level of likely criminal activity at each location within their boundaries of jurisdiction.

This is accomplished by combining population and crime data and feeding it into machine learning algorithms. To lessen crime and its effects on the public, the prescriptive analyzer provides process reengineering techniques that efficiently allocate police resources. Opinions and points of view are obviously crucial to preventing crime since they allow the police to focus resources on high-risk regions.

METHODOLOGY

As the field of Earth and atmospheric sciences continues to evolve with the rapid growth of artificial intelligence, data-driven models powered by deep learning are becoming more prevalent to effectively analyze and predict weather patterns .There are two main classes to it. Regression is one, while pattern categorization is the other.

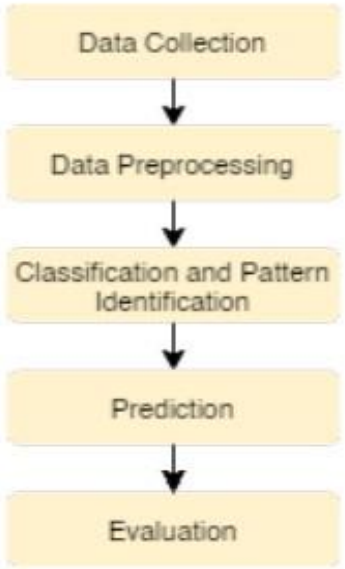
Analysis of the existing relationships serves as the foundation for regression models.in order to forecast the continuous variables, one must consider the relationship between trends and variables.

As the prediction's result, classification's task is to allocate a certain class label to a given data value. There are two methods for classifying patterns: supervised learning and unsupervised learning.

In supervised learning, it is already known which class labels should be applied for creating classification models.

Unsupervised learning does not know these class labels. The topic of this paper is supervised learning.

Gathering and pre-processing data Information is obtained from numerous sources during the process of data collection, which is subsequently employed in the development of machine learning models.



A Data collection

Here, we have gathered a dataset from the local police station that includes details on the type of offense, section, witness name, accused person name, age, gender, report date, report time, occurrence date, and occurrence time, among other information. Decision tree techniques are used to the training set of the data model, which was generated using Random Forest, and test set prediction is carried out depending on the correctness of the test results.

B Model Selection

1) Support Vector Machine: This algorithm works well for classification, temporal prediction series, and regression tasks. The performance of support vector machines can be evaluated in comparison to recurrent neural networks.

As a result, SVM has been used to forecast crime hotspots [16] as well as illnesses like diabetes and pre-diabetes.because it is capable of creating coherent prototypes for nonlinear relations.

It works effectively for time series prediction. It must use K-clustering to choose a subset of data for a given degree of crime

2) K-nearest Neighbor: The correlation between the train and test sets is ascertained using this technique.

If the test set is close to the train set, the train set's class label is applied to it. The primary limitation is visible when the training set contains fewer data points.

Numerous techniques, including the K-NN algorithm, have been used to enhance it. This approach is classified as supervised learning. Among its applications are pattern recognition, intrusion detection, and data mining.

In this instance, class membership is the result. An object is categorized by a collective vote among neighbors, where the most familiar of its k-nearest neighbors receives the most votes.

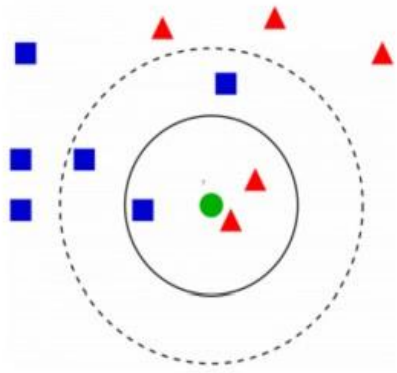


Fig: Example of KNN

3) Decision Tree: One of the most often used and effective tools for prediction and categorization is the decision tree. It is organised like a tree, with every intermediary nodesymbolises a test on a peculiarity, with each branch denoting the test's outcome and each leaf node holding the class label. Most of the time, the desired variable is categorical. Decision trees can be used to classify records (by placing records in the most comparable class) or to determine the likelihood that a certain record falls into each category.

C .Training and testing

In this stage, following the verification of the selected algorithm's underlying assumptions. Based on the provided training Sample, the model is trained. The model's accuracy

and error rates are used to evaluate its performance after training. Finally, the trained model is tested using some previously unseen data, and its performance is evaluated using a range of problem-specific performance metrics

4) Data processing.: The stages in this technique are designed to get rid of any null or infinite values that can affect the correctness of the system. The main steps are cleaning, sampling, and formatting. It could be necessary to use the cleaning process to amend or eliminate any incomplete data. It is possible to reduce the algorithm's execution time by using sufficient data for sampling. In Python, the preprocessing is done. There may be missing values in the collected data, which could lead to conflicting results. Preprocessing the data is necessary to improve the algorithm's performance and yield better outcomes.

Variable conversion must be done after the outliers are removed. Taking into account the correlation between the attributes, it was found that the most significant of all, credit history, is followed by property area, education, loan amount, and credit history. Contrary to popular assumption, a number of variables, including the income of the applicant and co-applicant, do not matter when examined independently. To make analysis easier, the data is reduced to a predetermined minimum number of records. The online dataset is updated and maintained by the Indian police agency.

Creation of an Anticipatory Model Machine learning requires the collecting of data, and a lot of historical data is available. Enough raw and historical data are available for data collecting. You cannot utilize raw data directly without first preparing it. Next, the type of algorithm used with the model undergoes preprocessing. After training and testing, this model produces precise predictions with little mistakes. In order to improve accuracy, a tailored model is constantly modified. Finally, do an analysis of the provided dataset and describe how to automatically modify it in light of the inconsistent historical data on bank accountants after the model has been constructed, tested, and put into use.

.

PROPOSED SYSTEM AND ARCHITECHTURE

Proposed System

The solution is given in the form of a statistical and machine learning model that makes use of clustering, regression, and classification algorithms. These algorithms include K-NN, Bayes Nave, and regression algorithms, and they can be used to characterize the functional relationships between victim, economic, social, and demographic variables. by looking for trends in sets of criminal data. When combined with the previously discussed methods, temporal series techniques have shown to be advantageous in enabling the model to accurately predict criminal incidents based on features that change over time and exhibit temporal growth

.

Predictive modeling is a mathematical process that involves analyzing patterns in data to make predictions about future events or outcomes. Predictive modeling is a mathematical technique that involves analyzing patterns in existing data to predict future events or outcomes .Predictive modeling is a statistical approach that uses patterns in past data to predict future behavior or outcomes.

Predictive modeling is a statistical technique that involves analyzing patterns in existing data to predict future events or outcomes. Predictive modeling is a mathematical process that involves analyzing patterns in data to predict future events or outcomes.

Predictive modeling, which involves analyzing patterns in existing data to predict future events or outcomes, is a valuable tool in today's rapidly changing world. Predictive modeling is a mathematical process that involves analyzing patterns in data to predict future events or outcomes.

In order to characterize the functional relationships among demographic, economic, social, victim, and geographic variables, a statistical and machine learning model utilizing classification, clustering, and regression algorithms—K-NN, Bayes Nave, and regression algorithms—is supplied as the solution. using criminal data sets' trends for analysis. methods for temporal series have demonstrated their usefulness when combined with the previously discussed techniques to enable the model to high degree of accuracy by predicting criminal episodes based on variables that change and evolve over time.

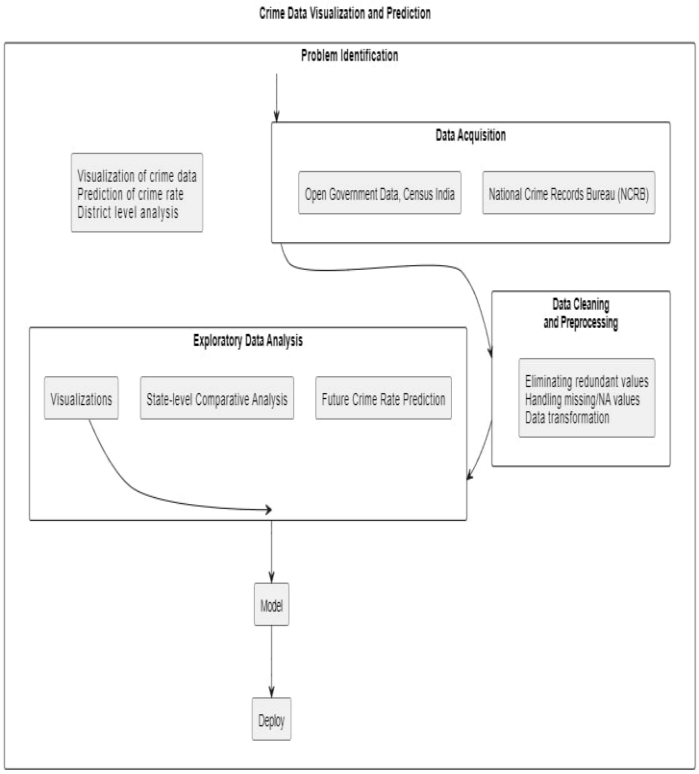


Fig: PROPOSED SYSTEM AND ARCHITECHTURE

RESULTS AND DISCUSSIONS

Data Set

Our dataset contains information from 2001 to 2020. The first set of primary data for crimes is gathered through fieldwork-based primary data gathering. This set consists of more than 500 more than 10 rows of data. Name, Years, Months, Crime Types, Crime Areas, Victims Genders, Victim Ages, Victim Areas, and Year are the main components. Three groups comprise our dataset: state-level IPC cases, cases involving women, and cases involving children. Using this dataset, we projected state-level crime.

Algorithms Used

Random Forest algorithm:

The Random Forest algorithm is a supervised learning method that builds a forest with several trees . Multiple decision trees can be built to use it in training for tasks like regression, classification, and other tasks. Random Forest can be used to rank the variables according to their priority.

Brief description Random forest algorithm fills in like an enormous assortment of connected decision trees. We use lot of selected trees, but this irregular random forest algorithm makes a bunch of selected trees and uses them to classify . This is why the random technique-dependent procedure is implemented.

$$M = \begin{matrix} f_{A1} & f_{B2} & f_{C3} \\ f_{A2} & f_{B2} & f_{C3} \\ f_{An} & f_{Bn} & f_{n} \end{matrix}$$

displays the M matrix In this instance, let's say that matrix S represents a matrix of training samples that we may provide to the algorithm in order to build a classification model. In this situation, the characteristics are represented by the values f A1, f B1, and f C1. If we move on to all the samples up to N with A1 being the feature A of the first sample, Accordingly, the f B N is the feature B of the Nth sample, and we also have the C1 and CN in these last columns, indicating that we have a training class and a large number of features. Thus, the goal is to construct an arbitrary random forest to serve as an illustration of how the method operates. Assemble a randomized subset by:

The Random Forest creation pseudo-code is displayed by the creator immediately away: Pick "M" highlights at random from all-out "N" highlights where M ≥ k Determine the hub "b" among the "M" highlights by using the best part point.

Using the optimal split, divide the hub into smaller girl hubs. Continue the a through c endeavors until the "l" number of hubs is attained. Build woods by repeating steps one through n a number of times to create a number of trees.

The random forest model is trained annually, and as the model's accuracy rises—from 0% in 2001 to 79.37 percent in 2015—the model is trained. In a similar vein, there is a rise in the rate of genuine positive values and a drop in the rate of FP.

Time Series Analysis Through LSTM:

One attractive RNN architecture variation that can be used to model sequential data is the LSTM architecture. Because of its structure, LSTM provides an efficient countermeasure to RNNs' vanishing gradient issue. It makes use of memory that can depict the long-term dependencies found in sequential data. By more naturally capturing the structure of sequential data, LSTM offers better learning for time series. For complex temporal applications, it even performs hierarchical processing. For the past 20 years, time series classification tasks have been regarded as difficult in the field of data mining because they differ from typical classification and regression predictive modeling problems . Time-series data is needed for classification in practically all real-world applications, from cybersecurity to electronic health records .

The data were first preprocessed to remove noise and then converted into stationary data before LSTM implementation. In order to facilitate handling and improve categorization, time series data must be converted from their typical non-stationary form into stationary form. As a result, the Dickey-Fuller test is used to further assess the suitable error scores and to check for stationary data in a standard manner . After processing the data and training the LSTM model on a collection of time-series data, the results offer detailed recommendations. Different forms of mistakes, such as the scale-dependent error and percentage error, are typically measured for time-series data. Here, the batch size and number of epochs were combined with two well-known scale-dependent error measures: the RMSE and the MAE.

Data Preproccessing

One crucial stage in the analysis of data is data pre-processing. Preparing the data for use in a machine learning algorithm involves transforming it into a format that is compatible with the algorithm and fixing or replacing any missing values in the data set. The mean of the values in the associated column is used to replace NA values (missing values) in a column according to our ICDA method. For regression algorithms to produce the appropriate regression analysis, label encoding is the incorrect method. Regression algorithms require numerical data.

For every state, district, and territory, dummy variables are created using the Dummy data frame method from the R dummies library. It results in 808-1 district dummy columns and 35-1 state dummy columns. To escape the dummy variable trap for a feature with n class levels and n number of dummy variables, one dummy variable is not created for each categorical variable. When a record is set to 0, it indicates that it does not belong to the associated class level, and when it is set to 1, it indicates that it does. This is the meaning of a dummy variable.

From 2001 to 2012, district-level registered cognizable IPC crimes are included in the raw data set, which is sourced from the NCRB's official website. Next, using R and MySQL workbench, data transformation and cleaning are carried out as needed. R is used to replace the missing values in the data set with the mean value of the corresponding column. After that, MySQL Workbench is used to import the data set into a database. Crime data was

generated both state- and year-wise with the aid of structured queries. The categorical variables in these data sets are then subjected to dummy coding before being used to construct regression models. Following the production of appropriate derived data sets, desirable regression models are constructed using these data sets. derived on a state, district, and annual basis

	A	B	C	D	E	F	G
1	STATE/UT	DISTRICT	YEAR	MURDER	ATTEMPT TO MURDER	CULPABLE HOMIC	RAPE
2	ANDHRA PR	ADILABAD	2001	101	60	17	50
3	ANDHRA PR	ANANTAPU	2001	151	125	1	23
4	ANDHRA PR	CHITTOOR	2001	101	57	2	27
5	ANDHRA PR	CUDDAPAH	2001	80	53	1	20
6	ANDHRA PR	EAST GODA	2001	82	67	1	23
7	ANDHRA PR	GUNTAKAL	2001	3	1	0	0
8	ANDHRA PR	GUNTUR	2001	182	88	2	54
9	ANDHRA PR	HYDERABA	2001	111	113	7	37
10	ANDHRA PR	KARIMNAG	2001	162	85	6	56
11	ANDHRA PR	KHAMMAM	2001	93	60	1	47
12	ANDHRA PR	KRISHNA	2001	65	51	0	37
13	ANDHRA PR	KURNOOL	2001	133	72	4	29
14	ANDHRA PR	MAHABOOI	2001	157	67	26	59

Fig: state wise data

selecting one of the 29 dependent variables (one variable representing all cognizable IPC offenses, and the other 28 variables representing various categories of cognizable IPC crimes). Subsequently, state-, district-, and year-specific regression models are constructed by modifying R scripts in accordance with the selected dependent variable. Every R script produces a regression model that has been trained, tested, and assessed. This study compares the accuracy and goodness of fit of regression models that predict district-wise total cognizable IPC offenses and thefts using MAPE, R squared, and adjusted R squared measures in the results section. With the use of comparison , the comparison is displayed. Based on the, the best model is selected.

XGBoost

The boosting method XGBoost employs bagging, which involves training several decision trees and combining the outcomes. It not only offers XGBoost an advantage in scenarios where there are numerous attributes to take into account, but it also enables it to learn faster than other algorithms.

Model performance and execution speed are the two main uses of XGBoost. Operating speed is critical since handling big datasets requires it. You can deal with datasets that are larger than what would be feasible to utilize with other algorithms when you use XGBoost because there are no limitations on the size of your dataset.Because it enables you to build models that can outperform other models, model performance is also crucial. A variety of methods, including random forest (RF), gradient boosting machines (GBM), and gradient boosting decision trees (GBDT), have been compared to XGBoost. These comparisons demonstrate how XGBoost performs better than these other algorithms in terms of model performance and execution speed.
A machine learning approach called gradient boosting builds a sequence of models and then combines them to produce an overall model that is more accurate than any of the individual models.

It supports challenges involving predictive modeling for both classification and regression. Gradient boosting, a gradient descent algorithm, is used to add ew models to an existing one. The XGBoost package, sometimes referred to as multiple additive regression trees, stochastic gradient boosting, or gradient boosting machines, implements gradient boosting.

Outcomes

The "Indian Crime Analysis" system has Presently accessible software has been specifically created for criminal investigations to carry out operations that cannot be completed by any other means. Consequently, it is evident that even while multiple solutions to the problem have been posted, an ideal one has not yet been created for every type of user in every city, state, and country. The accurate system would display the analysis using animated graphics and forecast the crime rate More specifically, the system would alert users to the lack of data or the likely reason for the inaccuracy if it was unable to produce accurate results.

This section provides an overview of upcoming crime data sets and algorithmic crime bases. A number of variables, such as age, gender, geography, and monthly ratios, are taken into consideration while evaluating the crime rate. Predictions are made using a variety of data sources and techniques, such as surveys of common personal data, literature evaluations, and statistical models that forecast future patterns in crime. The data imbalance resulting from the inclusion of certain minority groups in the crime classification process led to a substantial miss rate in the prediction model. Thus, we used random oversampling to fix the problem.

Predict future crime trends by extrapolating from a time series analysis of present crime trends. An algorithm is used to anticipate future crime patterns based on a time series analysis of current crime trends.algorithm Predict future crime trends with algorithms that extrapolate from a time series analysis of the current trends in crime. It is possible to predict future patterns in crime by analyzing the behavior of previously recorded data

The goal of any predictive model is to show how one predictor and one dependent variable are related to one another. These models need to be able to identify and anticipate the range of situations that could potentially impact victimization and criminal activity in the future if they are to be more accurate. This analysis projects future crime rates in a far more detailed and accurate way.

Test Cases

The output of the proposed model is presented in figure 2 to 5, Figure 2 illustrates the visualization of the data set from 2001 to 2015

.

Case 1: Forecast for Haryana's 2020 Prohibition of Child Marriage, as indicated in Figure 3.

Case 2: Forecast for Manipur murder cases through 2020, as indicated in Figure 4

Case 3: Forecast for Kerala's ARMS ACT cases through 2020, as indicated in Figure 5.

.

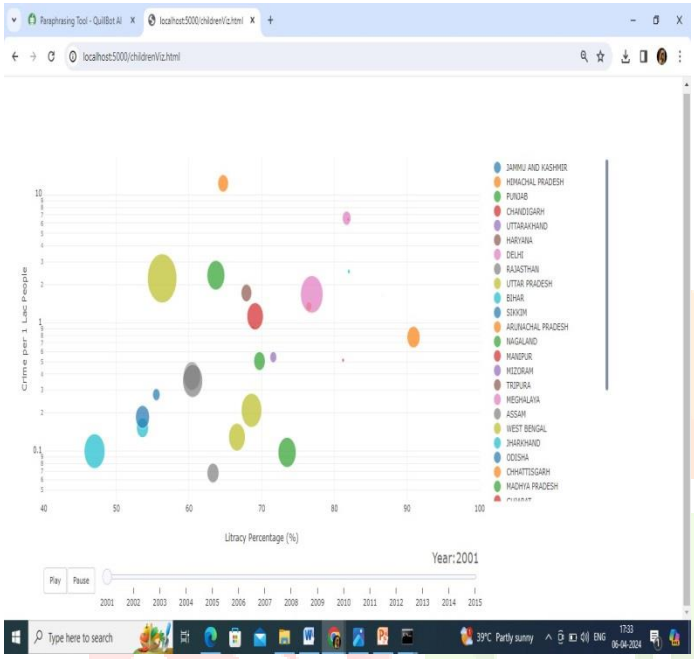


Figure 2. Visualization of the data set from 2001 to 2015

The application of machine learning technology has greatly simplified the process of Identifying links and patterns among diverse data. Finding the kind of crime that could happen given the location where it has previously happened is the main goal of this study. Using the machine learning concept, we have created a model using a training set of data that has undergone data transformation and cleansing. The model's accuracy in determining the type of crime is 0.789.

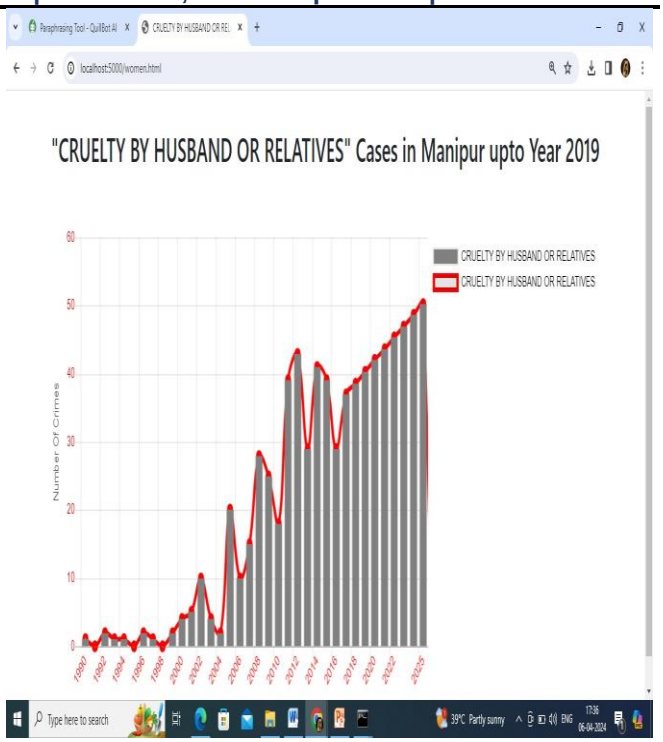


Figure 3. Predicted graph of Murder case in manipur

Prediction models try to show how one particular predictor and one dependent variable are related to each other. The range of scenarios that could influence victimization and criminality in the future must be identified and anticipated by these models in order for them to be more accurate. This analysis provides a far more detailed and accurate projection of future crime rates.

Data visualization facilitates the process of analysing a set of data. Bar, pie, line, and scatter diagrams are among the graphs; each has special characteristics of its own. We produced a number of graphs and found interesting information that helped analyze crime datasets in order to find the components that can contribute to maintaining a clean society.

Based on the locations of crimes and the most frequent types of crimes that occur there, almost all crimes may be anticipated. Different types of crimes that are committed in India are included in the data collection based on the state and year. This article's output, which takes crime kinds as its input, is the area where crimes are committed. The preprocessing of data includes data cleansing, feature selection, null value removal, and data scaling through normalization and standardization. After data preparation, null values that could significantly affect the accuracy of the model are removed from the data, and feature selection is used to choose just the attributes that are absolutely necessary and won't have an impact on the quality of the model.

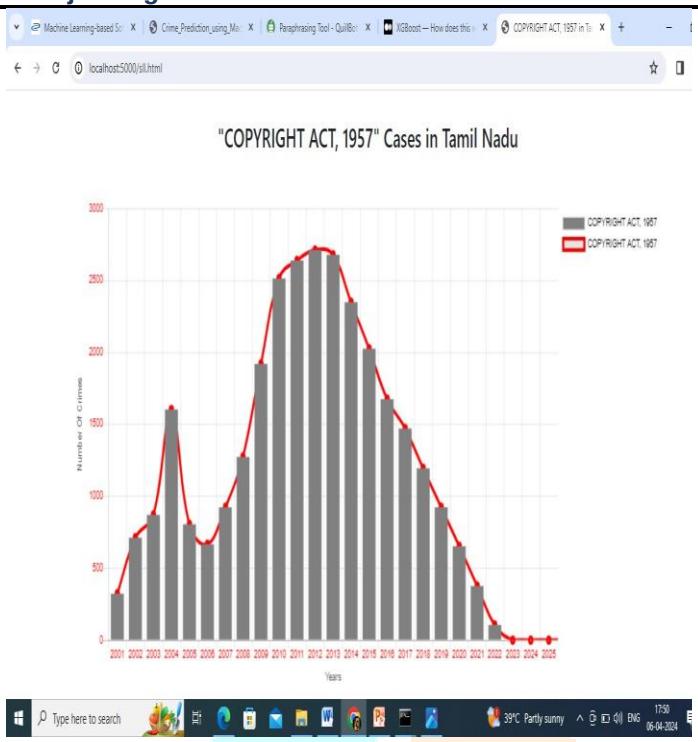


Fig4 : predicted for copyrightact in tamilnadu

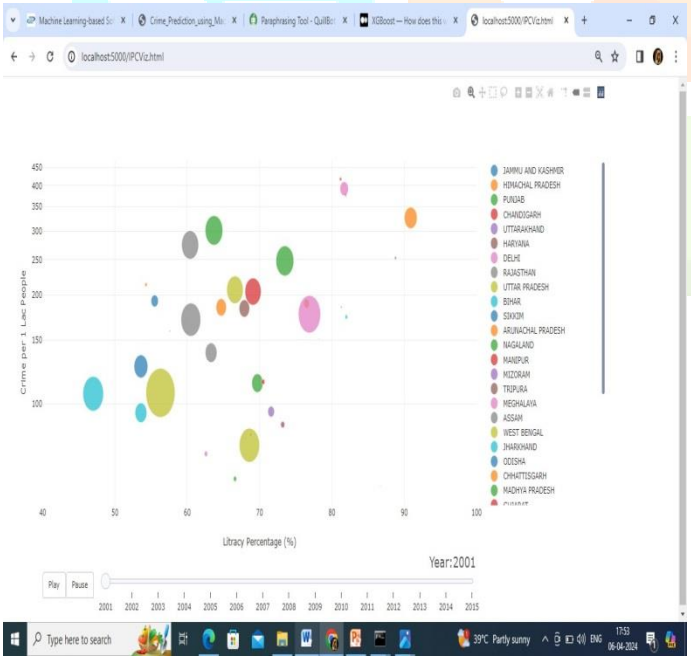


Figure 5. Visualization of the data set from 2001 to 2015

CONCLUSION AND FUTURE SCOPE

Conclusion

One of the prevailing social trends at the moment is crime prediction.

The goal of crime prediction is to lower the frequency of crimes. It achieves this by forecasting the kind of crime that might happen in the future. Here, a variety of techniques are used to analyze and predict crime, including KNN, artificial neural networks, decision trees, support vector machines, and extra trees.Based on the results, we can conclude that SVM should not be used for this dataset because of its extremely long training time. Since MLP has a relatively low accuracy, it does not perform well with this dataset. As we can see, decision trees, CNNs, and extra tree classifiers perform well on this data set in terms of training efficiency and accuracy.

Future Scope

Future iterations of this work could include improved categorization algorithms that have a higher accuracy rate in identifying criminals.

In order to safeguard the data set we use, we can also strengthen privacy and implement additional security measures. In addition, face recognition technology can be used to expand this work and forecast the identity of the criminal. If there is any suspicious alteration in the habitual movements or behavior, the system will identify it. A person may be a pickpocket, for instance, if they frequently move within the same area. This technology can also follow a person over time

References:

[1] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pentland, “Once upon a crime: towards crime prediction from demographics and mobile data”,IEEE,Proceedings of the 16th international conference on multimodal interaction, 2014, pp. 427-434.

[2] Ubon Thansatapornwatana,”A Survey of Data Mining Techniques forAnalyzing Crime Patterns”, Second Asian Conference on Defense Technology ACDT, IEEE,Jan 2016, pp. 123–128.

[3] H. Adel, M. Salheen, and R. Mahmoud, ”Crime in relation to urban design. Case study: the greater Cairo region,” Ain Shams Eng. J., vol. 7, no. 3, pp. 925-938, 2016.

[4] J. L. LeBeau, ”The Methods and Measure of Centrography and the spatial Dynamics of Rape” Journal of Quantitative Criminology, Vol.3, No.2, pp.125-141, 1987.

[5] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, Alex Pentland.”Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data”,in ACM International Conference on Multimodal Interaction (ICMI 2014)

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

[7] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav,”Crime pattern detection,analysis and prediction,International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2017.

[8] Amanpreet Singh,Narina Thakur,Aakanksha Sharma,”A review of supervised machine learning algorithms”,3rd International Conference on

Computing for Sustainable Global Development,2016.

[9] Bin Li,Yajuan Guo,Yi Wu,Jinming Chen,Yubo Yuan,Xiaoyi Zhang,"An unsupervised learning algorithm for the classification of the protection device in the fault diagnosis system",in China International Conference on Electricity Distribution (CICED),2014.

[10] Varshitha D N Vidyashree K P,Aishwarya P Janya T S,K R Dhananjay Gupta Sahana R,"Paper on Different Approaches for Crime Prediction system",International Journal of Engineering Research Technology (IJERT), ISSN: 2278-0181, 2017

[11] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," Indian J. of Sci. and Technol., vol. 6, no. 3, pp. 4219- 4225, Mar. 2013.

[12] Malathi. A,Dr. S. Santhosh Baboo,"An Enhanced Algorithm to Predict a Future Crime using Data Mining",International Journal of Computer Applications (0975 – 8887) Volume 21– No.1, May 2011.

[13] T. Beshah and S. Hill, "Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia," Proc. of Artificial Intell. for Develop. (AID 2010), pp. 14-19, 2010.

[14] K.B.S. Al-Janabi, "A Proposed Framework for Analyzing Crime Data Set using Decision Tree and Simple K-Means Mining Algorithm," in Journal of Kufa for Mathematics and Computer, Vol. 1, No. 3, 2011, pp. 8-24.

[15] A. Malathi, S.S. Baboo, "An Enhanced Algorithm to Predict a Future Crime using Data Mining," in International Journal of Computer Applications, Vol. 21, 2011, pp. 1-6. [16] C.H. Yu, "Crime Forecasting Using Data Mining Techniques," in 11th International Conference on Data Mining Workshop, 2011, pp. 779-786.