CRIME ANALYSIS AND TRACKING SYSTEM USING MACHINE LEARNING

Prof. Zarina Shaikh, Abhijeet Chhajed, Shiromani Mulgir, Ritwik Kumar Singh, Manthan Kate

Department of Computer Engineering Savitribai Phule Pune University

ABSTRACT

As we have known how crime has become anobvious way of making people and society under trouble. Due to increasing crime factors it is leading to an imbalance constituency in the country. So in orderto analyze and have a response ahead of types of criminal activities, it is necessary to understand thecrime patterns and its occurrences. The purpose of this study one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of mostly occurring crimes. The key aspect of this project is to estimate which type of crime contributes the most along with time and location, where it happened and its Occurrences. Some machine learning algorithms such as Naïve Bayes and Random Forest Classifier is implied in this work in order to classify among various crime patterns. The mainobjective of the project is to create proper Criminal analysis and Tracking System for predicting crime type, give analysis on occurrences and marking crimehotspots. Keep a reliable tracking system for proper administration based on datasets and records provided by Crime branch. Improve the efficiency and workflow of these activities and also to create social awareness among people regarding crimes and its occurrences.

Keywords- Crime, Analysis, Crime patterns, Kaggle, Naïve Bayes, Random Forest Classifier, Accuracy, Occurrences .

INTRODUCTION

Crime has become a serious thread imposed which is taken into account to grow relatively high. Crime maybe a social nuisance and price our society nearly in someways. In our society the rate is growing very rapidly especially women face many of such problems. the rationale for this may be the low pronouncement of guilt. About 10% of criminals commit 50% of crimes. this method helps keep over the crime related patterns. The crime pattern analysis requires a knowledge within the different aspects of criminology and in indicating patterns. Therefore the use of machine learning techniques and its records are required to predict the categories of crime and patterns. It imposes the uses of existing crime data and predicts the crime type and its occurrence bases on the placement and time. Researchers undergone many studies that helps in a veryanalyzing the crime patterns together with their relationsin a specific location. a number of the hotspots analyzed has become easier way of classifying the crime patterns. This ends up in assist the officials to resolve them faster. This approach uses a dataset obtained from Kaggle opensource supported various factors together with the time and space where it occurs over a specific period of your time. We implied a classification algorithms that helpsin locating the sort of crime and hotspots of the criminal

activities that takes place on the certain time and day. during this proposed system to impose a machine learning algorithms to search out the matching criminal patterns together with the assist of its category with the given temporal and spatial data and give visualization of crime analysis supported its occurrences

PROBLEM STATEMENT

Due to lack of reliable system for crime record analysisand tracking of crime occurrences there is less social awareness among people. The criminal Justice system are inefficient and need proper criminal investigation analysis. So to progress the flow of cases and achieve speedier and improve crime rate analysis system is required. So the introduction and deployment of Information technologybased case crime analysis and tracking System is needed.

The purpose of this project is to design and implement a crime analysis and occurrences prediction. This will aid thepolice in crime analysis ,tracking crime occurrences and create social awareness among people.Considering the factors and records given in datasets this system as a result depicts the crime predictions, Crime analysis and data visualizations using various supervised and unsupervised Machine learning algorithms with increase in its accuracy level and precision level

LITERATURE REVIEW

[1] Spatio-Temporal Crime HotSpot Detection and Prediction:A Systematic Literature Review Authors(UmairMuneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi , proposed IEEE Transactions on September 2016.)Description - the first objective of this study is to accumulate, summarize, and evaluate the state-of the- art for spatio-temporal crime hotspot detection and prediction techniques by conducting a scientific literature review (SLR). The authors were unable to search out a comprehensive study on crime hotspot detection and prediction while conducting this SLR. Therefore, to the most effective of author's knowledge, this study is that the premier try and critically analyze the present literature together with presenting potential challenges faced by current crime hotspot detection and prediction systems.The SLR is conducted by thoroughly consulting top five scientific databases (such as IEEE, Science Direct, Springer, Scopus, and ACM), and synthesized 49 different studies on crime hotspot detection and prediction after review. This study unfolds the subsequent major aspects:
1) the impact of knowledge mining and machine learning approaches, especially clustering techniques in crimehotspot detection; 2) the utility of your time series analysis techniques and deep learning techniques in crime trend prediction; 3) the inclusion of spatial and temporalinformation in crime datasets making the crime prediction systems more accurate and reliable; 4) the potential challenges faced by the state-of-the-art techniques and alsothe future research directions. Moreover,

the SLR aims to produce a core foundation for the research on spatio temporal crime prediction applications while highlighting several challenges associated with the accuracy of crime hotspot detection and prediction applications.

[2] Crime analysis through machine learning. Authors(Suhong Kim, Param Joshi, Parminder Singh Kalsi ,Pooya Taheri , proposed IEEE Transactions on November 2018.) Description- This paper investigates machine-learning-based crime prediction. during this work, Vancouver crime data for the last 15 years is analyzed using two different data-processing approaches. Machine-Learning predictive models, K- nearest neighbour and boosted decision tree, are implemented and a criminal offense prediction accuracy between 39% to 44% is obtained when predicting crime in Vancouver.

[3] Criminal Combat: Crime Analysis and Prediction Using Machine Learning. Authors(Amar Shukla , Avita Katal , Saurav Raghuvanshi, Shivam Sharma,proposed IEEE Xplore August 2021) Description- Crime is one among the foremost critical issues that the whole world is facing nowadays. the speed of crime should be minimized by using different techniques of machine learning so as to safeguard the world community from getting trapped into the activities of the criminals or anti-social elements. The paper identifies the crime patterns by utilizing the various mathematical and statistical models to forecast the probability of the crime. Crime datasets of the State of North Carolina are used for this purpose. The paper aims to test different statistical parameters and work out the foremost common factors that affect crime. The univariate and bivariate exploratory analysis is employed for extractingthe foremost dominant features. The Akaike InformationCriteria (AIC) method is employed to drop out unimportant attributes followed by testing of model through Mean Absolute Error (MAE), Median Squared Error (MSE) and Root Mean Squared Error (RMSE) techniques. The work worn out the paper concludes that crime predictability and criminology will be very useful in eliminating menace of crime from our society. These mathematical and statistical models can aid us within theprocess of creating our society a safer place to measure in.

[4] Survey of research of crime detection techniquesusing data processing and machine learning. Authors(S . Prabakaran and S. Mitra , IOP J. of Physics: Conf . Series, vol. 1000, 2018) Description- data processing is that the field containing procedures for locating designs or patterns during a huge dataset, it includes strategies atthe convergence of machine learning and database framework. It is applied to varied fields like future healthcare, market basket analysis, education, manufacturing engineering, crime investigation etc.Among these, crime investigation is a noteworthy application to process crime characteristics to assist the society for a more robust living. This paper survey various data processing techniques employed in this domain. This study could also be helpful in designing new strategies for crime prediction and analysis.

[5] Intelligent Crime Investigation Assistance Using Machine Learning Classifiers on Crime and Victim Information. Authors(Saqueeb Abdullah, Farah Idid Nibir ,Suraiya Salam , Akash Dey , Md Ashraful Alam and Md Tanzim Reza, IEEE Xplore May,2021) Description- so as to ascertain peace and justice in an exceedingly society , it's essential to form proper and proper investigation of crime incidents. With the expansion of the use ofcomputerized system to trace crime and violence, computer applications can help enforcement officers during a significant way. In most cases, crime incidents arekept in police database and these will be used for various helpful purpose. during this experiment, we've got collected data of crime scenario from Bangladesh Police that had features like area of crime, kind of crime, number of victims then on. Then we applied machine learning algorithms on the dataset for prediction of some attributes like criminal age, sex, race, crime method etc. We used four different algorithms for our research: K-Nearest Neighbor (KNN), Logistic Regression (LR), Random Forest Classifier (RFC), Decision Tree Classifier (DTC). Using the aforementioned algorithms with 10 fold cross validation, we achieved different accuracy from all four attribute labels starting from a mean of approximate 75%to a median of approximate 90%. Despite the clear need of further improvement, the results give clear implicationsthat it's possible to attain well performing automated system for suspect attribute prediction with further work. Finally, we ended the research by comparing and analyzingall the achieved results.

[6] A Web-based Geographical system for Crime Mappingand Decision Support. Authors(Guiyun Zhou, Jiayuan Lin, Wenfeng Zheng, IEEE 2012) Description- Timely mapping of crime locations and accurate detection of spatial concentrations of crime help to spot where crime tends to concentrate in space and time and thus provide important information for enforcement crime reduction efforts. The most objective of this work is to style and implement a Web-based Geographic data system (GIS) for crime mapping and decision support. Four hotspot mapping techniques, i.e., choropleth mapping, grid mapping, spatial ellipse mapping and kernel density mapping, are implemented within the system. The system could be a rich Internet application and is entirely supported open source software, making it affordable and efficient for several small and medium-sized police departments in developing countries. Results from the prototype development demonstrate that for a Web-based crime hotspot mapping system, rich Internet application technology together with open source software is an efficient solution.

[7] Comparison of Machine Learning Algorithms for Predicting Crime Hotspots. Authors (XU ZHANG , LIN LIU, LUZI XIAO, AND JIAKAI JI , proposed IEEE Access October 2020) Description- Crime prediction is of great significance to the formulation of policing strategies and therefore the implementation of crime prevention and control. Machine learning is that the current mainstream prediction method. However, few studies have systematically compared different machine learning methods for crime prediction. This paper takes the historical data of property crime from 2015 to 2018from a piece of an oversized coastal city within thesoutheast of China as research data to assess the predictive power between several machine learning algorithms. Results supported the historical crime data alone suggest that the LSTM model outperformed KNN,random forest, support vector machine, naive Bayes, andconvolutional neural networks. Additionally , the built

environment data of points of interests (POIs) and concrete road network density are input into LSTM model as covariates . It 's found that the model with builtenvironment covariates has better prediction effect compared with the initial model that's supported historical crime data alone. Therefore, future crime prediction should profit of both historical crime data andcovariates related to criminological theories. Not all machine learning algorithms are equally effective in crime prediction.

[8] An Implementation of Naive Bayes Classifier. Authors (Feng-Jen Yang, proposed IEEE 2018) Description- the need of classification is very demanded in world. As a mathematical classification approach, the Naïve Bayes classifier involves a series of probabilistic computations for the aim of finding the best-fitted classification for a given piece of information within a controversy domain. During this paper, an implementation of Naive Bayes classifier is described. This classifier will be used as a general kit and applicable to numerous domains of classifications. To confirm the correctness of all probabilistic computationsinvolved, a sample data set is chosen to check this classifier.

BASIC DEFINITIONS

Data Filtering or Data Cleansing is the process of detecting and correcting (or removing) corrupt orinaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleansing may be performed interactively with data wrangling tools, or as batch processing through scripting.

Data preprocessing can refer to manipulation or dropping of data before it is used in order to ensure orenhance performance, and is an important step in the data mining process. The phrase "garbage in, garbageout" is particularly applicable to data mining and machine learning projects.

Data slipting the data after pre-processing is split into test and trained attributes. The data should ideally be divided into 2 sets – namely, train set and test set. Train Set :The train set would contain the data which will be fed into the model. In simple terms, our model would learn from this data. For instance, a Regression model would use the examples in this data to find gradients in order to reduce the cost function. Then these gradients will be used toreduce the cost and predict data effectively. The sample of data used to fit the model, that is the actual subset of the dataset that we use to train the model (estimating the weights and biases in the case of Neural Network). The model observes and learns from this data and optimize its parameters. Test Set :The test set contains the data onwhich we test the trained and validated model. It tells us how efficient our overall model is and how likely is it going to predict something which does not make sense. There are a plethora of evaluation metrics (like precision, recall, accuracy, etc.) which can be used to measure the

performance of our model. The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. It is only used once the model is completely trained using the training and validation sets. Therefore, test set is the one used to replicate the type of situation that will be encountered once the model is deployed for real-time use.

Mapping : The crime features such as crime type, the date on which the crime has been occurred including the time of occurrence are first segregated. It is then mapped to an integer for easy labelling. The labelled details are further analyzed and used are used in graph plotting. Python is chosen as programming language in implementing the proposed work since it is well suited for machine learning process.

Machine learning approaches are traditionally divided into three broad categories, depending on the nature of the "signal" or "feedback" available to the learning system:

Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure inits input. Unsupervised learning can be a goal in itself(discovering hidden patterns in data) or a means towards an end (feature learning).
Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). As it navigates its problem space, theprogram is provided feedback that's analogous to rewards, which it tries to maximize.

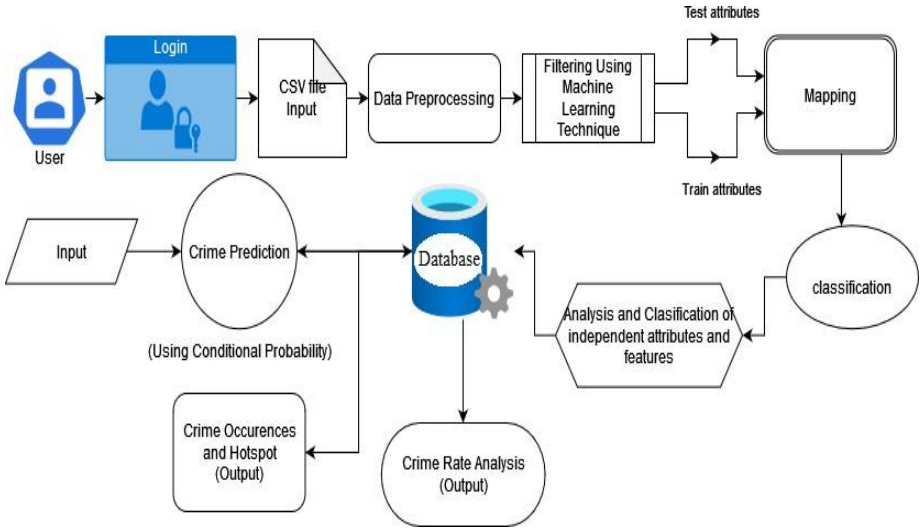
Model Generation: The process of training an ML model involves providing an ML algorithm (that is, the learning algorithm) with training data to learn from. Theterm ML model refers to the model artifact that is created by the training process.
The training data must contain the correct answer, whichis known as a target or target attribute. The learning algorithm finds patterns in the training data that map the input data attributes to the target (the answer that you want to predict), and it outputs an ML model that captures these patterns.

Random forest is a technique used in modeling predictions and behavior analysis and is built on decision trees. It contains many decision trees representing a distinct instance of the classification of data input into the random forest. The random forest technique considers the instances individually, takingthe one with the majority of votes as the selected prediction.

Naïve Bayes Classification : The reason behind the application of Naïve Bayes is that crime prediction usually concerns with the temporal and spatial data. The independent effect among the attribute values is first analyzed since the selected crime attributes possess an independent effect

upon them. They are used in creating a model by providing a training using crime data that are related to robbery, burglary, murder, sexual abusing, armed robbery, chain snatching, gang rape .

PROPOSED SYSTEM



The data obtained from Kaggle is first undergoes through the process of data cleansing or we can say data filtering in this process all corrupted and incorrect data are removed this process is part of data pre-processing, using machine learning techniques filter and wrapper in order to remove irrelevant and repeated data values the data is filtered.

The data further undergoes into splitting process. The dataset is then classified into two types of sets, test and trained data set. The model is trained by dataset both training and testing .The model observes and learns from this data and optimize its parameters.

It is then followed by mapping process the crime type, year, month, time, date, place are mapped to an integer values for ensuring classification easier and efficient for prediction process further. The independent effect between the attributes are analyzed initially by using Naïve Bayes. Bernouille Naïve Bayes is used for classifying the independent features extracted. The crime attributes arelabelled which allows the system to analyze the occurrenceof crime at a specific time and location.

Finally, the crime which occurs the most along with spatialand temporal information is gained. Further the in prediction using condition probability the expected crime type is predicted by extending the supported crime features. The performance of the prediction model is find out by calculating accuracy rate. For designing such a system there is use of python language and libraries

such as NumPy, Pandas, Matplotlib are used for data analysis, visualization and machine leaning modelling and training.

ADVANTAGES

1. The proposed algorithm in this system is suited for the crime pattern detection since most of the attributes dependson the time and location factors.

2. It also overcomes the problem of analyzing independent effect of the attributes and give unbiased outcome.

3. The accuracy of this system is relatively high when compared to other machine learning prediction model.

4. Since this system allows the user to have a reliable source for crime analysis and have precise and efficient info which could be handy in social awareness regardingcrime hotspots.

CONCLUSION

The difficulty in dealing with the nominal distribution and real valued attributes is overcome by using three classifierssuch as Multinominal NB and Gaussian NB and Random forest. Excessive training time is not required and it serves to be best suited for real time predictions. Thus the crime that occur the most could be predicted and spotted using Classification. The performance of the algorithm is also calculated by using some standard metrics. The metrics include average precision, recall, F1 score and accuracy are mainly concerned in the algorithm evaluation. The accuracy value could be increased much better byimplementing machine learning algorithms.

REFERENCES

1. Crime analysis through machine learning (Suhong Kim, Param Joshi, Parminder Singh Kalsi,Pooya Taheri, IEEE Transactions on November 2018.)
2. Spatio-Temporal Crime Hotspot Detection and Prediction: A Systematic Literature Review (Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi, IEEE Transactions on September 2016.)
3. Comparison of Machine Learning Algorithms for Predicting Crime Hotspots (XU ZHANG , LIN LIU, LUZI XIAO, AND JIAKAI JI , IEEE AccessOctober 2020)
4. Crime Prediction Using K-Nearest Neighboring Algorithm (Akash Kumar, Aniket Verma, Gandhali Shinde, Yash Sukhdeve , Nidhi Lal, IEEE Xplore May, 2020)
5. An Implementation of Naive BayesClassifier(Feng-Jen Yang, IEEE 2018)
6. Criminal Combat: Crime Analysis and Prediction Using Machine Learning (Amar Shukla,

Avita Katal, Saurav Raghuvanshi, Shivam Sharma, IEEE Xplore August 2021)

7. Intelligent Crime Investigation Assistance Using Machine Learning Classifiers on Crime and Victim Information. (Saqueeb Abdullah, Farah Idid Nibir, Suraiya Salam, Akash Dey, Md Ashraful Alam and Md Tanzim Reza, IEEE Xplore May,2021)
8. Survey of analysis of crime detection techniques using data mining and machine learning (S. Prabakaran and S. Mitra, IOP J. of Physics: Conf. Series, vol. 1000, 2018)
9. A Web-based Geographical Information System for Crime Mapping and Decision Support (GuiyunZhou, Jiayuan Lin, Wenfeng Zheng, IEEE 2012)