

Crime Mapping with Automatic Classifier System using Machine Learning and GIS

M. Deepika, S.Ancy

Abstract— Crime rate is increasing over the years, and it remains a great challenge for the government to track the crimes convicted. Each area has a pattern of which type of crime is happening and the crime knowledge is inevitable to control from the crime happening. Crime occurs in a sequence leaving hidden patterns. Thus the crime data is to be processed for finding underlying patterns, this project finds the patterns and insights about the crime data. Majorly being an unstructured data, this is been preprocessed and checked for future values. Crimes convicted are collected from a particular area (Indore in our case) and checked for predictions using Multi class Classification Algorithms like Random Forest and the future crime to be convicted in an area is predicted and visualizations are made accordingly. Many classification algorithms like support vector machines, decision trees, and random forest are used to classify and random forest shows better accuracy. Features to be given as input and output are selected by visualizing the data by graphs and plots.

Keywords: Criminal-network analysis, Geographic Information System, B plots, multi-class classification problem.

I. INTRODUCTION

In this environment analysis of patterns in crimes explains the reason for occurrence of crime in certain areas. For this analysis of data efficient algorithms are used that help in identifying the patterns in crime. The inputs to our algorithms are time (hour, day, month, and year), place (latitude and longitude), and class of crime: Act 379 – Robbery, Act 13 – Gambling, Act 279 – Accident, Act 323 – Violence, Act 302 – Murder, Act 363 – Kidnapping. The output is depicted in the form of various classes of crime that has been predicted to have occurred earlier. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests. We also perform multiple classification tasks - First the prediction is made to identify which among the six classes of crimes have occurred already. Later the differentiation is made between violent and non-violent crimes. The scope of paper is given below:

Analysis of crime using smart phones is designed for patrol officers who are assigned in particular areas to ensure safety of people. A trial on randomized controlled methods was conducted. Some of entry and exit level surveys were conducted for evaluating the usage of device and applications. Another benefit of this survey is to attain knowledge on crime clusters. Various data sources are used to identify crime clusters that every individual perceives the most. Several focusing groups were used for exploring the

applications in context manner.

Machine learning detects the patterns automatically in data. Information extraction from huge volumes of data has become a highly demandable.

Predicting Future Crime Spots: We can predict the occurrence of future crimes by using the existing data. For example robberies that took place one area could match with more robbery that is likely to take place in the surrounding areas in the future. This system will prioritize the areas where the police should patrol more.

Predicting Who Will Commit a Crime: Using Face Recognition to predict if a individual will commit a crime before it happens. The system will detect about the changes in the activities or unusual behavior. A pickpocket or chasing in the particular area can be identified if a person is walking back and forth at the same place again and again. area for a future crime. It will also track individual overtime.

Pretrial Release and Parole: Once the individual is charged with a crime, their release on bail should be done by judges with their good judgement because the individual's pretrial release may harm the society.

This judgement will attempt to determine if someone is at a flight risk, or some danger is about to occur to society, or at risk to harm a witness if released. It is an imperfect system that must be reconsidered.

II EXISTING SYSTEM & METHODOLOGY

The Existing system deals with structured crime data where they come from tabular column data tables and they are preprocessed for making predictions using classification algorithms and this system uses features without analyzing the parameters that have a higher degree of correlation for predicting the crime.

The disadvantage of crime mapping is that it requires a large number of data in order to be effective, and this information must be retrieved by police data bases.

This information must then be processed into the GIS. As crime mapping software become more and more advanced and sophisticated and increasingly critical to police forces, these departments will need to update their hardware and software, as well as hire more trained and professional crime analysts.

III PROPOSED SYSTEM

A Multiclass Classification system is developed where the preprocessed unstructured data is fed into the system and the features are selected according to the higher degree of correlation. This system predicts and analyzes using

Revised Manuscript Received on 14 September, 2019.

M. Deepika, Assistant Professor, Department of Information Technology, St.Peter's Institute of Higher Education and Research, Chennai, Tamilnadu, India.(E-mail: deepikamanickam8@gmail.com)

S.Ancy, Assistant Professor, Department of Information Technology, Jeppiaar Institute of Technology, Chennai, Tamilnadu, India. (E-mail: sancyit@gmail.com)

Random Forest Algorithm where which set of crime has take place.

Advantages of proposed system

Crime mapping is tool used mark location in victim well as offenders in order to locate and prevent localized crime with this information, the police departments are able to identify where the criminal activity is more prominent, and also where the victims are located.

This information would show which neighborhoods are the most critical spots.

With this information, the department can track how the criminals move from location to location.

This is done by breaking down the crime by time of day and by day of weeks and by day of the month.

IV SYSTEM ARCHITECTURE

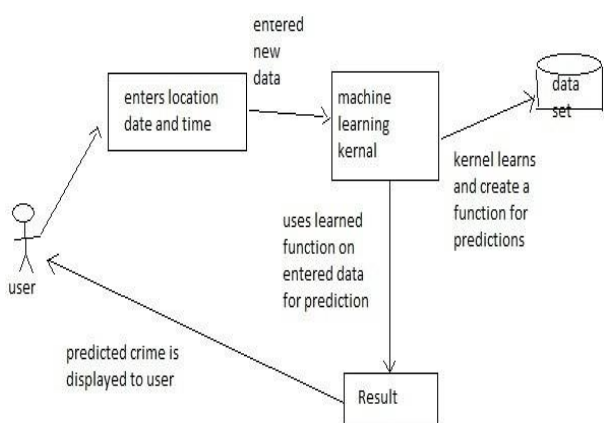


Fig 1: Identifying the subsystem for control and communication

In the figure 1 system architecture diagram show, the design process for identifying the subsystems making up the system and framework for subsystem control and communication.

V MODULES & RESULTS

1. Predicting crimes

Raw dataset obtained from the police stations and government offices are collected and preprocessed. Preprocessing techniques includes data cleansing, addition of necessary values etc. Features are tested for correlation and each feature is tested against the crimes.

Visualization of graphs are made for finding the maximum correlation of features relating to the crimes convicted. Features with lower correlation are discarded and the dataset is cleaned.

Features to be taken as input are decided (Latitude, Longitude, and timestamp in our case) and the output features

i.e., classes of crime are fed as the output. Multiclass classification algorithms are used for the prediction of the crimes convicted. Random Forest Algorithm showed best accuracy for the prediction for crimes convicted .

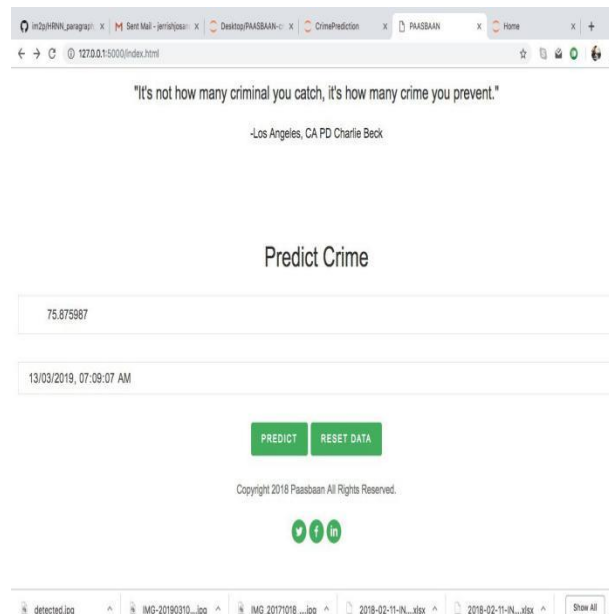


Fig 2: Crime Predication

2. Analyzing and visualizing crimes

Crimes are categorized in the dataset itself. Graphs are plotted by plotting independent variables against dependent variables.

Box plots and scatter plots shows the correlation between the features and the plots are visualized.

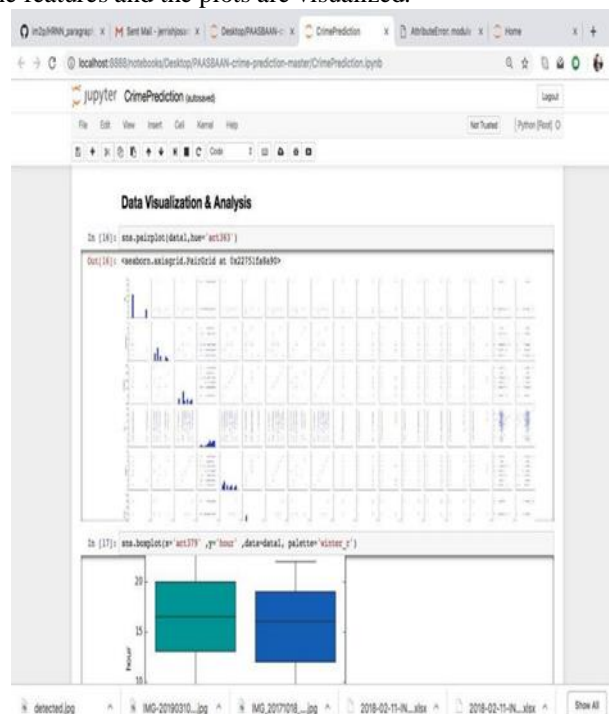


Fig 3: Data Visualization And Analysis

The overall dataset is visualized and insights are retrieved from it.

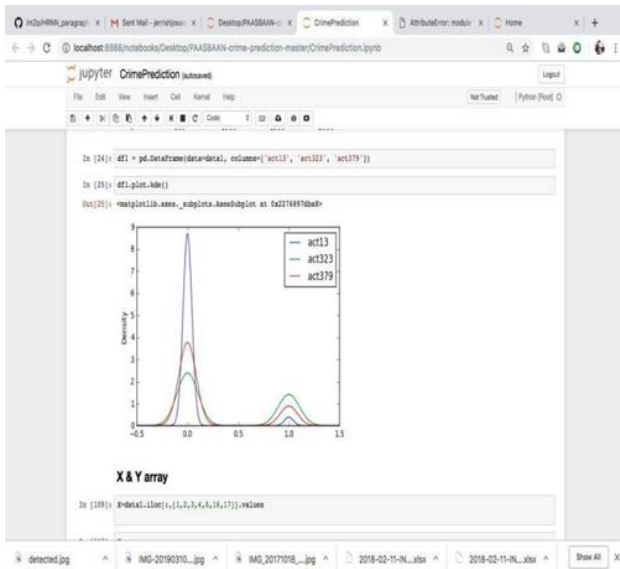


Fig 4: Analysis Crime Wall

Multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests.

Decision tree visualizations have the decision node and its features are compared with the target nodes distributions. The feature decision node is tested and the decision nodes splits the target nodes. Leaf node purity, which affects the Leaf with low variance among the target nodes or target nodes distributions (classification) are much more reliable predictors.

High-crime- density areas are commonly known as hot spots. These can be detected by using a effective method known as mapping of crimes. Crime hot spot represents area where the number of criminal events is more than other places, or the area where people higher risk of victimization.

The method which is used most widely for detecting crime hot spots is the spatial clustering in the original crime data.

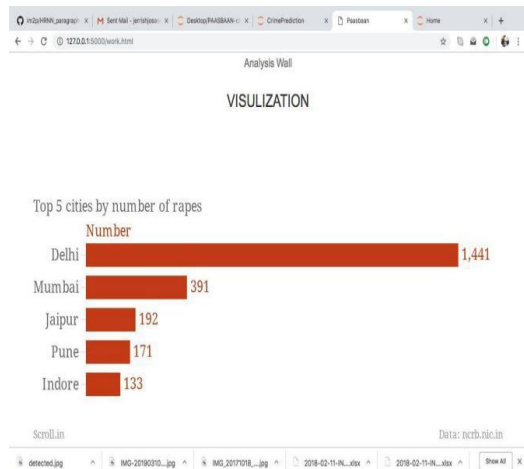


Fig 5: Density of Crime

The analysis and recording of crime data becomes more difficult because the crime pattern are complex.

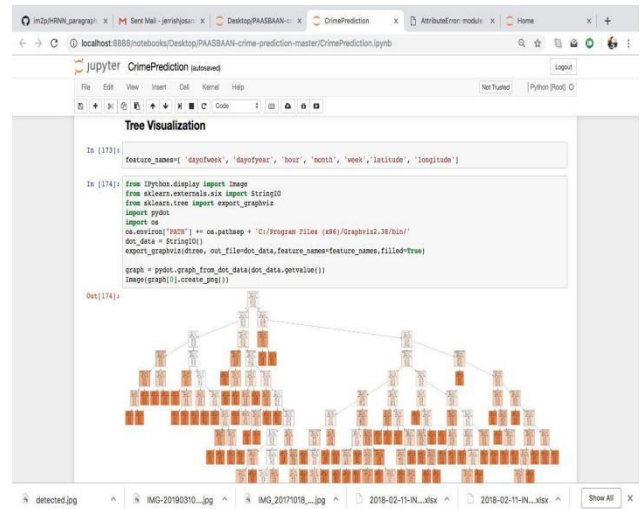


Fig 6 : Tree visualization

In this paper analysis of crime is designed in a better system to overcome the above mentioned difficulties. The application of spatial ellipse attempts to differentiate crime hot spots.

VI CONCLUSION AND FUTURE ENHANCEMENT

The major challenge was to classify the crime categories (ie) multi-category classifying problem. Initially high accuracy in our initial data set was not obtained because there was not that much prediction on the data set. We found that our meaningful classifier system categorized the fewer group of classes of crime into larger groups of classes of crime. This categorization enabled us to identify the structures in the data set. High accuracy was achieved on prediction.

However classification on categories of crime like violent and non-violent did not yield accurate results with similar classifiers. This classification problem was very difficult in achieving accuracy. Better choice and consideration is required to consolidate violent and non-violent classes of crime.

The extension of this work includes modeling of the data in time series by understanding temporary connections that could be used to predict sudden changes in classes of crime. Other additional areas to work on includes implementing much more accurate multi-category classifier for classifying violent and non-violent classes of crimes.

REFERENCES

1. Bogomolov, Andrey and Lepri, Bruno and Staiano, Jacopo and Oliver, Nuria and Pianesi, Fabio and Pentland, Alex. 2014. Once upon a crime: Towards crime prediction from demographics and mobile data, Proceedings of the 16th International Conference on Multimodal Interaction.
2. Yu, Chung-Hsien and Ward, Max W and Morabito, Melissa and Ding, Wei. 2011. Crime forecasting using data mining techniques, pages 779-786, IEEE 11th International Conference on Data Mining Workshops (ICDMW)

- 3 Kianmehr, Keivan and Alhadj, Reda. 2008. Effectiveness of support vector machine for crime hot-spots prediction, pages 433-458, Applied Artificial Intelligence, volume 22, number 5.
- 4 Toole, Jameson L and Eagle, Nathan and Plotkin, Joshua B. 2011 (TIST), volume 2, number 4, pages 38, ACM Transactions on Intelligent Systems and Technology
- 5 Wang, Tong and Rudin, Cynthia and Wagner, Daniel and Sevieri, Rich. 2013. pages 515- 530, Machine Learning and Knowledge Discovery in Databases