IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CRIME DATA ANALYSIS USING MACHINE LEARNING

1st Mr. Dr. S. SASIKALA
AP - Cyber security Engineering
Paavai Engineering College
(Anna University Affiliated)
Namakkal, India

2nd K. MADHAN
Cyber security Engineering
Paavai Engineering College
(Anna University Affiliated)
Namakkal, India

3rd S. SENTHAMARAI KANNAN
Cyber security Engineering
Paavai Engineering College
(Anna University Affiliated)
Namakkal, India

Abstract: Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict the type of crime activity which have high probability for given location in terms of latitude and longitude and date and also we can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Instead of focusing on causes of crime occurrence like criminal background of offender, political enmity etc. we are focusing mainly on crime factors of each day.

Keywords – Sign language, Arduino, ASL, Sound Glove, Dumb, Gesture recognition, Deaf, Flex.

I. INTRODUCTION:

Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also, the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc. have been decreased while crimes like murder have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime sensitive areas. So, for building such a powerful crime analytics tool we have to collect crime records and evaluate it.

Crime analysis is a crucial task for law enforcement agencies, as it helps in understanding patterns, identifying trends, and developing effective strategies to combat criminal activities. Clustering algorithms can be employed as an approach for crime analysis, enabling the identification of similar patterns among criminal incidents. Here's an outline of an approach that utilizes clustering algorithms for crime analysis.

Data collection: Gather relevant data on criminal incidents, including information such as the type of crime, location, time, and any other relevant attributes. This data can be obtained from crime reports, incident databases, or other reliable sources.

Data preprocessing: Clean and preprocess the collected data to ensure its quality and suitability for analysis. This step involves handling missing values, removing inconsistencies, and standardizing data formats.

Feature selection: Identify the relevant features or attributes that contribute to crime patterns. These could include geographical coordinates, time of occurrence, crime type, modus operandi, and any other relevant information.

II. EXISTING SYSTEM:

Copylink describes as an integrated information and knowledge management environment trying to manage massive amount of information on criminals. It has been developed at the University of Arizonas Artificial Intelligence Lab in collaboration with the Tucson 6 and Phoenix Police departments. The main aim of this project is to develop information and knowledge management systems, technologies, and methodologies appropriate for capturing, accessing, analyzing, visualizing, and sharing law enforcement-related information in social and organizational contexts.

COPLINK consists of two main parts namely, COPLINK connect and COPLINK detect. COPLINK connect is responsible for sharing of data from different police departments, while COPLINK detect is used to uncover different crime associations that exist in police databases. Detect is mostly concerned with creating associations and linkages among various aspects of a crime. The following drawbacks can be identified in this system. (a) COPLINK is complex and requires user training. (b) Although the system is able to identify linkages among specific concepts residing in existing database, it does not support data mining.

Another solution is Ozguls Crime Prediction Model (CPM) , which predicts offenders of terrorist events based on location, date, and modus operandi attributes. It uses both solved and unsolved crime information, learning from attributes of each crime. These crimes are clustered according to the attributes. As an example, similar located crime clusters.

These clusters contain crimes according to specific attributes. Similarity scores are measured for each crime and total similarity between two crimes is measured using Euclidean distance. Using these values similarly behaving solved and unsolved crimes are put into more accurate clusters. CPM looks for perpetrators of crimes, with the assumption that majority of crimes in a single cluster were committed by the same single offender group.

III. PROPOSED SYSTEM:

The platform which is going to be developed directly targets the crime domain. It has the ability to analyze crime data in three different ways, namely Descriptive, prescriptive, and predictive analytics. The most important part of this platform is that, it is designed to be scalable to support different types of crime data analysis.

Different user requirements can be achieved through developing simple plug-ins to the system and scale the platform. Descriptive Analyzer use both quantitative and qualitative data along with analytical techniques. Qualitative data and analytical techniques refer to non-numerical data, as well as the

examination and interpretation of observations for the purpose of discovering underlying meanings and patterns of relationships.

Descriptive analyzer basically provides relationships between crimes and identifies the pattern of crimes and temporal and spatial relationships between crimes. In this platform, predictive analytics methods are mainly used for predicting category of a crime which can be occurred somewhere at a given time. Predict category system integrates population and race data accordingly to the given crime data set. Using machine learning techniques, it can predict the category of crime that can occur.

IV. SOFTWARE REQUIREMENTS:

1. Pandas

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open-source data analysis / manipulation tool available in any language. It is already well on its way toward this goal.

2. NumPy

NumPy is a library for Python, adding support for large, multidimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. NumPy is an open-source software and has many contributors. In comparison, MATLAB boasts a large number of additional toolboxes, notably Simulink, whereas NumPy is intrinsically integrated with Python, a more modern and complete programming language.

3. Scikit-learn

Scikit-learn (formerly scikits. learn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines(svm), random forest, gradient boosting, k means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. It has been built on top of NumPy, SciPy and matplotlib.

4. Scikit-plot

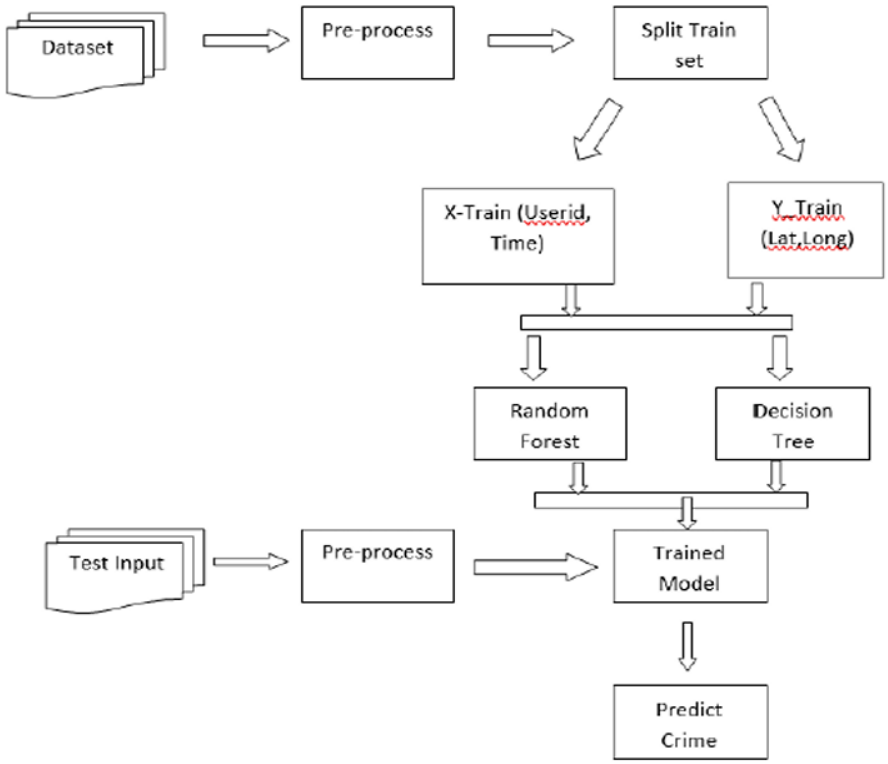
Scikit-plot is the result of an unartistic data scientist's dreadful realization that visualization is one of the most crucial components in the data science process, not just a mere afterthought. Gaining insights is simply a lot easier when you're looking at a colored heatmap of a confusion matrix complete with class labels rather than a singleline dump of numbers enclosed in brackets. Besides, if you ever need to present your results to someone 15 (virtually any time anybody hires you to do data science), you show them visualizations, not a bunch of numbers in Excel. All in all, it is an intuitive library used to add plotting functionality to a scikit-learn object.

5. Matplotlib's pyplot

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits. It provides an object-oriented API for embedding plots into applications using generalpurpose GUI toolkits like Tkinter, wxPython, Qt or GTK.matplotlib. Pyplot provides a MATLAB-

like plotting framework pylab combines pyplot with NumPy into a single namespace. This is convenient for interactive work, but for programming it is recommended that the namespaces be kept separate.

V. SYSTEM ARICHITECTURE:



Takes clarity around a specific goal, and when a larger team is involved, collaboration and alignment. For that reason, the design process is crucial because it's the disciplined approach that guides creative teams towards a clear and meaningful goal for their design work. The design is a plan or drawing produced to show the look and function or workings of an object before it is made. Unified Modeling language (UML) is a standardized modeling language enabling developers to specify, visualize, construct and document artifacts of a software system. Thus, UML makes these artifacts scalable, secure and robust in execution. UML is an important aspect involved in object-oriented software development. It uses graphic notation to create visual models of software systems.

The above figure represent use case diagram of proposed system, where user inputs dataset, we pre-process dataset, the algorithm Decision Tree and K-means clustering to generate the trained model to predict the crime type. The actor and use case is represented. An eclipse shape represents the use case namely input image, pre-process, Split features, prediction and output.

In the deployment diagram the UML models the physical deployment of artifacts on nodes. The nodes appear as boxes, and the artifacts allocated to each node appear as rectangles within the boxes. Nodes may have subnodes, which appear as nested boxes. A single node in a deployment diagram may conceptually represent multiple physical nodes, such as a cluster of database servers.

Collection and analysis of crime related data are imperative to secure agencies. the use of a coherent methods to classify these data based on the rate and location of occurrences, detection of the hidden pattern among the committed crimes at different times, and prediction of their future relationship are the most important aspects that have to be addressed. One of the most popular approaches is hot spot analysis. Some of the most popular approaches used for this purpose of point pattern analysis and clustering/distances statistics. Another popular approach is the discovery of pattern or trends through various techniques from data mining, text mining and spatial analysis, and self-organizing maps. An

crime analysis tool should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action.

VI. CONCLUSION:

Examined the accuracy of class and prediction based totally on different check sets. Classification is done based on the Bayes theorem which showed more than 90% accuracy. Using this algorithm, we trained numerous news articles and build a model. For testing we are inputting some test data into the model which shows better results. Our system takes elements attributes of an area and preprocessing offers the frequent patterns of that place. The pattern is used for constructing a model for decision tree.

Corresponding to each place we build a model by training on these frequent patterns. Crime patterns cannot be static since patterns change over time. By training means we are teaching the system based on some particular inputs. So the machine automatically learns the converting patterns in crime through examining the crime patterns. Also the crime elements trade over time.

By sifting through the crime data we have to identify new factors that lead to crime. Since we are considering only some limited factors full accuracy cannot be achieved. For getting better results in prediction we have to find more crime attributes of places instead of fixing certain attributes. Till now we trained our system using certain attributes but we are planning to include more factors to improve accuracy.

Our software predicts crime prone regions in India on a particular day. It will be more accurate if we consider a particular state/region. Also another problem is that we are not predicting the time in which the crime is happening.

VII. RESULT AND DISCUSSION:

7.1 RESULT:

After finishing the development of any computer based system the next complicated time consuming process is system testing. During the time of testing only the development company can know that, how far the user requirements have been met out, and so on. Software testing is an important element of the software quality assurance and represents the ultimate review of specification, design and coding. The increasing feasibility of software as a system and the cost associated with the software failures are motivated forces for well-planned through testing.

System testing is done for checking the server name of the machines being connected between the customer and executive. The product information provided by the company to the executive is tested against the validation with the centralized data store. System testing is also done for checking the executive availability to connect to the server. The server name authentication is checked and availability to the customer proper communication chat line viability is tested and made the chat system function properly. Mail functions are tested against the user concurrency and customer validate.

Unit testing focuses on verifying the effort on the smallest unit of software module. The local data structure is examined to ensure that the date stored temporarily maintains its integrity during all steps in the algorithm's execution. Boundary conditions are tested to ensure that the module operates properly at boundaries established to limit or restrict processing

Data can be tested across an interface. One module can have an inadvertent, adverse effect on the other. Integration testing is a systematic technique for constructing a program structure while conducting tests to uncover errors associated. It begins after the integration testing is successfully assembled. Validation succeeds when the software functions in a manner that can be reasonably accepted by the client.

In this the majority of the validation is done during the data entry operation where there is a maximum possibility of entering wrong data. Other validation will be performed in all process where correct details and data should be entered to get the required results.

Collection and analysis of crime related data are imperative to secure agencies. the use of a coherent methods to classify these data based on the rate and location of occurrences, detection of the hidden pattern among the committed crimes at different times, and prediction of their future relationship are the most important aspects that have to be addressed.

One of the most popular approaches is hot spot analysis. Some of the most popular approaches used for this purpose of point pattern analysis and clustering/distances statistics. Another popular approach is the discovery of pattern or trends through various techniques from data mining, text mining and spatial analysis, and self-organizing maps. An crime analysis tool should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action.

VIII. ACKNOWLEDGEMENT:

Certainly, acknowledging the support and contribution of Paavai Engineering College, particularly the teaching and non-teaching staff of the Department of Cyber security, is essential. Additionally, expressing gratitude to our parents, friends, and all those who have provided direct or indirect support for this research is imperative.

IX. REFERENCE:

[1] De Bruin ,J.S.,Cocx,T.K,Kosters,W.A.,Laros,J. and Kok,J.N(2006) Data mining Approaches to criminal carrier analysis ,"In Proceeding of the Sixth International Conference on Data Mining (ICDM"06),Pp. 171-177.

[2] Manish and M. P. GuptaGupta1, B.Chandra1 1,200Information System.

[3] Nazlena Mohamad Ali1, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and Shahrul Azman Mohd Noah2, 2010 Visual Interative Malaysia Crime News Retrival system 7 crime data mining for Indian Police.

[4] Chung-Hsien Yu, Max W.Ward, Melissa Morabito and Wei Ding, "Crime Forecasting Using Data Mining Techniques",2011 11th IEEE International Conference on Data Mining Workshops.