International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 650-654 www.ijaem.net ISSN: 2395-5252

Crime Suspect Prediction Using Machine Learning

[1]Sharath H A Associate Professor, [2]Krithika B S, [3]L.Anisha, [4]Neha Fathima, [5]Nidhishree S
[1,2,3,4,5] *Dept. of ISE MITM.*

Date of Submission: 19-06-2020 Date of Acceptance: 05-07-2020

ABSTRACT: In recent years the ML is data analyzing techniques that used to analyze crime data previously stored from various sources to find patterns, suspects and trends in crimes. In additional, it can be applied to increase efficiency in solving the crimes faster and also can be applied to automatically notify the crimes. However, there are many ML techniques. In order to increase efficiency of crime detection, it is necessary to select the ML techniques suitably. This paper reviews the literatures on various ML applications, especially applications that applied to solve the crimes. Survey also throws light on research gaps and challenges of crime ML. In additional to that, this paper provides insight about the data mining for finding the patterns, suspects and trends in crime to be used appropriately and to be a help for beginners in the research of crime ML.

Keywords-Random forest algorithm, Naïve Bayes algorithm,

I. INTRODUCTION

Crime prevention and detection become an important trend in crime and a very challenging to solve crimes. Several studies have discovered various techniques to solve the crimes that used to many applications. Such studies can help speed up the process of solving crime and help the computerized systems detect the criminals automatically. In addition, the rapidly advancing technologies can help address such issues. However, the crime patterns are always changing and growing. The crime data previously stored from various sources have a tendency to increase steadily. As a consequence, the management and analysis with huge data are very difficult and complex. To solve the problems previously mentioned, ML techniques employ many learning algorithms to extract hidden knowledge from huge volume of data. ML is data analyzing techniques to find patterns and trends in crimes. It can help solve the crimes more speedily and also can help alert the criminal detection automatically.

ML, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While ML and knowledge discovery in databases (or KDD) are frequently treated as synonyms, ML is actually part of the knowledge discovery process.

PROBLEM STATEMENT

Determining the crime suspects is a major challenge in today's world to reduce the crimes and to take the precautionary measures to avoid crimes. The system helps to decrease the local people traffic in police station, it also helps police to increase investigation rate that in turn decrease the crime rate, the system also does analysis of criminal data which helps to predict crime in particular region. System is equipped with real time information. The system can integrate number of modules to increase the efficiency and problem solving of police, such as forensic report, chemical report, ballistic report and it can integrate the court reports.

PROPOSEDSYSTEM

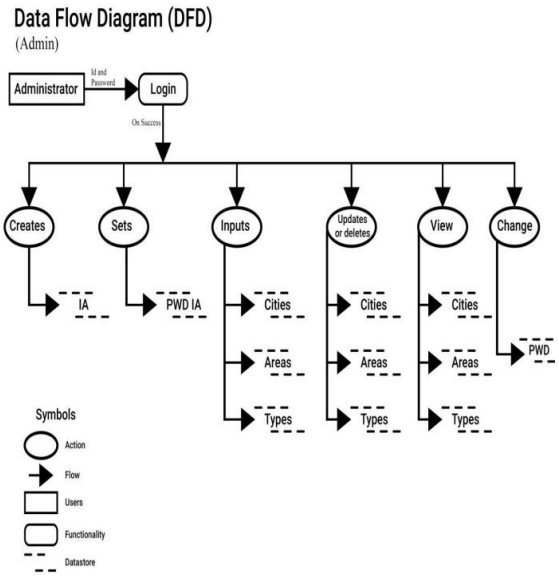
Proposed system is applicable in the field of crime. Proposed system includes modeling of crimes for finding suitable algorithms to detect the crime, precise detection, data preparation and transformation, and processing time. Proposed system identifies crime behavior, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Proposed system is automation for complaints registration, crime pattern prediction based on the previous crime details collected from various sources.

Data Flow Diagram: A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system. DFDs can also be used for the visualization of data processing (structured design). On a DFD, data items flow from an external data source or an internal data store to an internal data store or an external data sink, via an internal process.

A DFD provides no information about the timing of processes, or about whether processes

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 650-654 www.ijaem.net ISSN: 2395-5252

will operate in sequence or in parallel. It is therefore quite different from a flowchart, which shows the flow of control through an algorithm, allowing a reader to determine what operations will be performed, in what order, and under what circumstances, but not what kinds of data will be input to and output from the system, nor where the data will come from and go to, nor where the data will be stored (all of which are shown on a DFD).



Use Case Diagram:

A use case diagram is the primary form of system/software requirements for a new software program underdeveloped. Use cases specify the expected behaviour (what), and not the exact method of making it happen (how). Use cases once specified can be denoted both textual and visual representation (i.e. use case diagram). A key concept of use case modelling is that it helps us design a system from the end user's perspective. It is an effective technique for communicating system behaviour in the user's terms by specifying all externally visible system behaviour.

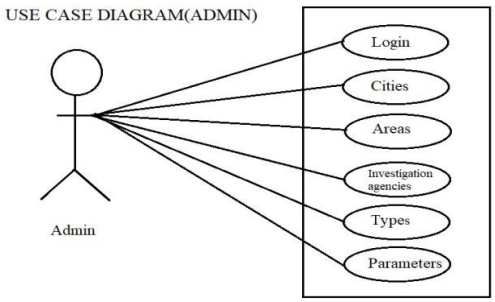


Fig: Use case diagram for admin

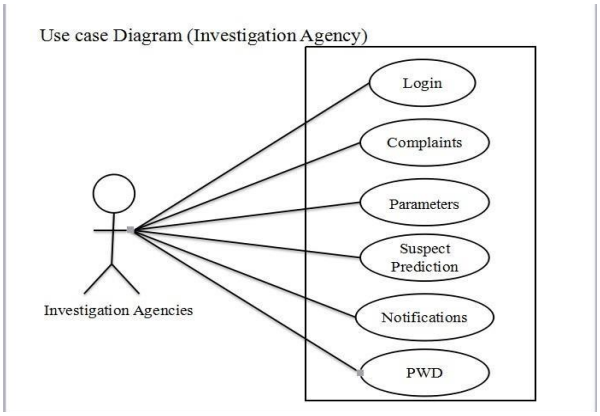


Fig: Use case diagram for investigation agencies

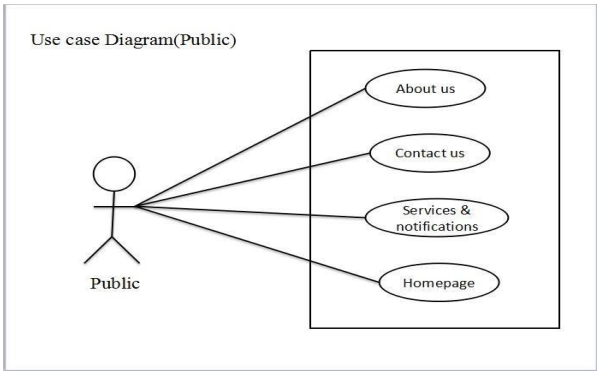


Fig: Use case diagram for public

II. METHODOLOGY

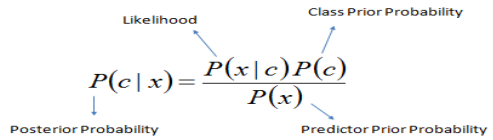
The project involves the following algorithms:

- **Naïve Bayes Algorithm**
- **Random Forest Algorithm**

Naïve Bayes Algorithm :In machine learning, **naïve Bayes classifiers** are a family of simple "probabilistic classifiers" based on applying Bayes'

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 650-654 www.ijaem.net ISSN: 2395-5252

theorem with strong (naïve)independence assumptions between the features. They are among the simplest Bayesian network models. But they could be coupled with Kernel density estimation and achieve higher accuracy levels.Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Random Forest Algorithm: 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 overfitting to their training set. The general method of random decision forests was first proposed by Ho in 1995.[1]Ho established that forests of trees splitting with oblique hyper planes can gain accuracy as they grow without suffering from overtraining, as long as the forests are randomly restricted to be sensitive to only selected feature dimensions. A subsequent work along the same lines concluded that other splitting methods behave similarly, as long as they are randomly forced to be insensitive to some feature dimensions. Note that this observation of a more complex classifier (a larger forest) getting more accurate nearly monotonically is in sharp contrast to the common belief that the complexity of a classifier can only grow to a certain level of accuracy before being hurt by overfitting. The explanation of the forest method's resistance to overtraining can be found in Kleinberg's theory of stochastic discrimination.

The early development of Breiman's notion of random forests was influenced by the work of Amit and Geman who introduced the idea of searching over a random subset of the available decisions when splitting a node, in the context of growing a singletree. The idea of random subspace selection from Ho was also influential in the design of random forests. In this method a forest of trees is grown, and variation among the trees is introduced by projecting the training data into a randomly chosen subspace before fitting each tree or each node. Finally, the idea of randomized node optimization, where the decision at each node is selected by a randomized procedure, rather than a deterministic optimization was first introduced by Dietterich.

III. RESULT ANALYSIS

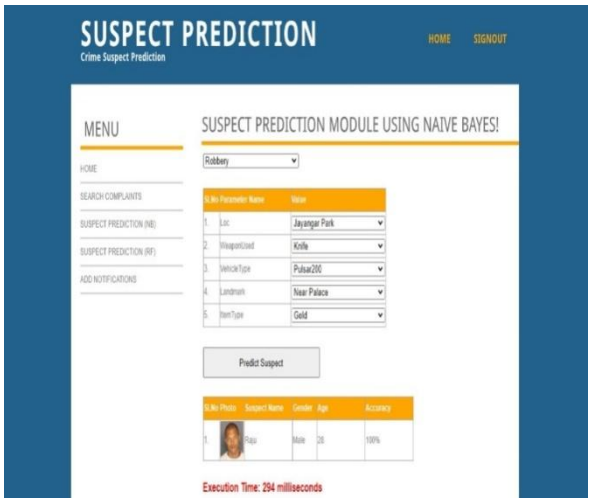


Fig: Execution of Navie Bayes

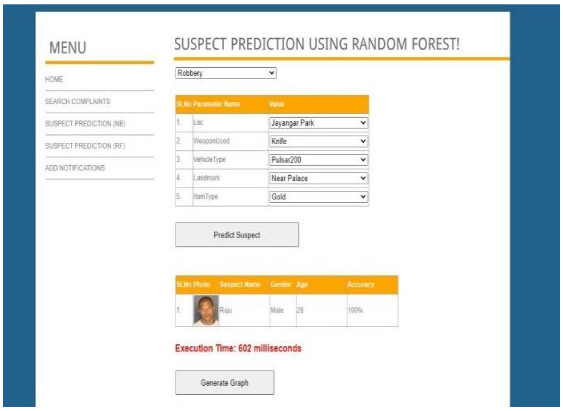


Fig: Execution of Random forest

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 650-654 www.ijaem.net ISSN: 2395-5252

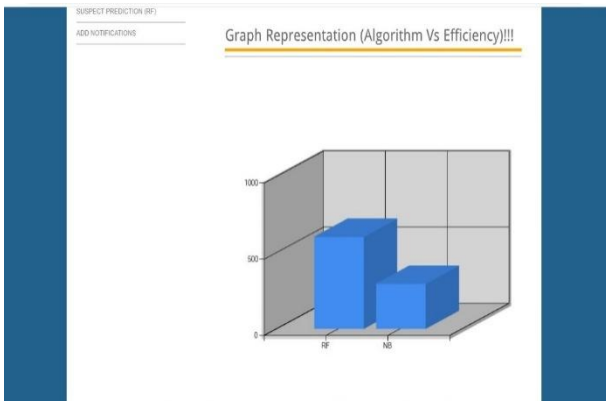


Fig: Graphical representation

IV. CONCLUSION

System is useful to Police Departments. Reduces the number of crimes and Increase efficiency in solving the crimes faster and also can be applied to automatically notify the crimes. Crime prevention and detection become an important trend in crime and a very challenging to solve crimes. Helps speed up the process of solving crime and help the computerized systems detect the criminals automatically. Determines the crime suspects and reduce the crimes and to take the precautionary measures to avoid crimes. Proposed system is applicable in the field of crime. Proposed system is an innovative crime suspect detection system.

V. FUTURE ENHANCEMENTS

Instead of registering complaints through a piece of pen and paper online complaint registration can be started. By doing so, the complaints will not be lost and can be kept safe. Also another module called the query module can be added where in the public can post queries to the administrator regarding any public issues.

ACKNOWLEDGEMENTS

We are indeed grateful to many groups of people who have helped us with various aspects of this study. We would like to thank Prof Sharath HA as well as special thanks for our head of the department DrSharath Kumar Y H for giving us the opportunity to do this wonderful project on “Crime suspect prediction using machine learning”.

REFERENCES:

[1]. Jay Patel, HareshWala, Deepak Shahu, Hazal Lopes, Computer engineering, Universal College Of Engineering Mumbai, India, jaypatel2799.jp@gmail.com,IEEE.

[2]. T. Bosse, C. Gerritsen and J. Treur, “On the relation between cognitive and biological modelling of criminal behavior”, Computers in Human Behavior, vol. 27, no. 5, pp. 15931611, 2011.

[3]. L. Mookiah, W. Eberle, A. Siraj: Survey of Crime Analysis and Prediction, *The 28th International Flairs Conference*, pp. 440–443 (2015).

[4]. RasoulKiani, Siamakmahdavi, Amin Keshavarzi, “Analysis and Prediction of Crimes by Clustering and Classification”, International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No. 8, pp 11-17, 2015.

[5]. R. Clarke and J.E. Eck,” Become a Problem-Solving Crime Analyst”, Routledge, 2014.

[6]. Abu-Hana, R. O., Freitas, C. O., Oliveira, L. S., &Bortolozzi, F. (2008). Crime scene classification.

[7]. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 419–423).

[8]. Hussain, Durairaj et al., “Criminal behavior analysis by using data mining techniques”, International Conference on Advances in Engineering, Science and Management (ICAESM), IEEE, March 2012. Keyvanpour, Javideh, et al., “Detecting and investigating crime by means of data mining: a general crime matching framework”, World Conference on Information Technology, Elsvier B.V., 2010.

[9]. Peng Chen, Justin Kurland, “Time, Place, and Modus Operandi:A Sample Apriori Algorithm Experiment for Crime Pattern Detection”, IEEE International Conference, 2016.

[10]. Federico Neri, Paolo Geraci “Online Police Station a state-of-the-art Italian Semantic Technology against cybercrime”, Advances in Social Network Analysis and Mining, 2009.

[11]. T.Abraham and de Vel, O., “Investigative profiling with computer forensic log data and association rules”, Data Mining, Proceedings, IEEE International Conference, 2002.

[12]. N. HanmantRenushe, R. PrasannaRasal& S. Abhijit Desai, “Data mining practices for effective investigation of crime”, IJCTA, May-June 2012.

[13]. Ravi Kumar &BharathiNagpal, “Analysis and prediction of crime patterns using big data”, Springer, 5 july 2017.

[14]. Masila Abdul Jalil, Chai Pui Ling, Noor Maizura Mohamad Noor, FaithahMohd,

"Knowledge Representation Model for Crime Analysis", 2nd Internation Conference on Computer Science and Computational Intelligence 2017, ScienceDirect, 2017.

[15]. Gnaneswara Rao Nitta, B.Yogeshwara Rao, T.Sravani, M.BalaAnand, N.Ramakrishna, "LASSO-based feature selection and naïve Bayes classifier for crime prediction and its type", Springer, 7 October 2018.

[16]. Teng Li, "Criminal Behavior Analysis Method based on Data Mining Technology", IEEE International Conference, 2016.

[17]. Mugha Sharma, "Z - CRIME: A Data Mining Tool for the Detection of Suspicio Criminal Activities Based on Decision Tree", IEEE, 2014.

[18]. UbonThongsatapornwatana, Akkarach Kawbunjun, WoraphonLialkiatsakum, Tossapon1Boongoen, "Analysis of Criminal Behaviors for Suspect Vehicle Detection", IEEE, September 2017.

[19]. KamalTaha,PaulD.yoo, "A Forensic System for Identifying the Suspects of a Crime with No Solid Material Evidence", IEEE Computer Society, 2018.

[20]. MaryShermila A, Amrith Basil Bellarmine, Nirmala Santiago, "Crime Data Analysis and

[21]. Sunil Yadav, Meet Timbadia, AjitYdav, RohitVihwakarama&Nikhilesh Yadav, "Crime Pattern Detection, Analysis & Prediction".

Sharath H A, et. al. "Crime Suspect Prediction Using Machine Learning". *International Journal of Advances in Engineering and Management (IJAEM)*, 2(1), 2020, pp. 650-654.

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 02 Issue: 01 DOI: 10.35629/5252

www.ijaem.net
Email id: ijaem.paper@gmail.com