"Bodh", BPIT's International Journal of Technology & Management
ISSN: 2454-8421, Volume 2, January 2017, Page 33-37

Crime Cases and Ways of Criminal Detection

Dr. Bhawana Suri [1], Mrs. Deepti Jain[2], Alakshender Kumar [3], Abharika Kamra [4], Rajesh [5], Shweta Singh [6]
Dept. of Computer Science Engineering
Bhagwan Parshuram Institute of Technology, Rohini
New Delhi.
{Bhawnasuri12@gmail.com [1], deepti.kheterpal@gmail.com [2], alakshender@outlook.com [3], abharika2@gmail.com [4], rajeshkaushik3108@gmail.com [5], shwetasingh1996j@gmail.com [6]}

***Abstract*—** In today's scenario crime rates are rising tremendously and hence we have put forward the requirement of a computer technique that helps in criminal detection and prevention. The frequent and repeated robbery, rape, murder, dowry etc. have made common citizens to have sleepless nights, the criminals may have been operated in an organized way and sometimes have nationwide or international links. The research paper of quantities instead of quality would be the best way to review the high crime rates because you are open to farther information that assist you to relate to any new crime detection The paper will deal with the same. The data is based on a record of 130 criminal cases rather than dummy data set which would otherwise lead to a wishful analysis. It is a known fact that with the increase in the criminals we have a shortage in the police force and need a computer technique that helps in criminal detection and prevention in the future crimes by the criminals. The paper will include juvenile and non-juvenile criminal cases i.e. all age groups. Graphical approach has been used in the paper plotting age vs. education. Following that, knives'-Nearest Neighbor Algorithm (k-NN) is used to find the nearest plot. Further, we use the Naïve Bays algorithm to detect the future rating for each attribute and detect the further chances for the crime to take place.

Keywords—k-Nearest Neighbor Algorithm, Naïve Bays, Memory Based Reasoning.

I. INTRODUCTION

The population growth is a global scenario, and with this growth there is a larger competition for the gain of resources and other problems like unemployment proper education, nutrition etc. to each person. This directly influences over the criminal activities. With growing crime rates we need a mechanism that could detect criminals and hence lead to prevention. In this modern lifestyle there is a need to automates the criminal identification and techniques, so that we could decrease the human work load and we could work our minds in doing something more productive and use time more efficiently. The paper uses data mining algorithms for identification, classification and future scope of crime by any criminal. This will help us develop a code through which criminal identification will be easy. The paper uses data mining algorithms for identification, classification and future scope of crime by any criminal. This will help us develop a code through which criminal identification will be easy. The algorithms used in the paper are the Apriori, *k*-NN and Naive Bayes.

II. ALGORITHM DESCRIPTION

The paper uses two data mining algorithms for the detection and prevention of future crimes described below:-

k-NN is an instance-based algorithm for learning, where any local plotted point is approximated locally depending upon the nearest point to that point i.e. all computation is deferred and classification depending on the nearest point. The *k*-NN algorithm is among the simplest and one of the most used of all machine learning algorithms. The training examples includes vectors in a multidimensional feature space, each of which having a class label. In the training of the algorithm we store the vectors with the class labels of the samples. In the

classification phase, we define k as a user-defined constant with an unlabelled vector i.e. a query or test point is classified by assigning the label that is the most frequent in the k samples used nearest to the query point.

$$\text{Distance}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \quad (1)$$

Other distances that may be used are 1) Manhattan Distance (x= (a, b) and y=(c, d) then Manhattan distance between x and y is |a−c|+|b−d|) and 2) Minkowski Distance is

$$d(p, q) = \left(\sum_{i=1}^{n} |p_i - q_i|^c\right)^{1/c} \quad (2)$$

The Naive Bayes algorithm is a prediction and classification Algorithm. It uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data. Naïve bays algorithm is used in data mining process. Data mining is a process of analysing patterns from historical information and transform it into an understandable structure for future use. Typical use of data mining process is in Science fields where analysts identifies the patterns based on historical data available and use those patterns to predict future activities. It is also used in medical fields, like whether a patient has heart disease or not from his historical data like patient's age, blood sugar level and other symptoms.

(A) Description:

Naïve bays algorithm is based on three concepts:
Prior: Past experience
Likelihood: chance of event could happen. **Posterior**: predicting the event will occur

Prob. (B given A) = Prior * Prob. (A and B)/Prob. (A) (3)

Example: Support you would like to determine the possibility that people over 60 ages are more prone to heart disease. In this case, prior condition (A) would be over 60 and dependent condition (B) would be having heart disease.

If there are 100 persons randomly tested for heart disease and before testing it is already known that out of them 25 are having heart disease,

Probability of A and B, (means people are tested and have heart disease previously) = 25%
If 75 of the 100 patients are over 60, then Probability of (A) = 75%
Then in this case, Bayes Theorem would predict that that 33% of the patients over 60 are likely prone to heart disease (25/75).

III. WORKING AND IMPLEMENTAION
JJJ.

The implementation of the paper includes a java code based over the k-NN algorithm, implementing memory based reasoning (MBR), firstly we need to put the attributes of the criminals on the basis of the criminal records that we have.

The attributes includes the age, punishment, education, gender of the criminal, gender of the victim. The data entered is flexible and later be changed or expanded as more the data, more better we could predict the crime. As the program runs we enter the attributes of the person who expected to be involved in any crime and depending on the prior data we could detect the crime done by that person.
For the use of the system in any other country one may need to change the prior data totally as the data is on the basis of the judicial system of India, which is different with any other country. In order to explain the working of the program we need to consider a five dimensional graph on which we plot each attribute that we have

considered i.e. age, punishment, education, gender of the criminal, gender of the victim. Now using the k-NN algorithm we find the nearest neighbor of the unknown point using the Euclidean distance formula that we have modified as:-

$$\text{Distance}((v, w, x, y, z), (a, b, c, d, e)) = \sqrt{(v - a)^2 + (w - b)^2 + (x - c)^2 + (y - d)^2 + (z - e)^2} \quad (4)$$

A five dimensional graph won't be possible thus a simple representation on a two dimensional graph is as follows:-

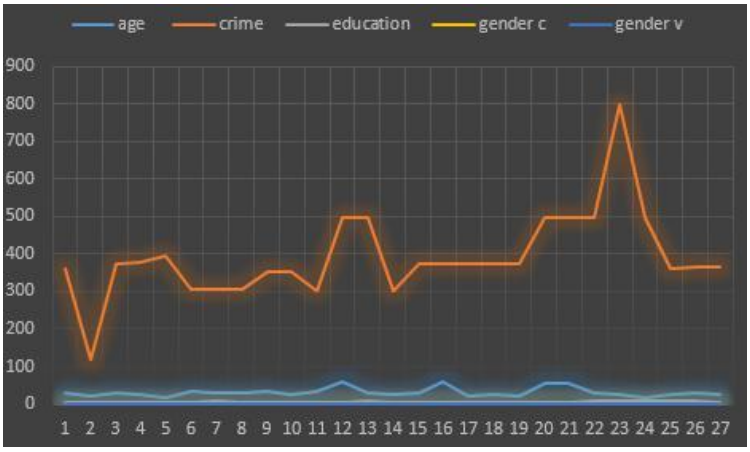

Figure 1: 2D Representation of attributes of criminal

The technique allows us to determine the relation between the crime, criminal and the punishment given to the criminal, depending on the prior decisions taken by the court. Due to the privacy in the judicial system of India we are able to add only five attributes in the program that can be increased if needed in the future by making small changes and the research is restricted to only one hundred and thirty cases. The cases used in the paper are real and the information regarding them is obtained from different courts and police stations, spending lot of effort and time.

Example Result:

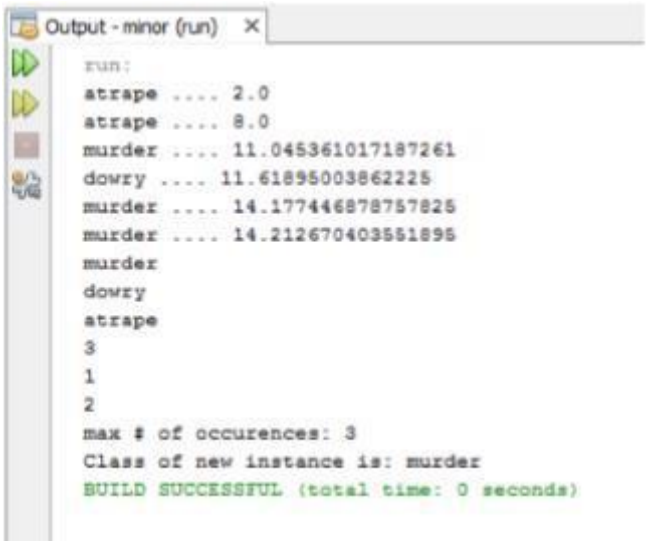

Figure 2: Result of murder criminal from past records

IV. MATH

Equations of Naïve Bayes algorithm:

Prior: past experiences (p)

Likelihood: chance of event could happen (l) Posterior: predicting the event will occur (P) Evidence: number of cases where event occur alone (e)

$$P = (p*l)/e \quad (5)$$

Euclidean distance formula:

$$\text{Distance}((v, w, x, y, z), (a, b, c, d, e)) = \sqrt{(v - a)^2 + (w - b)^2 + (x - c)^2 + (y - d)^2 + (z - e)^2} \quad (6)$$

C. THE CONCEPT

We have used association rule to map criminals and the crimes.

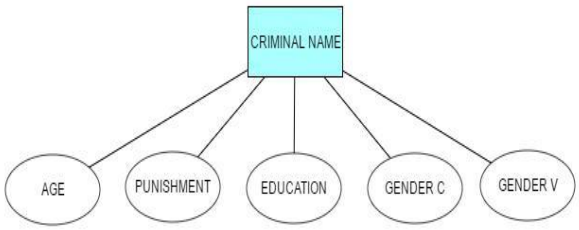

Figure 3: criminal and attributes

For any criminal we divide its attributes and rate them, plotting it over the graph and using the k-NN algorithm we can get the nearest plot, and correspond it to the criminal, whose plot is nearest to it.

Example a new criminal's attributes are near to that of a murderer, then we could conclude that the new criminal, (according to the logic of paper) is most probably a murderer.

The paper also deals with any future detection of crime by the criminal i.e. use Naïve Bayes algorithm. According to the rating of Prior experiences, Likelihood and Evidence it provides the Posterior rating of occurring of the crime.

Specifically here we have 5 attributes i.e. 1) age of criminal, 2) punishment that he received 3) education 4) gender of criminal 5) gender of victim. We have already used the data and using that we will decide the new attribute to which category of crime does the crime belongs, depending on the information provided, more information in the database will make the implementation results more nearer to the actual result. Presently we deal with crimes like rape, murder, robbery, dowry demand etc. covering more crimes can make us cover more no. of crime and could have a usable

implementation of the research.

VI. CONCLUSION

Automatize the crime detection of a criminal and prediction of crime by the criminal. Practically the system may fail, if:

D. The data set provided is not enough.
E. The criminal may vary in terms of its attributes from the previous listed criminals.
F. Records need to be updated.
G. The attributes included may not be enough.
H. The same data not valid in any other country.
I. System gives output on the basis of past experiences only (MBR), not to be relied on it as the results may vary

REFERENCES

[1]Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep hosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S., Yu · Zhi-Hua Zhou · Michael Steinbach, David J. Hand, Dan Steinberg "Top 10 algorithms in data mining" Knowl Inf Syst (2008).

[2] Qusay Bsoul, Juhana Salim, Lailatul Qadri Zakaria "An Intelligent Document Clustering Approach to Detect Crime Patterns" The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013).

[3] MohammadReza Keyvanpoura, Mostafa Javidehb, Mohammad Reza Ebrahimia "Detecting and investigating crime by means of data mining: a general crime matching framework" WCIT 2010.

[4] W. Buikhuisen, E. H. M. Bontekoe, C. v.d. Plas-Korenhoff and S. van Bum-en "Characteristics of Criminals: The Privileged Offender" lnlernational Journal 01 Law and Psychfarry 1984.

[5] F.I.R copies across North India.

[6] Science Direct http://www.sciencedirect.com/

[7] Wikipedia http://www.wikipedia.com/