International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Crime Prediction System

Pankaj Kushwaha[1], Kiran Kunwar[2], Archna Verma[3], Simili Shetty[4], Sampada Watane[5]

[1,2,3,4]*Student* [5]*Assistant Professor*

[1,2,3,4,5]*Department of Computer Engineering, Shree L. R. Tiwari college of Engineering, Mumbai University, India*

Abstract - *Nowadays crime rates has touched the peak level, so to cease the rising level of committing crimes or to find the suspect easily our system will be very helpful to the investigators. This paper introduces a solution to the criminal and crime prediction problem using naïve bayes theory. Acquiring the crime dataset for the attributes which we required for our proposed system is quite difficult process in practice due to confidentiality principle. So we created the data considering the attributes which we want like criminal name, location, weapons used, date of crime, type of crime. The result of the procedure is the maximum probability of the suspect or the crime type.*

Key Words: **Data Mining, Naïve Bayes, Crime dataset, Crime and Criminal Prediction.**

1. INTRODUCTION

The crime incidents area unit continued to grow in rate and complexness. Crime can not be foreseen since it's neither systematic nor random. In many criminal investigation, crime analysis includes processes to find the criminal of incidents. The discovery of possible suspects in advance is an important requirement for security forces to optimize the usage of human and technical resources, and carries out important functions in the prediction of crime. Crime Records Bureau crimes like burglary, arson etc. have been decreased while crimes like murder, sex abuse, gang rape etc. have been increased. Even though we cannot predict the result with 100% accuracy but the application gives the result having more chances of attempting it.

2. LITERATURE REVIEW

There are number of papers which we have analyzed in order to determine the technology used and data mining techniques. The review of literature will involve efficient and usable techniques such as fuzzy system and weka tool.

[1] In the studied paper, tool for forecasting the crime was built. The discrete choice model will be used in this case. The discrete choice model takes into account the choice of the criminal and location in order to forecast the crime.

[2] In this paper a specific location Malaysia is considered. The crime location and mindset of the criminal will be considered in this case. Crime Forecasting is rarely used globally by police including Malaysia. In practice usually the police would target persons with their criminality and studying their strategy of implementing crime. The police will also monitor the current crime situation and will take necessary action when the crime index increases. Both of these scenarios require action taken after crime incurred. Therefore if crime forecasting can be adopted perhaps early crime prevention can be enforced. The aim of this study is to identify crime patterns in Kedah using univariate forecasting technique. Seventy-six recent monthly data (January 2006 – April 2012) were obtained from IPK Alor Star with the permission from PDRM Bukit Aman. Exploratory Data Analysis (EDA) and adjusted decomposition technique were conducted in order to fulfill the objectives of the study. The findings revealed that total crimes in Kedah were mainly contributed by type of property crime (80-85%) while violent crime has a small proportion only. Fortunately due to the productiveness of the police the property crime trend indicated curve declining pattern.

[3] Special section of the crime forecasting is considered in this case. The crime forecasting is used so that crime can be controlled. No specific location is considered in this case. The GIS(Geographical Information System) is used in order to predict the location where crime is occurred.

[4] This is book in which crime analysis is conducted. The mentality of criminal is considered in this case. The help of crime forecasting is also considered in this case. There are several works that analyze the social network for the crime analysis, but to the best of our knowledge, there is no study analyzing the social network of the criminals using naïve bayes. Although SNA [5-6] is a widely used tool in criminal investigation, the automatic criminal analysis over such a network is very limited again due to data accessing difficulties. In the literature, there exists a work over the terrorist network and other suspects [7,8,9]. However, these works only consider SNA analysis without analyzing the other crime related data such as location, date and type of crime.

3. PROPOSED SYSTEM

The steps included in prediction are:

1. Data collection
2. Comparison of historical data and collected data.
3. Prediction using naïve bayes algorithm.
4. Database updated.

Firstly, if there has occurred a crime at a particular location, the investigators would collect all the evidences which comes under data collection. Now it'll compare the evidences with the historical data that whether the same type of case has occurred in past with the same set of attributes by using Naïve bayes Algorithm. Naive Bayes doesn't always provide the accurate result instead it give the maximum probability of the particular attribute.

Fig. 1 shows the actual work flow of the system.

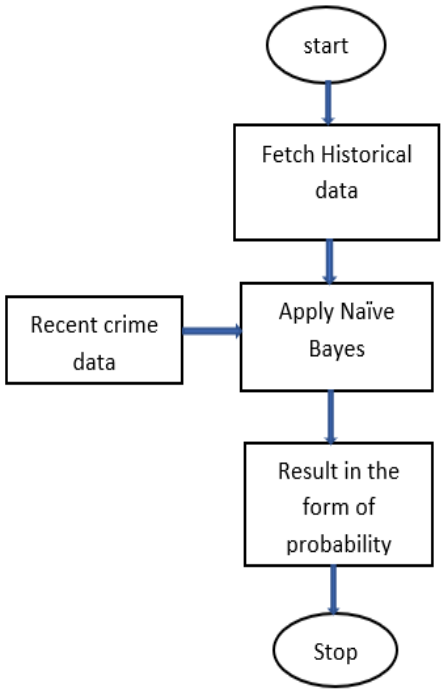

Fig. 1 Flowchart of Proposed system

4. METHODOLOGY USED

4.1 DATA GENERATION

The expected dataset shall look something of this kind on the basis of which we shall perform prediction.

name	crime location	gender	crime type	evidence and weapon us	crime reason
pankaj	mira road	male	murder	knife	financial matter
rohit	vasai	male	half murder	gun	dispute
kiran	bandra	female	robbery	knife	financial matter
jatin	virar	male	kidnapping	gun	family matter
simili	bhayandar	female	drug trafficking	drug	financial matter
pankaj	mira road	male	rape	gun	dispute
rohit	dadar	male	half murder	knife	politics
archna	dadar	female	robbery	chloroform	financial matter
simili	bhayandar	female	drug trafficking	drug	financial matter
karan	dahisar	male	kidnapping	chloroform	affair
rahul	borivali	male	rape	gun	politics
pravin	thane	male	murder	knife	family matter
hena	mira road	female	suicide	poison	affair
amit	dahisar	male	rape	drug	affair
pranjal	borivali	male	suicide	knife	family matter
kiran	bandra	female	drug trafficking	gun	politics
prachi	andheri	female	drug trafficking	drug	financial matter
pawan	mira road	male	rape	marks	affair
rohit	vasai	male	murder	knife	dispute

Fig. 2 Dataset for Crime Prediction System

4.2 NAÏVE BAYES ALGORITHM

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where A and B are events.

- Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.
- P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance (here, it is event B).
- P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.

Algorithm 1 Pseudocode

1. Given training data set D which consists of documents belonging to different class say class A and B.
2. Calculate the prior probability of class A=number of objects of class A / total number of objects
   Calculate the prior probability of class B=number of objects of class B / total number of objects
3. Find ni, the total number of word frequency of each class.
   na= the total number of word frequency of class A.
   nb= the total number of word frequency of class B.
4. Find conditional probability of keyword occurrence given a class.
   P(word1 / class A) = wordcount / ni(A)
   P(word1 / class B) =wordcount / ni(B)
   P(word2 / class A) = wordcount / ni(A)
   P(word2 / class B) =wordcount / ni(B)
   .....
   .....
   P(wordn / class B) =wordcount / ni(B)
5. Avoid zero frequency problems by applying uniform distribution.
6. Classify a new document C based on the probability P(C /W).
   a) Find P(A / W) = P(A) * P(word1/ class A) * P(word2/ class A) ……* P(wordn / class A).
   b) Find P(B / W) = P(B) * P(word1 / class B) * P(word2/ class B) ……* P(wordn / class B).
7. Assign document to class that has higher probability.

Fig. 3 Pseudo code for Naïve Bayes

5. IMPLEMENTATION

5.1 LOGIN PAGE

User will first have to login to access the system, if he/she already has an account. If not then will have to signup for the same.

Fig. 4 is the screenshot of the login page.

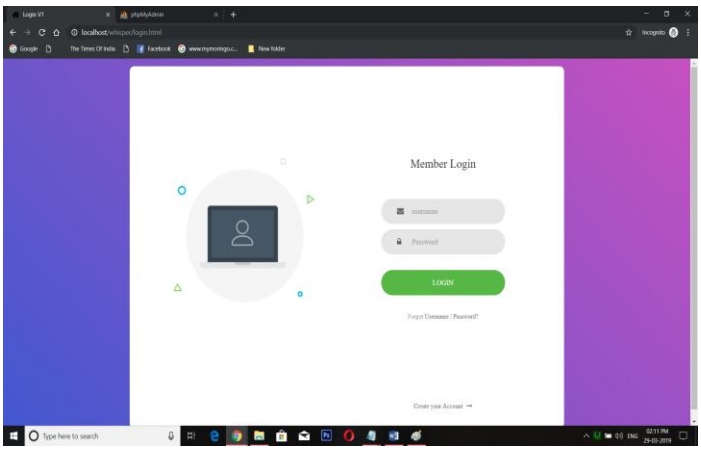

Fig. 4 Login page

5.2 SIGNUP PAGE

User has to provide a valid e-mail id, and choose username and password to signup.

Fig. 5 is the screenshot of the signup page.

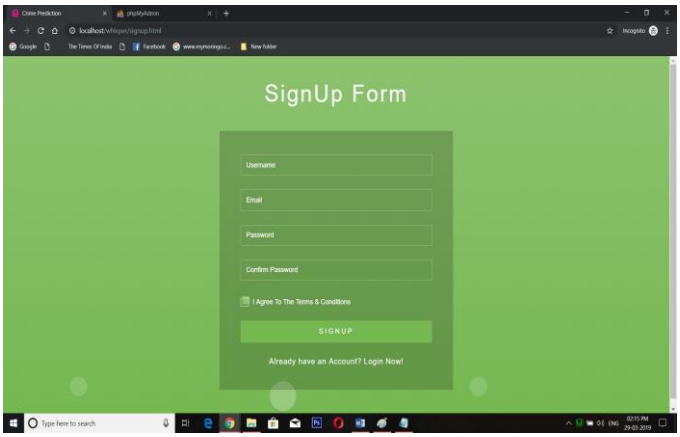

Fig. 5 Signup page

5.3 HOME PAGE

After a valid user gets logged-in, the page redirects to the home page. It is an interaction page where user can choose what task to be performed from the tabs.

Fig. 6 is the screenshot of the home page.

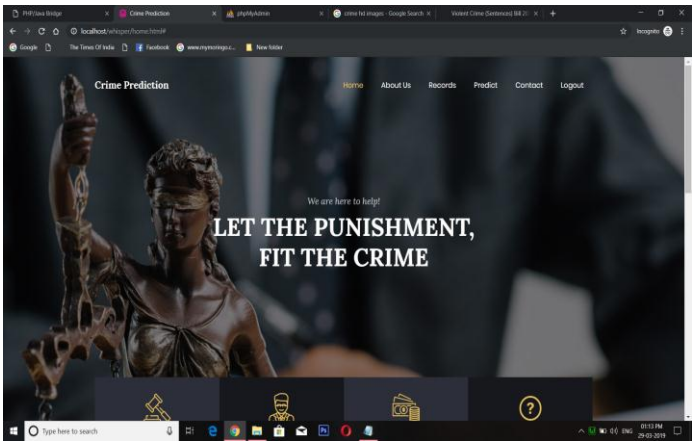

Fig. 6 Home page

5.4 ABOUT PAGE

About page tells us the information about the system.

Fig. 7 is the screenshot of the about us page.

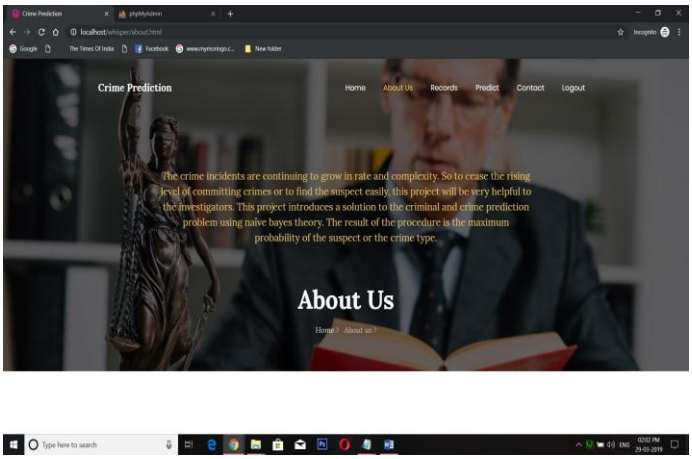

Fig. 7 About Us page

5.5 RECORDS PAGE

Here user can check the previous dataset by just downloading the file from the download option on the screen.

Fig. 8 is the screenshot of the Records page.

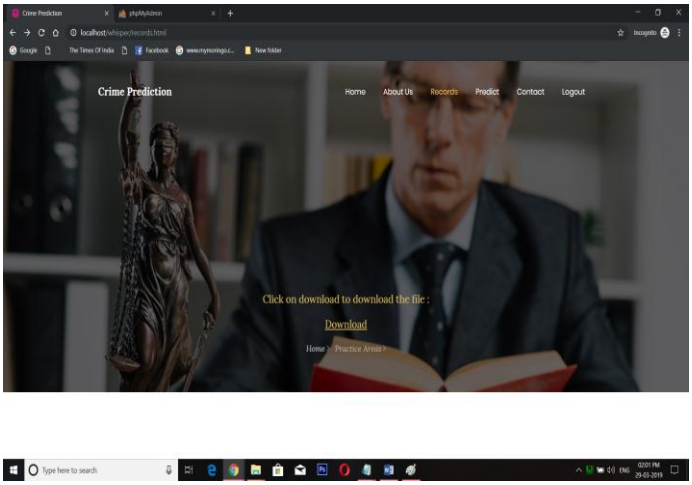

Fig. 8 Records page

5.6 PREDICT PAGE

For prediction, user has to enter the required details viz. location, gender, crime-type, weapon used and reason of the crime that has currently occurred.

For eg. The details entered here are mira road, male, murder, knife, affair.

Fig. 9 is the screenshot of the predict page.

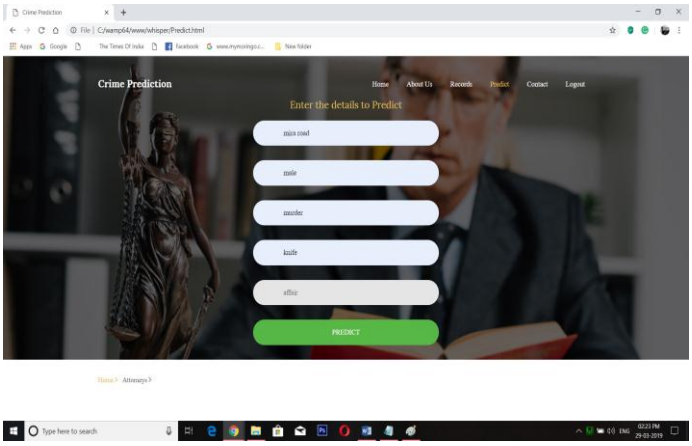

Fig. 9 Predict page

5.7 OUTPUT PAGE

After clicking on the predict button, it re-directs to the output page. Output page shows the prediction of the criminal which has the highest probability.

The details entered on the predict page gets compared with the historical dataset and using naïve bayes, it predicts the criminal. Here the criminal is jatin.

Fig. 10 is the screenshot of the output page.

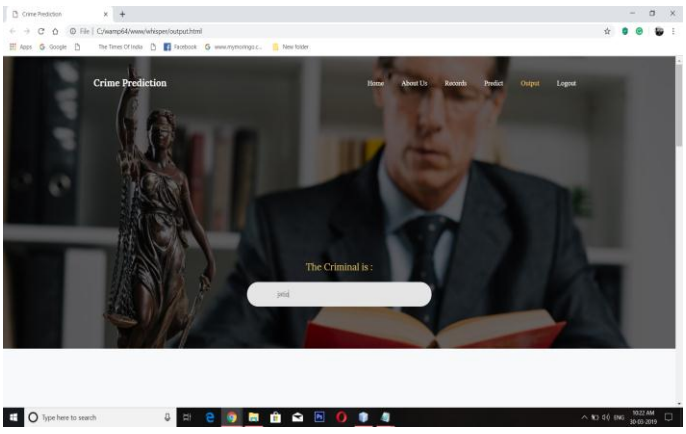

Fig. 10 Output page

6. CONCLUSION

People committing crime must be stopped or reduced as large number of people are getting affected due to this. Our system will help the agencies to tackle the problem of identifying the suspect and the crime that has occurred. Large amount of struggle is being done manually plus digitally for finding the actual criminal. But not every time there is success, so still some work can be done to improve the forecasting system.

REFERENCES

[1] M. B. Mitchell, D. E. Brown, and J. H. Conklin, "A Crime Forecasting Tool for the Web-Based Crime Analysis Toolkit," 2007 IEEE Syst. Inf. Eng. Des. Symp.,pp.1– 5,2007.

[2] S.Ismail and N. Ramli,"Short-Term Crime Forecast ting InKeedah," Procedia-Soc. Behav. Sci., vol.91,pp.654–660,2013.

[3] W. Gorr and R. Harries, "Introduction to crime forecasting," vol. 19, pp. 551–555, 2003.

[4] K. H. Vellani, "Crime Analysis: for Problem Solving Security Professionals in 25 Small Steps," Facilities,pp.1–56,2010.

[5] Kenneth M, Stefan W, John M (2001) Knowledge extraction from local function networks. In: International joint conference on artificial intelligence. Lawrence Erlbaum Associates Ltd, pp 765–770

[6] Calvo-Armengol A, Zenou Y (2006) Social networks and crime decisions: the role of social structure in facilitating delinquent behavior. Int Econ Rev 45:939–958 52. Xiao L, Fan M (2014) Does social network always promote entrepreneurial intentions? An empirical study in China. Neural Comput Appl 24:21–26

[7] Khan NG, Bhagat VB (2013) Effective data mining approach for crime-terrorpattern detection using clustering algorithm technique. Int J Eng Res Technol 2:2043–2048

[8] Dombroski MJ, Carley KM (2002) Netest: estimating a terrorist network's structure. Comput Math Org Theory 8:235–241

[9] Porter MD, Brown DE (2007) Detecting local regions of change in high-dimensional criminal or terrorist point processes. Comput Stat Data Anal 51:2753–2768

[10] Datasets, Center for Machine learning and Intelligent Systems, Donald Bren School of Information and Computer Sciences, University of California, Irvine.Available from: http://cml.ics.uci.edu/. 25/05/2015.

[11] References from www.google.com.