

CRIME PREDICTION SYSTEM USING BAYESIAN PREDICTION ALGORITHM

¹Sunil Yadav,²Mohit Sharma,³Yash Redkar,⁴Aditya Sawant,⁵Varun Achary

¹Professor(Department of Information Technology), ²Student, ³Student, ⁴Student, ⁵Student
Department of Information Technology, University of Mumbai, Shree L.R. Tiwari College of Engineering,
Thane, India

Abstract : We are now living in a materialistic society where human values are dying. Everyone wants to live in palatial homes, drive luxury cars and use the latest gadgets. When people can't earn enough money to satisfy their cravings for material goodies, they get into crime. Crimes are a social hindrance and cost severely to our societies. Studies reveal that only 10% of the criminals commit 50% of the crimes [1]. This information is very beneficial and can be used to trace criminals using Naïve-Bayesian and other algorithms. An important component of our project is 1) Creating an online portal for the registration of online FIRS which is not present in CIPA(Common Integrated Police Application). 2) Predicting the most probable criminals on the basis of clusters created in order to help the police with their investigation

Index Terms - Crime Prediction System, Bayesian Prediction

I. INTRODUCTION

The crime rates accelerate continuously and the crime patterns are constantly changing [2]. It is neither systematic nor entirely random [3]. In the existing system (i.e. CIPA(Common Integrated Police Application) only we can see a part of information related to the criminal in our state, the existing system has more workload for the legitimate person, but in the case of our proposed system, the authorised user can register in the application and send the crime report and complaint about a particular social crime. The proposed system allows a user to login through the registered username and password and file complaints. The proposed system fast tracks the existing processes by a huge margin, while also providing security, obscurity, better service via a user- friendly medium. After getting with registered complaints, the investigation process will take place. At the end of investigation, the previously processed government criminal records will be taken as referential training sets and witness record will be taken as a pattern. BAYES Theorem will be applied to find the accused involved in the crime/suspect. When implemented at larger scale, the system could increase the speed of the judicial process and make way for lesser crimes. The main module is the Administrator.

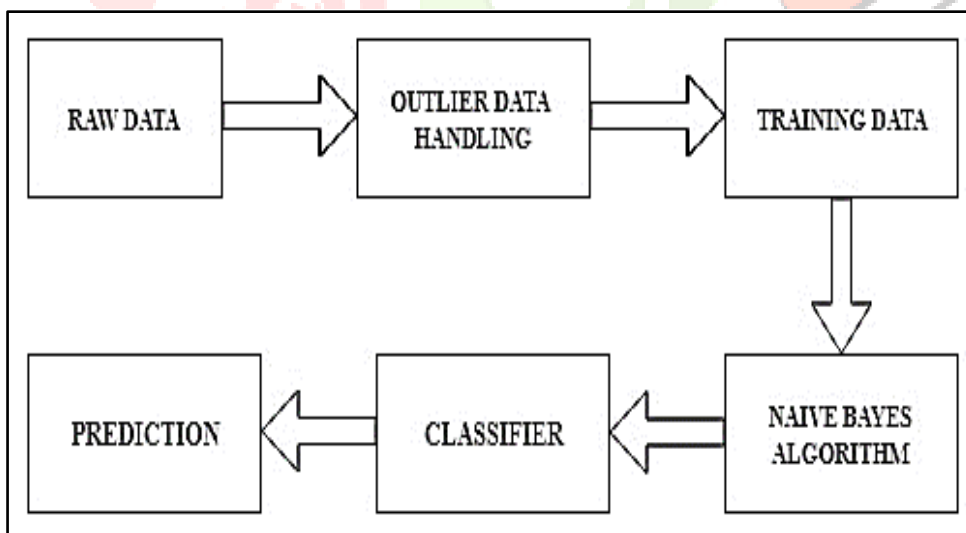


Fig 1. Block Diagram of the mining process

II. LITERATURE SURVEY

The use of clustering algorithm for a data mining approach to help detect the crimes patterns and speed up the process of solving crime was studied. Techniques such as K-mean algorithm was implemented to to aid in the process of identification of crime patterns. semi-supervised learning technique was also used here for knowledge discovery from the crime records and to

help increase the accuracy of the prediction. Also a weighting scheme for attributes to deal with limitations of various out of the box clustering tools and techniques was developed [8].

The review of the literatures on various data mining applications, especially applications that applied to solve the crimes was done. Also the research gaps and challenges of crime data mining was addressed. In addition to that, this paper provides insight about the data mining for finding the patterns and trends in crime to be used appropriately and to be a help for beginners in the research of crime data mining.[7]

A number of data mining classification techniques were employed to perform the crime forecasting. An analysis of variety of classification methods helped to determine which is best for predicting crime "hotspots". A proposal for the best forecasting approach to achieve the most stable outcomes was done. The model makes use of implicit and explicit spatial and temporal data to make reliable crime predictions.[4]

The identification of the data sources that should be used for knowledge discovery in crime data mining was done. Similarities and conflicts between crime data mining methodologies show that some forms of knowledge discovery are applicable for particular crime data mining methodologies. It is suggested that selecting the appropriate methodology depends on whether general or specific tasks required or high volume of crime data to be prepared. [5]

The proposed system could predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the resurgence of computerized systems, crime data analysts can help the police to accelerate 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 design 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.[12]

III. PROPOSED METHODOLOGY

Classification is one of the classic data mining techniques, which is used to classify each item in a set of data into one of the predefined set of classes or groups [4]. With the help of classification, existing datasets can be easily understood and it also helps to predict how new dataset will behave based on classification criteria [5].

The naive Bayes classifier is proposed for the machine learning task of classification, which has been found to give good performance. Naive Bayes and other standard algorithms are applied to the database application, in which the variables are highly non-normal and found that the algorithm performed well while predicting a class that had been derived from the same data.

The Bayes Theorem is as follows:

$$P(A/B) = [P(B/A) * P(A)] \div P(B) \quad (1)$$

P(A) -> Probability of the class which is going to be predicted.

P(B) -> Probability of the given pattern which predicts the class.[17]

For Bayes implementation, we need a Training set and Pattern. In the training set, we have to find the overall probability of the class that is going to be predicted. After finding the probability of the class the probability for all the given attributes in the training set is found. Here the Sample Space is taken to be 20.

Table 1. Training Set for Bayes Theorem

Sr No.	Mode of Transport	Weapon used	Crime type	Timing of the crime	No of persons	Motivated?	Age group	Criminal name
1	2-Wheeler	Knife	Chain	Evening	Gang	Yes	Middle	X
2	2-Wheeler	Knife	Assault	Morning	Gang	Yes	Young	Y
3	4-Wheeler	Rod	Murder	Night	Single	Yes	Young	X
4	2-Wheeler	Rod	Robbery	Night	Gang	Yes	Old	Y
5	2-Wheeler	Gun	Robbery	Night	Gang	No	Old	X
6	4-Wheeler	Knife	Robbery	Night	Single	No	Middle	X

7	Walk	Bat	Chain	Morning	Gang	No	Middle	X
8	Walk	Rod	Assault	Evening	Gang	No	Middle	Y
9	Walk	Gun	Murder	Evening	Single	No	Young	Z
10	2-Wheeler	Knife	Murder	Evening	Gang	No	Young	Z
11	2-Wheeler	Gun	Murder	Night	Gang	Yes	Young	X
12	2-Wheeler	Gun	Murder	Night	Gang	Yes	Old	Y
13	4-Wheeler	Rod	Murder	Morning	Single	Yes	Middle	Z
14	2-Wheeler	Bat	Assault	Evening	Gang	Yes	Middle	Y
15	Walk	Knife	Assault	Morning	Gang	Yes	Middle	X
16	Walk	Knife	Robbery	Morning	Gang	Yes	Middle	Y
17	2-Wheeler	Bat	Chain	Morning	Gang	No	Middle	X
18	Walk	Rod	Chain	Evening	Gang	No	Middle	Y
19	4-Wheeler	Gun	Assault	Night	Single	No	Young	X
20	Walk	Knife	Robbery	Evening	Single	Yes	Young	Y

Here the values entered in the Training set for the algorithm, as shown in the Table 1, are simulated government records. The sample size is 20 and constructed by taking into account the crimes committed by three serial offenders, namely :X,Y and Z .

The probabilities of the criminals being committing the crimes are as follows:

1. $X = P(c1) \rightarrow 9 / 20 = 0.45$
2. $Y = P(c2) \rightarrow 8 / 20 = 0.40$
3. $Z = P(c3) \rightarrow 3 / 20 = 0.15$

After entering the above details, the pattern will be created which then be evaluated with the training set (simulated government records) and starts its “prediction” i.e. the accused involved in the crime will be traced through the created pattern. Here the major seven attributes that influences the crime,

- Transport
- Weapon
- Crime Type
- Timing
- No of persons
- Motive
- Age

Table 2. Probabilistic Values of all Attributes

Attributes	Sub-Attributes	P(C1)	P(C2)	P(C3)
Mode of Transport	2-Wheeler	0.33	0.5	0.33
	4-Wheeler	0.33	0.0	0.33
	Walk	0.34	0.5	0.34
Weapon used	Knife	0.33	0.375	0.33
	Bat	0.25	0.125	0.00

	Rod	0.125	0.375	0.34
	Gun	0.22	0.125	0.33
Crime type	Assault	0.22	0.375	0.00
	Murder	0.23	0.125	1.00
	Robbery	0.22	0.375	0.00
	Chain	0.33	0.125	0.00
Timing of the crime	Morning	0.33	0.25	0.33
	Evening	0.11	0.50	0.67
	Night	0.56	0.25	0.00
No of persons	Single	0.33	0.125	0.67
	Gang	0.67	0.875	0.33
Motivated?	Yes	0.44	0.625	0.33
	No	0.56	0.375	0.67
Age group	Young	0.33	0.125	0.67
	Middle	0.56	0.625	0.33
	Old	0.11	0.25	0.00

Let us take the random pattern that is the investigated details about the accused involved in crime.

Pattern set= {2 Wheeler, Rod ,Assault, Night, Single, Non-motive, Young}

Using the Equation(1) we get the values as follows:

$$X = 3.099 \times 10^{-4}$$

$$Y = 1.02968 \times 10^{-4}$$

$$Z = 0.00$$

Since $X > Y$, Z the predicted accused is **X**.

This algorithm is especially useful in cases where the crimes are committed by serial offenders. That is, one criminal is involved in giving rise to multiple types of crimes, irrespective of the location. Hence, in such cases, we can use this method in order to predict the most likely suspect. This could give the police a headstart in the investigation.

IV. APPLICATION ADVANTAGES

This web-based application will be used for primarily providing training to the intelligence officers whoever attempts to identify the criminal. This software application has been computed successfully and was also tested successfully by taking “test cases”. It is user friendly, and has required options, which can be utilized by the administrator to perform the desired operations.

The goals that are achieved by the software are:

- 24*7 availability: Since this App is hosted online,It will be available all the time, excluding the maintenance and upgradation period.
- User Friendly Access: The user can file an FIR while sitting at his home, since it eliminated the need to go to police stations and waiting periods.

- Safety of Data: All the Data will be stored online, with backup being taken at regular intervals, eliminating the chances of data loss significantly
- Optimum Utilization of the available resources.

V. CONCLUSION AND FUTURE SCOPE

The biggest challenge of this whole project was to gather various records and data staging. As a further development into this project, The project would be able to generate crime hot-spots through which the police could increase their patrolling and deployment in these hot-spots. The developed model would definitely reduce the crime rate and occurrence and would help the law enforcement agencies in many ways, that is from apprehending the miscreants to reducing the crimes by carrying out various precautionary measures.

The developed model would also allow us to compare the crime rates in various other parts where this system would be implemented, which would allow the police to build new strategies to fight the crime rate in that particular area.

VI. ACKNOWLEDGMENT

This work is based on data collected as part of the project "Crime Detection and Prediction System" by research students of Shree L.R Tiwari College of Engineering, under the supervision of Prof. Sunil Yadav, Department of Information Technology, Shree L.R Tiwari College of Engineering, Thane, Maharashtra

VII. REFERENCES

- [1] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav, *Crime Pattern Detection, Analysis and Prediction*, ICECA, 2017, 978-1-5090-5686-6.
- [2] G.Patwa, P.Singh, Sunil Yadav, *Predictive Analysis using Association Classifier*, IJRTER, 2016, ISSN : 2455-1457
- [3] Sunil Yadav, Alok Nath Mishra, Swapnil Chowkekar, Rohit Dubeya and Ashish Raj, *Augmented Situation Awareness and Intelligence*, IEEE ASAI, 2018
- [4] Chung- Hsien Yu, Max W. Ward, Melissa Morabito, Wei Ding, *Crime forecasting using data mining technique*, 11th IEEE International Conference on Data Mining workshop, IEEE, 2011, ISBN : 978-0-7695-4409-0/11
- [5] Fatih OZGUL, Claus ATZENBECK, Ahmet CELIK, Zeki ERDEM, *Incorporating data sources and methodologies for crime data mining*, IEEE, 2011, ISBN : 978-1-61284-4577-0085-9/11.
- [6] Megha Sharma, *z-crime : A data mining tool for the sesection of suspicious activity based on decision tree*, IEEE, 2014, ISBN :978-1-4799-4674-7/14.
- [7] Ubon Thansatapornwatana, *A survey of Data Mining Techniques for Analyzing Crime Patterns*, Second Asian Conference on Defense Technology ACDT, IEEE, 2016, ISBN: 978-1-5090-2258-8/16.
- [8] Shyam Varan Nath, *Crime Pattern Detection using Data Mining*, International Conference on Web Intelligence and Intelligent Agent (WI-IAT2006 Workshops), IEEE, 2006, ISBN: 0-7695-2749-3/06
- [9] Rui Yang, Sigurdur Olafsson - "*Classification for predicting offender affiliation with murder victims*", Expert Systems with Applications, Volume 38, Issue 11, October 2011, Pages 13518- 13526.
- [10] Janick V. Frasch, Aleksander Lodwich, Faisal Shafait, Thomas M. Breuel - "*A Bayes-true data generator for evaluation of supervised and unsupervised learning methods*", Pattern Recognition Letters, Volume 32, Issue 11, 1 August 2011, Pages 1523-1531.
- [11] G. Alan Wang, Homa Atabakhsh, Hsinchun Chen - "*A hierarchical Naïve Bayes model for approximate identity matching*", Decision Support Systems, Volume 51, Issue 3, June 2011, Pages 413-423.
- [12] Shiju Sathyadeven, Deven M.S, Surya Gangadharan.S, *Crime Analysis and prediction using Data Mining*, IEEE, 2014

- [13] Emanuele Olivetti, Sriharsha Veeramachaneni, Ewa Nowakowska - "*Bayesian hypothesis testing for pattern discrimination in brain decoding*", Pattern Recognition, Volume 45, Issue 6, June 2012, Pages 2075-2084
- [15] Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining: Concepts and Techniques 3rd ed.*
- [16] Alessio Sancetta – "*Universality of Bayesian predictions*", International Society for Bayesian Analysis, Volume 7, Issue 1, October 2012, Pages 1-36.
- [17] J.Sethuraman , K.R.Sekar , N.Senthil Selvan : *Bayesian Prediction In Police-Crime Investigation Management System*, International Journal of Engineering and Technology (IJET), , Vol 5 No 5 Oct-Nov 2013, Pg 3660 – 3667
- [18] Lawrence M.Rudener & Tahung Liang - "*Automated essay scoring using bayes theorem*" JTLA - The Journal of Technology, Volume1, Number 2 , June 2002, Pages 1966-1994, University of Maryland.
- [19] https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- [20]<http://cipa.gov.in/cipa/index.jsp>

