



Performance Analysis of Naïve Bayes Algorithm on Crime Data using Rapid Miner

Waseem Ahmed
M.Tech Scholar
Department of CSE
Jamia Hamdard, New Delhi

Md. Tabrez Nafis
Assistant Professor
Department of CSE
Jamia Hamdard, New Delhi

Dr. Siddhartha Sankar Biswas
Assistant Professor
Department of CSE
Jamia Hamdard, New Delhi

Abstract: Crime, when someone does any unlawful activity, the intensity level of crime can be from very low to very high. In current society, crime exist everywhere in distinct form, and if we collect all the data related to different crime, that data would be very large in volume which can be managed through data mining techniques. Using various data mining techniques, lot of conclusion can be drawn like rise or fall in particular type of crime, percentage of particular crime, time when crime mostly or less happens, area in which maximum or minimum crime happens etc. In this paper, we use rapid miner data mining tool and naïve Bayes classification algorithm to show the different types of crime. By using crime data on classification algorithm with rapid miner tool we tries to demonstrate that how efficiently naïve Bayes algorithm can manage this data and the accuracy of result. As there is a probability of crime prediction and naïve Bayes algorithm gives result based on probability, thus the help of naïve Bayes classification algorithm the better accuracy of result can be achieved. This research will be very useful in crime control and crime detection but these all depends on the volume and correctness of the data, more data gives the more accurate prediction.

Keywords: Classification, Crime, Data mining, Naïve bayes, Rapid miner

I. INTRODUCTION

Now a day, a wide variety of data viz digital data, news source, digital library is available on different types of media. These data is in very huge quantity, so it falls under big data category and data mining is used to process and analyze such data. This data is of different type and amount of data is quit big, therefore it is difficult to organize, analyze and retrieve it manually [1].

Data mining is not relevant to one type of data rather it concerned with various kinds of information repositories. Data Mining is applied on Database-situated informational indexes and applications, information distribution centres, Relational databases, value-based databases and propelled informational collections and progressed applications, for example, Temporal Databases, Sequence Databases, and Time-Series Databases, Spatial Databases and Spatiotemporal Databases Text Databases, Object-Relational database and Multimedia Databases, Heterogeneous Databases and Legacy Databases, Data Streams and the Internet globally[2].

Crime investigation is characterized as expository procedures which give significant data in respect to crime examples and pattern relationships to help work force in concealment of criminal exercises.

The major purposes of crime analysis and prediction are as following:

- i) Using available criminal information and data about different crime, the crime can be investigated,
- ii) With the help of existing crime data with location, crime classification can be accomplished,
- iii) Help in crime acknowledgment [3].

Crimes affect the organizations and institutions when incident frequently in a region. Therefore it is required to find out the reasons, factors and relations. Investigation department use traditional manual process to analyze the record of crime and related criminals. On the other hand criminal uses advanced technological and sophisticated techniques to commit the crime [4].

Crime is an unpredictable incident. It is not linked with space and time. It is completely depends on human psychology. The range of crime is not limited; initially it started from breaking the rules of the road to the terrorist attacks. Thus large amount of information is available in various forms [5].

There are a lot of algorithms available in data mining. To select best algorithm among them is a challenging task. For selecting a data mining algorithm following criteria considered-

- Effortlessness in comprehension and understanding.
- Knowledge in distinguishing valuable data.
- Expansion of conceivable situations effortlessly.
- High data pick up.
- Simulation results on different datasets [5].

II. LITERATURE REVIEW

AnshulGoel et al.[1] introduced the performance of Naïve Bayes and J48 classification algorithm with the help of WEKA tool on bank dataset for the accuracy and cost analysis of class gender.

R.Kiani et al. in [3] presented a framework for clustering and predicting crimes by generation of training and testing data, they used Genetic Algorithm for optimizing of outlier detection operator parameters using RapidMiner tool.

S. Deshmukh et al. [5] in their paper shown theexaggerated behaviour of three data mining techniques(J48, JRip and Naïve bayes) for crime investigation.

S.Kapoor& A.Kalra [6] research work focused on crime investigation by executing clustering algorithm on crime dataset through rapid miner tool and they represented the framework of murder crime 1990 to 2000.

O. Georgina N. et al. [7] comparative analysis of classification techniques and shown that J48 higher accuracy.

S. SanthoshBaboo and Malathi. A [8] examine work on building up a crime prediction tool for Indian situation utilizing diverse information mining procedures that can help law implementation office to proficiently deal with crime prediction. The proposed device empowers offices to effectively and financially perfect, portray and break down crime data to recognize significant examples and patterns.

III. NAÏVE BAYES CLASSIFIER

To predict the class membership, Naïve Bayes statistical classifier used. Membership probability refers the probability of a given tuple related to a particular class, it is based on independence assumption it means that existence or not existence of a specific feature is independent of existence or not existence of any other feature. To conclude a suspect, naïve bayes classifier analyzes each feature independently [5].

"Naïve Bayes classification is based on Bayes theorem with naïve (strong) class conditional independence. Class conditional independence means the effect of an attribute value on a given class is independent of the values of other attributes.[5]"

The equation for Bayes theorem is given as follows:

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

IV. EXPERIMENTAL SETUP

For doing the proposed experiment, firstly Rapidminer is to be loaded with crime dataset. The sample dataset for this experiment on crime is taken from Indian government website [9]. This site provides the crime data on different parameter like acid attack, theft, kidnap, road rage etc. of year 2012, 2013 and 2014. The data is available in different format like excel, CSV, JSON, ODS etc. We have downloaded CSV format data file from this site as shown in Fig. 1.

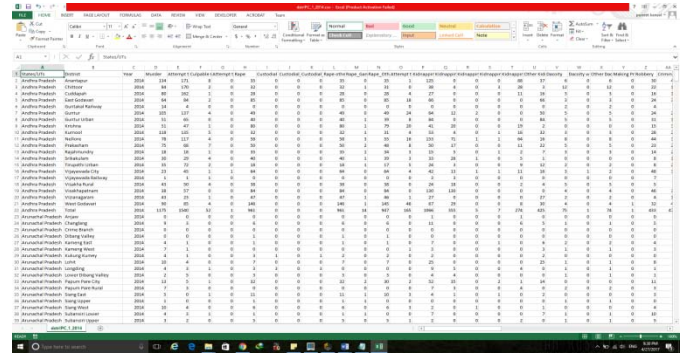


Fig. 1 CSV format of crime data

The downloaded CSV file is required to be uploaded in RapidMiner for further operation. The RapidMiner provides different type of processing scenario for modelling like Churn Modelling, Market Based Analysis, Outlier detection etc. We will use blank process for this experiment and create process model by putting different operators in Fig. 2 like retrieve, Numerical to Polynomial, Set Role, validation rule, naïve bayes classifier etc.

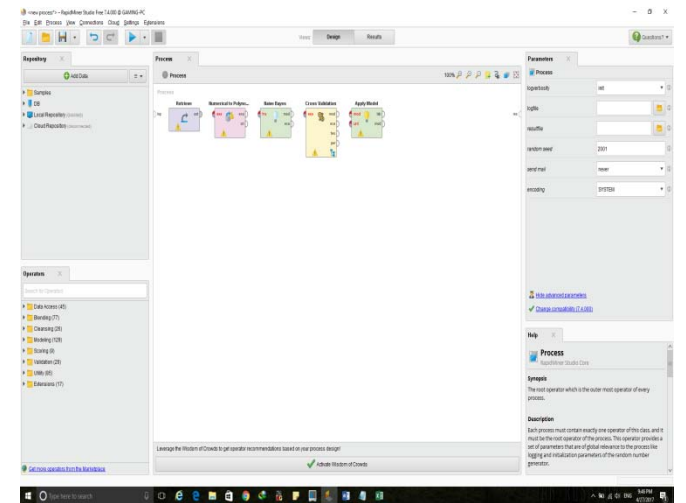


Fig.2Rapid Miner GUI with different operator

The aim of this experiment is to supply the input along with validation rule to the naïve bayes operaor and see the results. This setup will give the performance evaluation of naïve bayes classifier. Following steps are to be incorporated in Rapid Miner for setting up the process and getting desired result using naïve bayes classifier.

- i Create new Blank process in Rapid Miner.
- ii Load database by using “Add Data” option.
- iii Insert Operator Retrieve_Crime_Data, Set Role, Split Data, Naïve Bayes, Performance and apply model.
- iv Configuration is created in all inserted operator.
- v Finally, when process is run, this setup gives the output based on naïve bayes classifier.

V. METHODOLOGY AND RESULTS

The RapidMiner provides the naïve Bayes classifier, which is used to plot the crime data to see different statistical chart of crime in different states. The sample data can be downloaded from Indian government website. After downloading the data in CSV format as in Fig. 3, it should be loaded in RapidMiner for operation.

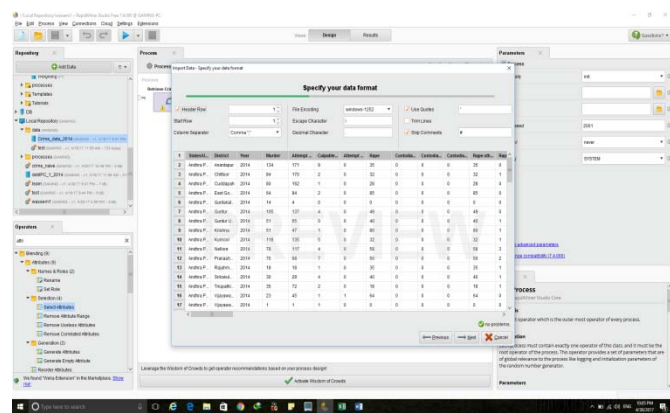


Fig. 3 Import of CSV file

The loaded data is available under ‘Local Repository’ option in Rapid Miner. For plotting of data, a new blank process is added in Rapid Miner.

After retrieving data, the role are set (States/UTs), further the data is split into ratio of 75/25 using split data operator. Now the Naïve Bayes operator is inserted and finally the performance operator is given to apply model operator Fig. 4

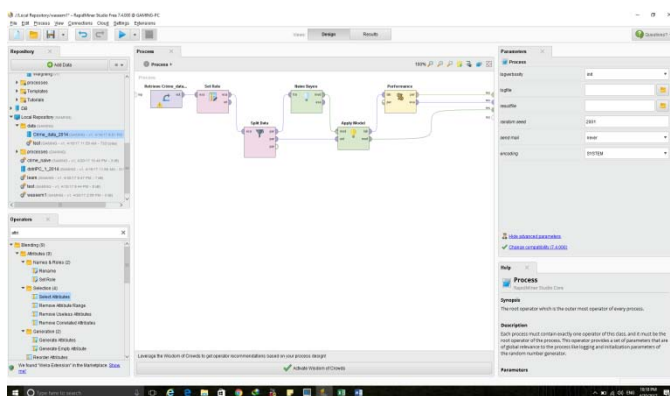


Fig. 4. Display of Naïve Bayes Classifier

When the process is run, then following output is generated in following:

- i) SimpleDistribution
- ii) ExampleSet (Apply Model)
- iii) PerformanceVector (Performance)

The above three categories having further classification which are following:

- i) SimpleDistribution:**
 - a. Description (Fig. 5)
 - b. Charts (Fig. 6)
 - c. Distribution Table (Fig. 7)
 - d. Annotations
- ii) ExampleSet (Apply Model)**
 - a. Data (Fig. 8)
 - b. Statistics (Fig. 9)
 - c. Charts (Fig. 10)
 - d. Advanced charts
 - e. Annotations
- iii) PerformanceVector (Performance)**
 - a. %Performance (Fig. 11)
 - b. Description
 - c. Annotations

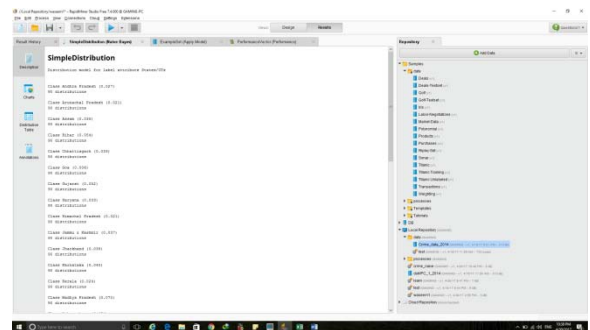


Fig.5 Description: SimpleDistribution

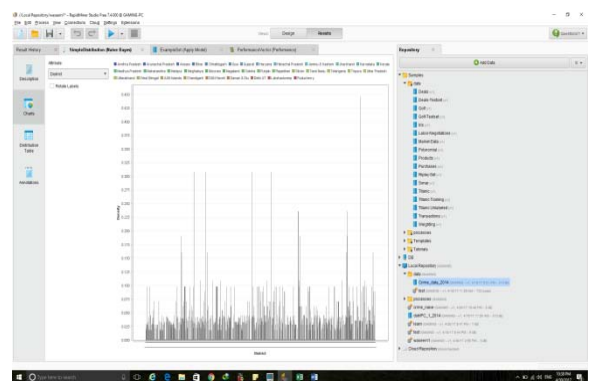


Fig. 6 Charts: SimpleDistribution

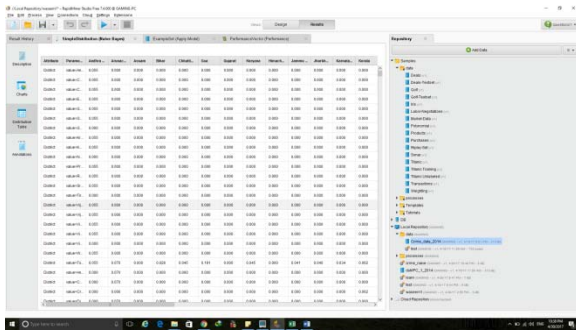


Fig.7 Distribution Table: SimpleDistribution

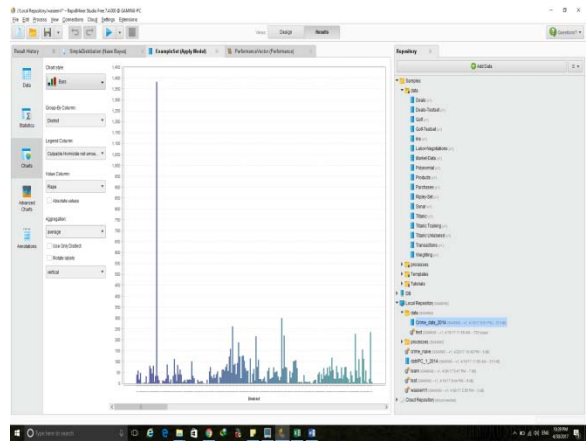


Fig. 10 charts: ExampleSet

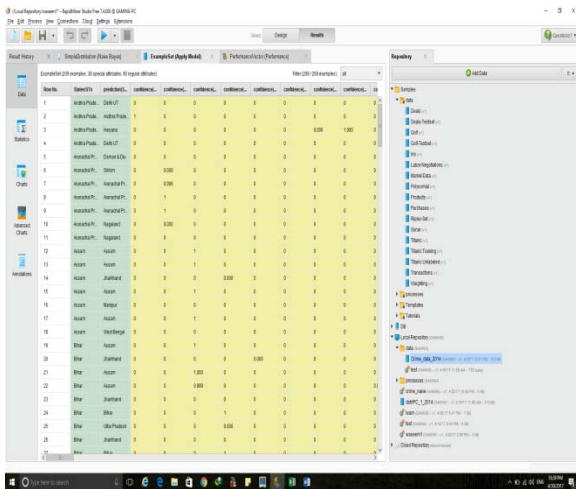


Fig.8 Data: ExampleSet

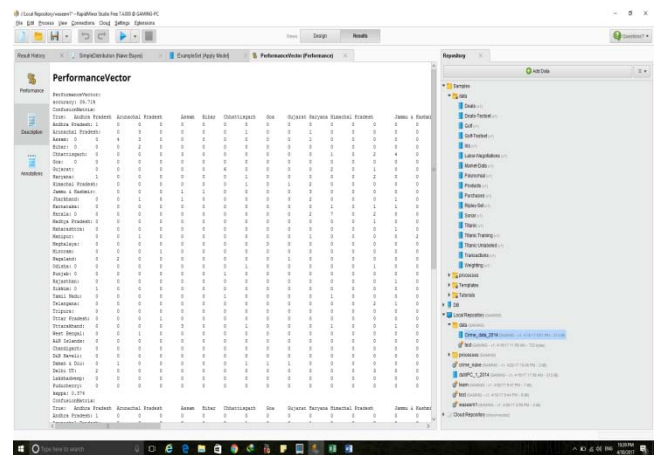


Fig.11 Percentile performance of different types of crimes

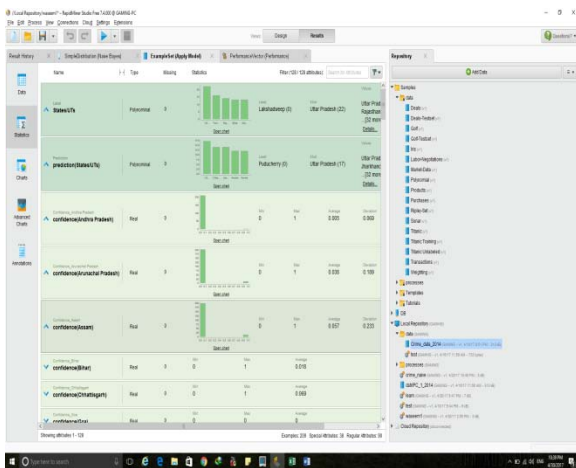


Fig.9 Statistics: ExampleSet

VI. CONCLUSION

This experiment shows how naïve bayes classifier in Rapid Miner data mining presentation tool can be used to display crime data. The naïve bayes is a predictive classifier, which can be used to take different decision based on the result. In this experiment the crime data is used as input and different chart are populated based on the input (refers figures). These charts can be used to take decision in controlling the crime by comparing the different type of crime in different states. The chart shows that which state having what crime less and what crime is more. The crime rate can be compared between different states. The important thing in this experiment using naïve bayes classifier is the availability of sample data. As the data is used in this is of government data, so if more data can be collected then more accuracy can be achieved and more conclusion can be drawn. The reason of using Rapid Miner, compared to other data mining tool like WEKA, orange and R is that, it provides the fully automatic parameter optimization of machine learning operator and provide good validation and cross validation.

VII. REFERENCES

- [1] Anshul Goyal, Rajni Mehta, "Performance Comparison of Naive Bayes and J48 Classification Algorithms", International Journal of Applied Engineering Research, Vol. 7, No. 11, 2012, ISSN 0973-4562
- [2] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, second Edition.
- [3] R. Kiani, S. Mahdavi, A. Keshavarzi, "Analysis and Prediction of Crime by Clustering and Classification", International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No. 8, 2015 pp. 11-17
- [4] Jyoti Aggarwal, Renuka Nagpal, Rajni Sehgal. "Crime Analysis using K-means Clustering", International Journal of Computer Applications, Vol. 83, No. 04, ISSN 0975-8887, pp. 1-4
- [5] S. Deshmukh, A.S. Dalvi, T. J. Bhalerao, "Crime Investigation using Data Mining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, No. 03, 2015, ISSN 2319-5940 pp. 22-24
- [6] S. Kapoor, A. Kalra, "Data Mining for Crime Detection", International Journal of Computer Engineering and Applications, Vol. 7 pp. 88-97
- [7] O. Georgina N., A. Isah and J. Alhasan, "Analytical Study of Some Selected Classification Algorithms in WEKA Using Real Crime Data", International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No. 12, 2015 pp. 44-48
- [8] S. Santhosh Baboo and Malathi A and Anbarasi A "An intelligent Analysis of a city Crime Data Using Data Mining" IPCSIT vol. 6, 2011 pp. 130-134
- [9] <https://data.gov.in/catalog/district-wise-crimes-under-various-sections-indian-penal-code-ipc-crimes>