ISSN No:-2456 –2165

Research and Improvement on using Enhanced Clustering Data Mining Algorithm for Crime Detection and Criminal Identification

Dr. Gurpreet Singh
Professor & Head
Department of Computer Sc. & Engg.
St. Soldier Institute of Engg. & Tech. Jalandhar (Punjab), India

Amarjit Singh
M.Tech Scholar
Department of Computer Sc.& Engg.
St. Soldier Institute of Engg. & Tech. Jalandhar (Punjab), India

Abstract: Data mining clustering algorithm is used for defining the type of crime and classification algorithm used for violence level of juvenile. The fight against crime is not a new one in humanity and it has, since the establishment of society, tried to bring crimes down. In this thesis work data extraction, data preprocessing are used to clean and pure database so that data mining techniques can easily applied on them. A clustering algorithm is used for grouping similar level of criminals and classification is used to predict criminal behavior.

The problems taken for this research work is divided into some objectives which are as follows:

- **To extract and preprocess the incomplete and inconsistent data.**
- **To accurately and efficiently analyze the growing volume of data.**
- **To explore and enhance clustering and classification algorithm to identify type of crime.**
- **To construct an efficient framework for predicting deviant behavior of juveniles**
- **To increase the outcomes with clusters superiority and performance.**
- **To decrease the error rate and achieve correctness.**
- **To reduce the computational time of execution.**

I. INTRODUCTION

An analytical process which is developed to examine data in form of patterns which are consistent is known as data mining. From ages only the physical excerption of patterns from data is going. The data collection, storage and manipulations have increased by the accretion and prevalence computer technology. Due to the grown size as well as complexity of datasets the direct manual analysis has amplified with indirect and automatic processing of data. These are various methods as clustering, neural networks, other genetic algorithms, vector support machines and decision trees which are applied to data with aim of not hiding the patterns which are hidden. Data mining is sometimes turns as knowledge discovery and its tools are here to predict behaviors and future trends making proactive business and knowledge driven decisions. These sophisticated data analysis tools used by data mining are to discover not the known, patterns and relationships of them in large datasets. Various models as statistical and mathematical along other machine learning methods can be involved as tools of data mining. These are basically the algorithms which improve their performance automatically by experience such as neural networks and decision trees.

II. DECISION TREES

There is assortment of algorithms presence used in grouping procedure. One if these are the DT attitude. To characterize together the worsening prototypes and classifiers result tree in the ceremonial of predicative exemplary is charity. Verdict tree really us the hierarchal prototypical of resolutions and their magnitudes. The arrangement of DT excludes subdivision, root knob and leaf swelling. Aspects test is designated on each wait node, the test result is signified by subdivision and course labels are exposed by leaf node. The uppermost node is the root node of the sapling. The tree scholarship is completed by dividing the spring into agreed which are commonly created on a assessment of feature worth. The top depressed method of DT sets an example of acquisitive algorithm. Apart from this bottom-up.

ISSN No:-2456 –2165

III. IMPLEMENTATION AND RESULT

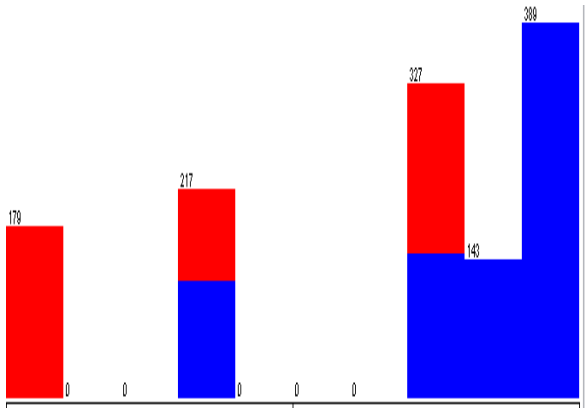

Figure 1 Month Attribute

Shows that the attribute is month. So its min, max ,mean and standard deviation values are depicted in the figure. Graph shows the distribution of crime in the attribute.

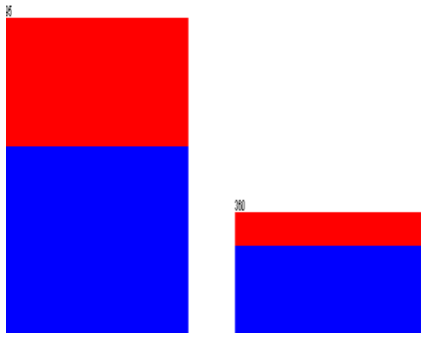

Figure 2 Location Attribute

Shows that the attribute is location type. So its (station & on train) deviation values are depicted in the figure. Graph shows the distribution of crime type in the attribute.

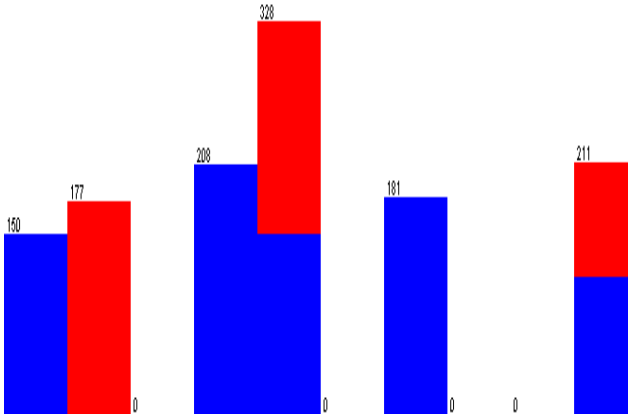

Figure 3 Age Attribute

Shows that the attribute is age. So its min, max ,mean and standard deviation values are depicted in the figure. Graph shows the distribution of crime type in the attribute.

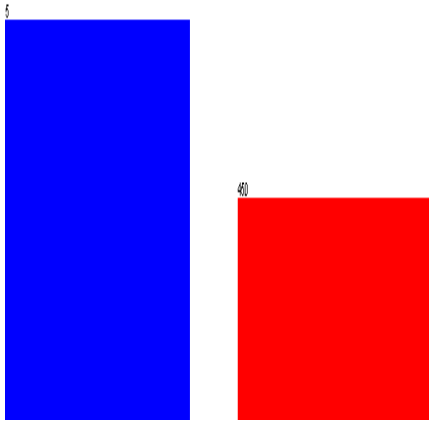

Figure 4 Nominal Attribute

Describes the reason attribute which is a nominal attribute. In case of nominal attribute WEKA specifies the various values (labels) under that attribute along with number of instances under that label of the attribute. Graph shows the distribution of crime type is reason attribute.

	KMeans	ECA
Iterations	11	19

Table 1: Results of Number of Iterations in Simple Kmeans & ECA

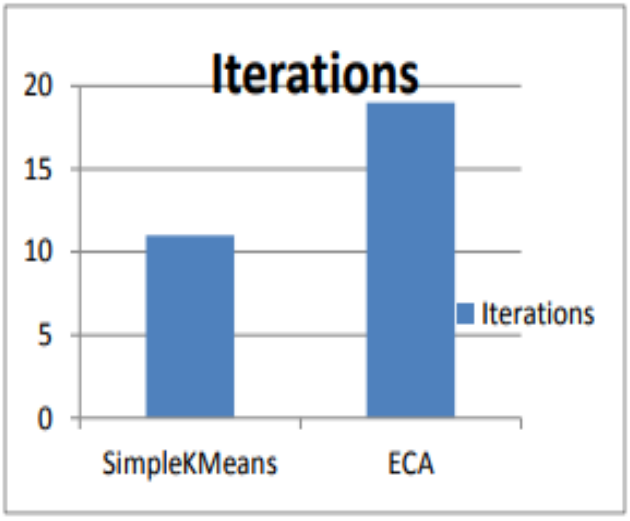

Figure 5 Graphical Representation of Iterations Results.

Shows the graphical representation of iterations results. Blue color depicts number of iterations. It shows 19 number of iterations are ECA and as 11 number of iterations are simple kmaens.

	KMeans	ECA
Error Rate	416	371

Table 2: Error Rate of Simple kmeans & ECA

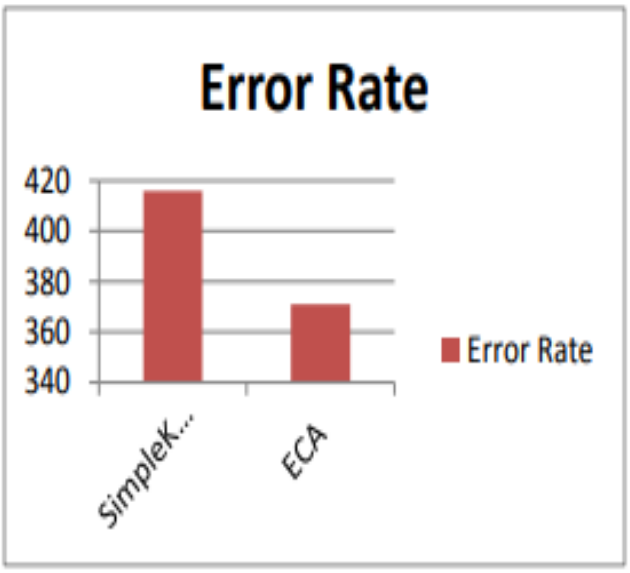

Figure 6 Graphical Representation of Error Rate

Shows the graphical representation of error rate. Red color depicts number of error rate. It shows as 416 number of erroe rate is simple kmean and as 371 number of error rate is ECA.

IV. CONCLUSION

In this explore; learn is being accomplished on DT algorithms. The form of kmeans and ECA algorithms are common and a new algorithm ECA is projected. The evaluation of projected algorithm is completed with the active algorithms and kmeans on mobile services dataset using WEKA measurements mining tool. The concern by altering the closely classified instances, incorrect classified instances and error rate doctrines identify that the projected process gives improved recital than kmeans and ECA by falling the error rate which signifies that ECA have elevated intra comparison and is additional precise. Also the expected algorithm can touch big datasets more successfully.

REFERENCES

[1]. Eung SakKwong, Lee Hong Kin, Wang Feng Jin, "Data Mining on DNA Sequences of Hepatitis B Virus", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 8, no. 2, march/april 2011.
[2]. MansooriG.Eghbal, "FRBC: A Fuzzy Rule-Based Clustering Algorithm", IEEE Transactions on Fuzzy Systems, vol. 19, no. 5, October 2011.
[3]. Malathi. A, Dr.BabooSanthosh S, " An Enhanced Algorithm to Predict a Future Crime using Data Mining", International Journal of Computer Applications (0975–8887) Volume 21– No.1, May 2011.