ANALYZING THE USES OF DATA MINING IN BIOLOGICAL AND ENVIRONMENTAL PROBLEMS

VRATNAKUMARI[1], DR. ARVIND KUMAR SHARMA[2]

Department of Computer Science and Engineering

[1,2]OPJS University, Churu (Rajasthan) – India

ABSTRACT

Data Mining is a knowledge discovery from data and it treats as mining of knowledge from large amount of data in every field. The algorithms are implemented using MATLAB and fuzzy logic tool box and results are evaluated based on performance parameter in both algorithms. After doing this research experiment results show that how k-means and fuzzy C means implemented on protein data set. FCM allows one piece of data to belong to two or more clusters. Results based on different clusters in both algorithms. K-means is the centroid based technique. We are also compared k-means and FCM results in this research. Comparison results show that the k-means is better than FCM. With the help of this research we can remove complexity from data sets in future. So the result shows that proteins are close to each other and k-means algorithm remove data set complexity with high accuracy and less consuming time and found large sum of distance in among the statistics peak's association to FCM algorithm. In this research work we present the problem that show proteins are highly affiliated to each other.

1. INTRODUCTION

Data mining is a knowledge discovery from data base. Knowledge discovery is the computer process of digging through analyzing of data and after that extracting the meaning of the data. Data mining tools helpful for answer business question that traditionally were too less time consuming. We can use data mining tool for other things []1. They scour data sets finding predictive and other information that experts may miss because it lies outside their expectations. Data mining tools and methods allowing business to make proactive, predict behaviors, finance, knowledge-driven decisions, future trends and bioinformatics.

Data mining is practical and chief for companies in a broad range of industries including retail, manufacturing transportation, health care, medical science; finance and aerospace are already using data mining tools and techniques totake advantage of historical data. With the help of model thanks technologies and statistical and mathematical techniques to sift through warehoused information, data mining analysts recognize significant facts, patterns, and relationship exceptions. For business purpose data mining is used to

relationships in the data in order to help make better business decisions and discover patterns. Data mining can predict customer loyalty, develop smarter marketing campaigns and it is help to spot sales trends [2].

2. DATA MINING

Data mining is a perfectly interdisplinary subject. Data mining can be explains in many different ways. Data mining should have been more unlike named first is knowledge mining from database and second is knowledge discovery from data set [3]. Many people treat data mining and concepts as a synonym. It's for another popularly used term and the others view data mining as simply an essential step in the course of data mining. The knowledge discovery process presents in figure1 as an iterative run of the following steps. In this figure we can see that how data mining works [4].
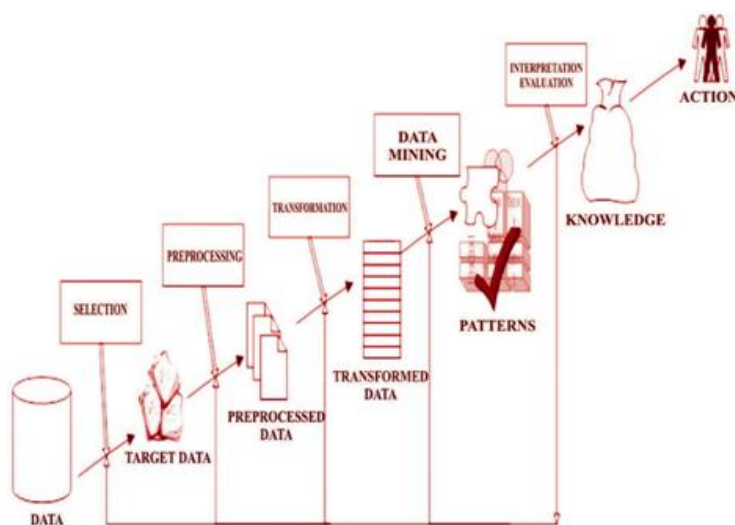

Figure 1: Data mining as a step in the process of knowledge discovery

2.1 Need of Data Mining

We live in a world where loads quantities of data are serene daily. Analyzing such data is a very key need. Data mining can meet this need by providing tools; methods and techniques to discover knowledge from real data .We examine how data mining can be viewed as a result of the progression of information technology, science and engineering [5].

Why data mining. We are living in the information age. This age is a usual saying; however, we are truly income in the data age which is group similar to a record. Data mining worn to the World Wide Web (www) and diverse data storeroom devices every day from business, medical science, medicine, bioinformatics, science, and engineering roughly every other piece of daily life,in these fields increase of offered data volume is a result of the mechanization of our society. In businesses worldwide create gigantic data sets, including stock, transactions, sales trading records, promotions, company profile and routine and customer pointer.

3. ENVIRONMENTAL AND BIOLOGICAL STUDIES IN DATA MINING

Firstly we are unfolding what is immediate. Immediate is the biotic and a biotic immediate of an organism, or population. Environmental includes mostly the factors that have an authority in their increase and progress. Green politics is a broad social lobby group that, in a large part, seeks to decrease and give back the downbeat outcome of human source activity on the biological environment [6]. The focal aim of alarm for environmentalists the more main ones being species extinction, pollution, and disaster and old growth forest loss.

3.1   Multi-disciplinarynature of environmental studies

Environmental nature is more difficult, its knowledge inputs from various disciplines of science like physics, chemistry, earth science, social science, law and engineering and technology are included in environmental studies in figure 2.
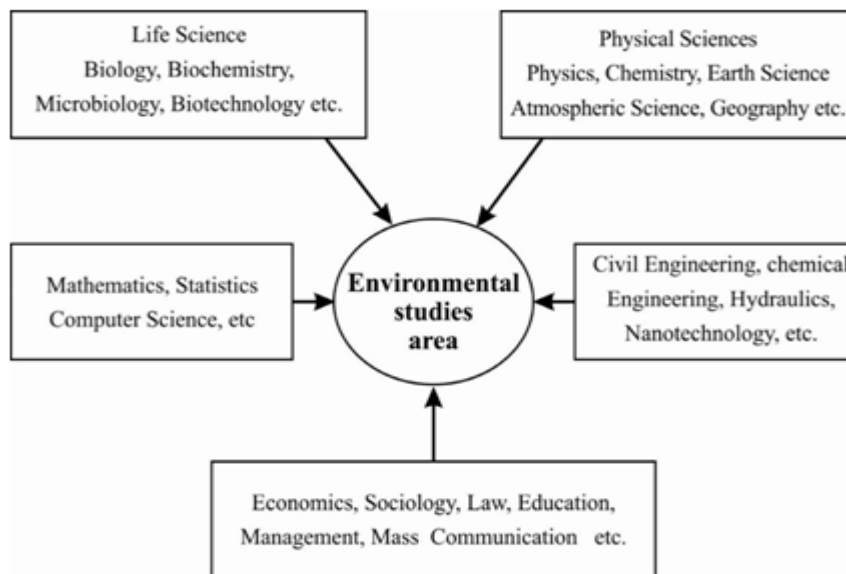

Figure 2: Environmental studies area

4. RESEARCH & DEVELOPMENT (R & D) IN ENVIRONMENT

Data mining has an important position to participate in probing various green problems and carry out R&D programme for budding cleaner technologies, methods and promoting sustainable development [7].

4.1 Emerald advocacy

With growing emphasis on implementing a choice of acts and laws linked to setting, has emerged. These types of acts should be able to plead the cases related to water and air pollution, forest etc.
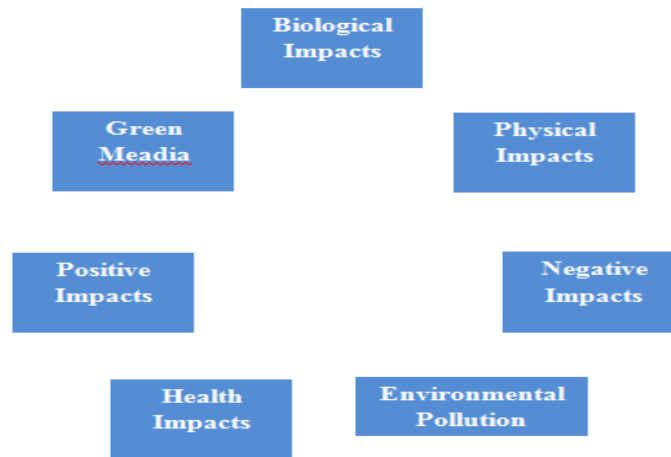
A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories

International Journal in IT and Engineering

http://www.ijmr.net.in **email id- irjmss@gmail.com   Page 98**

Figure 3: Impact of Mining activities

4.2 Olive media:

Green alertness can be spread among masses through mass media such as newspaper, radio, magazine, television, hording, advertisement, internet etc. Figure 3 shows that the how environmentally sophisticated someone are necessary.

Figure 3 shows that the blow of data mining activates in environment. Open pit mining in which equipment dig holes and do away with the ores (copper, iron, gravel, limestone).

Surroundings feel right to all and it is important to every person. Whatever period of a human being, person's body, hair etc? In situation many type of exertion are found. With the help of data mining these types of exertion we will remove. Such that Human people problem which is unnatural by any disaster [8]. In data mining field data interchange is most important quandary in any company. So we can remove this problem also.

➢ Most significant problems of environment

➢ Climate transform Global warming

➢ Changes in climate

➢ Contamination from factories

➢ Fumes from road interchange

➢ Deep interchange

➢ Recycling

➢ Forest fire

➢ Earth nippy

➢ Massive win

4.3 Classification techniques

Classification presents a set of preset classes. In arrangement each tuple/sample is held to be in the right place to a predefined class [9]. This predefined class is calculated by the class

label aspect. Training set worn for model manufacture and this model is represented as classification rules. Classification rules are decision tree. Test set is self-regulating of training set. Test sample is compared with the classified result from the model. Classification is the supervised learning. In supervised knowledge the training data (annotations dimensions etc.) are accompanied by labels indicating the class of the observations.

4.4 Classification methods

1. Predictive accurateness

2. Speed and scalability - time to make the model and time to use the mock-up

3. Robustness- handling clamour and absent values

4. Interpretability- accepting and insight proved by the representation

5. Goodness of rules- decision tree volume and smallness of sorting rules

The likely atmosphere hazards have its significance to two facts [10]. First the spoil and thrashing of life inflicted upon human society are often substantial catastrophic events, and second the environmental control, the large history of experience is found in the record of how people have coped with the natural hazards.

5. CONCLUSION

Biological sequences include protein progressions. Such chains are usually especially stretched and complicated. Proteins are close to every other. Here, gaps are usually important. Thus this work will be favorable for the data analysis effortlessly in bioinformatics. Currently the research of a species from its genomic database is a testing task. Data clustering is a well known quandary in a spatial data analysis. Natural environment hazards are those conditions or processes in the environment that present increases to monetary break or loss of life in person people.Only few researchers work on scientific data. Analyzing the methodical data is very different task. This research is important to eradicate the complexity of data set. Analysis of biological and ecological data set is very thorny assignment in data mining technique.Data Mining is a knowledge discovery from data and it treats as mining of knowledge from large amount of data in every field.

5. REFERENCES

[1] Anjali B Raut, Ankita A Nichat 2017. Students Performance Prediction Using Decision Tree Technique *International Journal of Computational Intelligence Research* ISSN 0973-1873 Volume 13, pp. 1735-1741

[2] BanerjiGeetali and SaxenaKanak (2012), "An Efficient Classification Algorithm for Real Estate domain", International Journal of Modern Engineering Research (IJMER), 2(4): 2424-2430.

[3] Cieslak D. and N. Chawla N.(2009), "A framework for monitoring classifiers performance: when and why failure occurs?)", *KAIS*. 18(1): 83–108.

[4] Dietterich T. (2002), "Ensemble learning, in The Handbook of Brain Theory and Neural Networks", 2nd ed., M. Arbib, Ed., Cambridge MA: MIT Press.

[5] Elena Makhalova, (2013), "Fuzzy C means Clustering in MATLAB", The 7th International Days of Statistics and Economics, Prague, 19(21): 905-914.

[6] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H., (2009), "The WEKA Data Mining Software)", An Update. SIGKDD Explorations, Newsletter, 11(1): 10–18.

[7] Jain Shreya and GajbhiyeSamta (2012), "A Comparative Performance Analysis of Clustering Algorithms", International Journal of Advanced Research in Computer Science and Software Engineering, 2(5):441-445.

[8] Opitz D. and Maclin R.(1999), "Popular ensemble methods: an empirical study.

[9] Skurichina. M. and Duin R.P.W. (1998), "Bagging for linear classifiers. Pattern Recognition", 31(7): 909–930.

[10] Sorokina D., Caruana R., Riedewald M.(2007), "Additive Groves Of Regression