www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 4 April 2016, Page No. 16216-16217

Role of K-Means Algorithm in Disease Prediction

Purvashi Mahajan[#1], Abhishek Sharma[#2]

M.Tech[#1], Assistant Professor[#2]

Department of Computer Science,

Shri Balwant Rai Institute of Technology

DCrust University, Murthal

purvashi.mahajan@gmail.com[#1], abhi.sbit@gmail.com[#2]

Abstract: ***Disease Prediction has always been a matter of research due to increasing number of health risks. Modern Medicine System produce huge amount of data which needs to be organized. Medical Data Mining plays a vital role in generating efficient results when it comes to prediction based analytics. Data Mining turns data into patterns which are useful for analyzing the diseases. This research paper focuses on role of K-Means Algorithm in disease prediction.***

Keywords: Data Mining, K-Means Algorithm, Medical Databases, Clustering Algorithm

1. Introduction

Data Mining infers knowledge extraction or knowledge discovery from huge and unorganized data. Goal of Data Mining is to derive the associated patterns as per requirement. Data Mining involves various steps: Data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation.

Medical Data Mining is a dedicated term which is only restricted to medical domain or predictive analysis of diseases. There are high risky consequences because of doctors' assumptions and lack of forte in a particular area. Data Mining then plays its part in deriving out the patterns from the historical data of the patients. These patterns are really useful in predicting the future aspects.

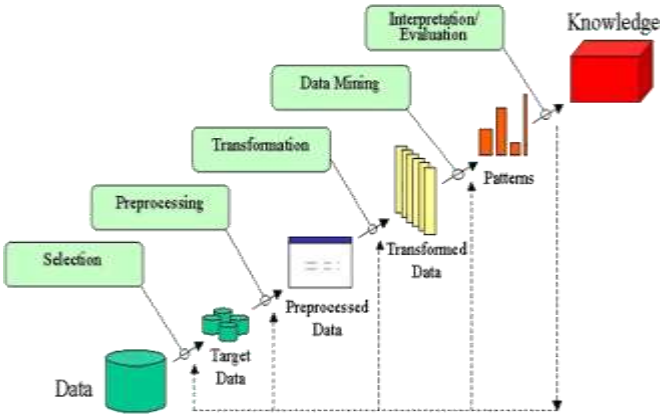

Figure 1: Steps in Data Mining Process

2. Literature Review

- Rebecca Hermon et al [1] presented Big Data in Healthcare: What is it used for. It explained the useful aspects of big data techniques in healthcare. Concluded was the impact of integration of data mining and medical informatics and its analysis using big data techniques on healthcare delivery costing and better heathcare results.
- G. Santhosh Kumar, Lakshmi K.S et al [2] proposed a new method of uncovering valid association rules from medical transcripts. The extracted rules describes association of disease with other diseases, symptoms of a particular disease, medications used for treating diseases, the most prominent age group of patients for developing a particular disease.
- Ji-Jiang Yang, Jianqiang Li, Jacob Mulder d, Yongcai Wange, Shi Chen f, Hong Wug, Qing Wang b, Hui Pan et al [3] proposed the main goal of this special issue and gives a brief guideline. Then, the present situation of the adoption of EMRs is reviewed. After that, the emerging information technologies are presented which have a great impact on the healthcare provision. These include health sensing for medical data collection, medical data analysis and utilization for accurate detection and prediction.
- Jyoti Soni et al [4] presented Predictive data mining for medical diagnosis. The experiments performed to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and some time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well.

3. Clustering Algorithms

Purpose of using clustering algorithms is to decrease the amount of data by grouping identical data items together.

Groups of similar objects or data are called clusters. Clustering Algorithms are useful in minimizing the human efforts and eventually deriving accurate and fast results. For Instance Medical Databases are so huge and complicated, here comes the part of clustering algorithms, which organize the data so

that patterns would be created for better analysis of diseases. Pattern Recognition and Data Analysis are the main advantages of clustering algorithms.

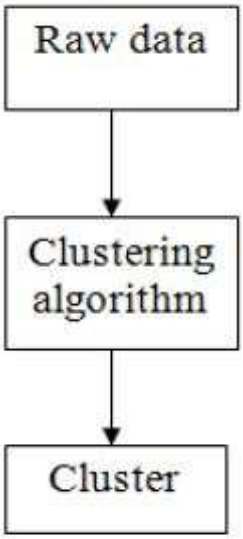

Figure 2: Clustering Stages

In the above figure, Clustering algorithm is being applied to the raw data (medical data, corporate data or any other form of data) to form clusters for further processing.

4. K-Means Algorithm

k-Means Algorithm is one of the simplest unsupervised learning algorithms that aids in solving the hugely known clustering problem. K-Means Algorithm is applied for dealing with medical databases for clustering. It is a partition based clustering technique. To increase the efficiency of mining process, some pre-processing needs to be done to the data. It is a semi-automatic process for analyzing huge databases to find patters that are related, valid, useful and user-friendly. The algorithm works on a set of *d*-dimensional vectors, $D = \{xi \mid i = 1, \ldots, N\}$, where $xi \in Rd$ denotes the i^{th} data point. The algorithm is initialized by picking *k* points in R*d* as the initial *k* cluster representatives or "centroids".

4.1 Steps

K-Means Algorithm follows the following steps:

1. K Value is fixed in advance and K center is chosen randomly.
2. Data Objects are being assigned to the nearest center.

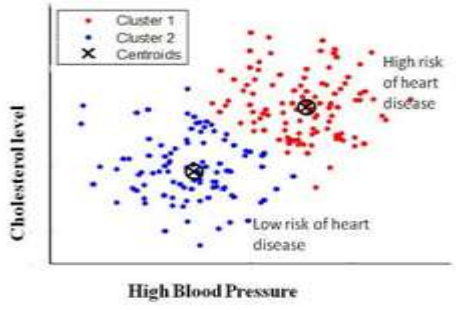

Figure 3: K-Means Clustering for Heart Disease Patients

5. Conclusion

Clustering Algorithm i.e. K-Means Algorithm can be quite beneficial in Medical Data Mining or Predictive Analysis of Diseases, but does not produce very accurate results as desired. In order to make K-Means worth for Medical Data Mining or Predictive Analysis of diseases, K-Means should be used in combination or integration with other algorithms so that accurate, relevant and beneficial results can be produced.

6. References

[1] S. Vijayarani and S. Sudha, "An Efficient Clustering Algorithm for Predicting Diseases from Hemogram Blood Test Samples", Indian Journal of Science and Technology, vol.8, pp. 1–8, Aug. 2015.

[2] Jyoti Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications, vol.17, pp. 43–48, Mar. 2011.

[3] K.Rajalakshmi, Dr.S.S.Dhenakaran and N.Roobini "Comparative Analysis of K-Means Algorithm in Disease Prediction", International Journal of Science, Engineering and Technology Research, vol.4, pp. 2697–2699, Jul. 2015.