

DATA MINING

Ansari Gulamraza*¹

*¹Department Of Information Technology, S.I.W.S. College, Wadala, Mumbai, Maharashtra-400031, India.

ABSTRACT

Data mining is a field of intersection of computer science and statistics used to discover patterns in the information bank. The main aim of the data mining process is to extract the useful information from the dossier of data and mold it into an understandable structure for future use. In 2000, India(31.7 million) topped the world with highest number of people with diabetes and it is also predicted that by 2030 type II diabetes may afflict to 79.4 million individuals in India. About 17.5 million people die each year in India form Heart disease. Many researchers are using different data mining tools to help medical professionals in the diagnosis of lifestyle diseases. Researchers reviewed literature on the prediction and diagnosis of heart disease and Type II diabetes by using data mining techniques and applied on healthcare data of patients. This paper highlights the important role played by data mining tools in analysis of huge volume of healthcare related data in prediction and diagnosis of lifestyle diseases.

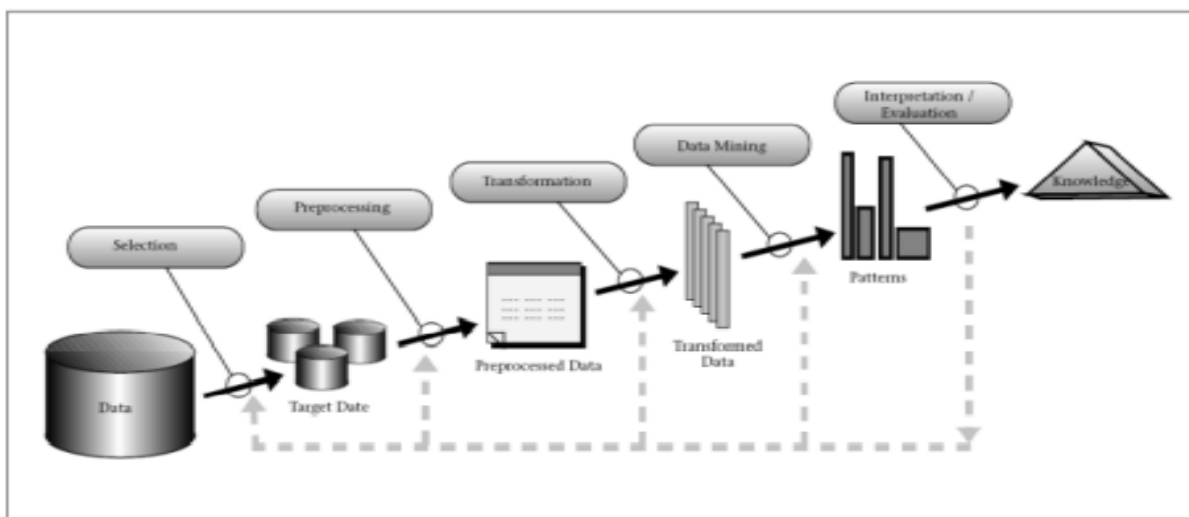
Keywords: Data Mining Techniques; Data Mining Algorithms; Data Mining Applications.

I. INTRODUCTION

Data mining Data knowledge the method of extracting patterns and different helpful info from giant knowledge sets. It's typically called information discovery in knowledge or KDD. Thanks to the rise of massive of massive in knowledge deposit technologies, the utilization of knowledge mining techniques has mature in recent decades, turning information into valuable information that firms will use. though technology has advanced to handle substantial datasets, executives still face automation and measurability challenges. data processing has improved company decision-making through clever knowledge analytics. knowledge knowledge become a valuable tool for businesses wanting to urge ahead within the info economy. during this article, we'll investigate what data processing is, discuss applications of knowledge mining and canopy some data processing techniques. We'll conjointly walk you bit-by-bit through the info mining method.

Overview of Data Mining

The development of data of knowledge Technology has generated great deal of databases and big data in varied areas. data {processing} could be a process of extraction of helpful data and patterns from Brobdingnagian information. The analysis in informationbases and knowledge technology has given rise to associate approach to store and manipulate this precious data for additional deciding. it's additionally referred to as as information discovery method, information mining from information, information extraction or information /pattern analysis.



Data mining could be a logical method that's wont to search through great deal knowledge[of knowledge|of information} so as to search out helpful data. The goal of this method is to search out patterns that were antecedently unknown. Once these patterns area unit found they will any be wont to ensure selections for development of their businesses. Pattern Identification: Once information is explored, refined and outlined for the precise variables the second step is to make pattern identification. establish and opt for the patterns that create the most effective prediction

II. LITERATURE SURVEY

Numerous works has been done connected to life-style illness identification victimisation totally different information mining techniques. The dataset, algorithms, strategies utilized by the authors and therefore the ascertained results beside the longer term work is applied find out economical strategies of diagnosing for varied life-style diseases. Here may be a transient discussion concerning 2 life-style diseases i.e. heart condition and sort II polygenic disease and therefore the work that has been already applied in past few years.

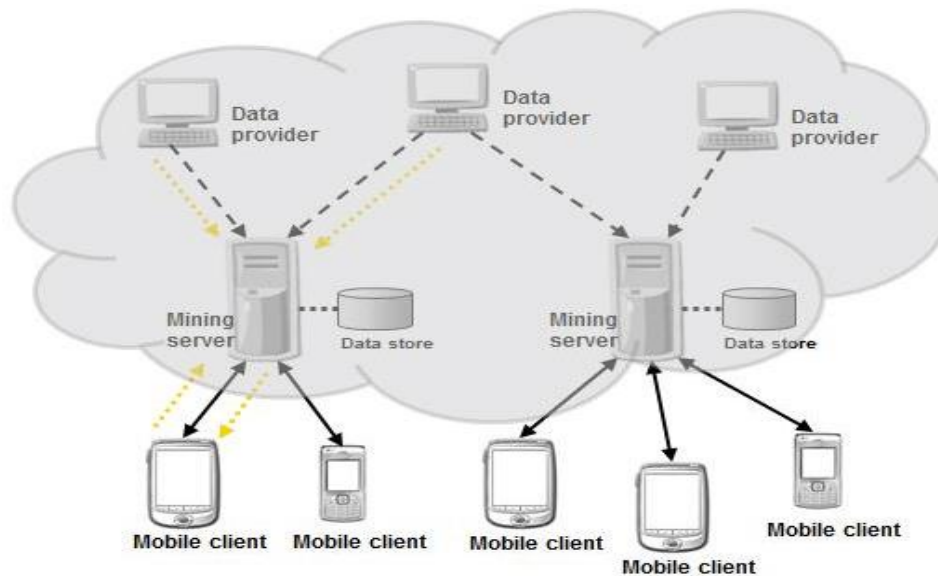
(1). Heart Disease Prediction System using Hybrid System

R. Chitra et.al. Present Hybrid Intelligent techniques for the prediction of heart disease. Some Heart disease classification system was reviewed in this study and concluded with justification importance of data mining in heart disease diagnosis and classification. The classification accuracy can be improved by reduction in features .

(2). Heart Disease Diagnosis Using Fuzzy Logic Approach

P.K. Anooj has proposed a weighted fuzzy rule based CDSS for the diagnosis of heart disease. It automatically obtains the knowledge from the patient clinical data. The proposed CDSS for risk of heart patients consists of two phases. First is an computerized approach for generation of weighted fuzzy rules and decision tree and the second is creating a fuzzy rule based decision support system.

(1)Data Mining in a Network Setting



Today's world is interconnected through many varieties of links. These links embody web content, blogs, and emails. several respondents contemplate community mining and therefore the mining of social networks as necessary topics. Community structures area unit necessary properties of social networks. The identification drawback in itself may be a difficult one. First, it's essential to possess the correct characterization of the notion of "community" that's to be detected. Second, the entities/nodes concerned area unit distributed in real-life applications, and thus distributed means that of identification are going to be desired. Third, a snapshot-based dataset might not be ready to capture the important picture; what's most significant lies within the native relationships (e.g. the character and frequency of native interactions) between the entities/nodes. below these circumstances, our challenge is to know (1) the network's static structures (e.g. topologies and clusters) and (2) dynamic behavior (such as growth factors, robustness, and practical efficiency). an identical challenge exists in bio-informatics, as we have a tendency to area unit presently moving our attention to the dynamic studies of

regulative networks. A queries associated with this issue is what native algorithms/protocols area unit necessary so as to find (or form) communities in a very bottom-up fashion (as within the real world). A concrete question is as follows. Email exchanges inside a corporation or in one's own mailbox over an extended amount of your time will be deep-mined to indicate however varied networks of common follow or friendly relationship begin to emerge. however will we have a tendency to acquire and mine helpful data from them.

Data Mining Algorithms and Techniques:

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

An algorithm in data mining (or machine learning) is a set of heuristics and calculations that creates a model from data. To create a model, the algorithm first analyzes the data you provide, looking for specific types of patterns or trends. The algorithm uses the results of this analysis over many iterations to find the optimal parameters for creating the mining model. These parameters are then applied across the entire data set to extract actionable patterns and detailed statistics.

Data mining is the process of finding patterns and repetitions in large datasets and is a field of computer science. Data mining techniques and algorithms are being extensively used in Artificial Intelligence and Data Science. There are many types of algorithms e.g. C4.5 Algorithm, K-mean Algorithm, Apriori Algorithm ect.

The approaches followed by every data mining technique are different. Researchers are using different data mining techniques for the diagnosis of many diseases. Some of the classification techniques are as under:

(A) C4.5 algorithm :

This algorithm is one of the types of decision tree that was introduced after upgrading the ID3 algorithm. This algorithm can classify the records with noisy and continuous amplitude. When the records are with discrete amplitude, this algorithm operates like ID3 algorithm but when the data amplitude is continuous, it will consider a threshold for all selectable modes and an effective standard is assessed for the threshold and then, the threshold with the highest rate is chosen as the decision index of that node.

(B). KNN algorithm:

K-Nearest Neighbor is an algorithm which is based on similarity with other items. The items which are similar to each other are called neighbors. Once a new item is found, its distance from other items in the model is calculated. This classification partitions the item to the nearest neighbor which is also the most similar one; so places the item in a group that includes the nearest neighbors

(C) Apriori algorithm

Apriori algorithm refers to the algorithm which is used to calculate the association rules between objects. It means how two or more objects are related to one another. In other words, we can say that the apriori algorithm is an association rule learning that analyzes that people who bought product A also bought product B. Apriori algorithm refers to an algorithm that is used in mining frequent products sets and relevant association rules. Generally, the apriori algorithm operates on a database containing a huge number of transactions. For example, the items customers buy at a Big Bazar.

Components of Apriori algorithm

The given three components comprise the apriori algorithm.

1. Support
2. Confidence
3. Lift

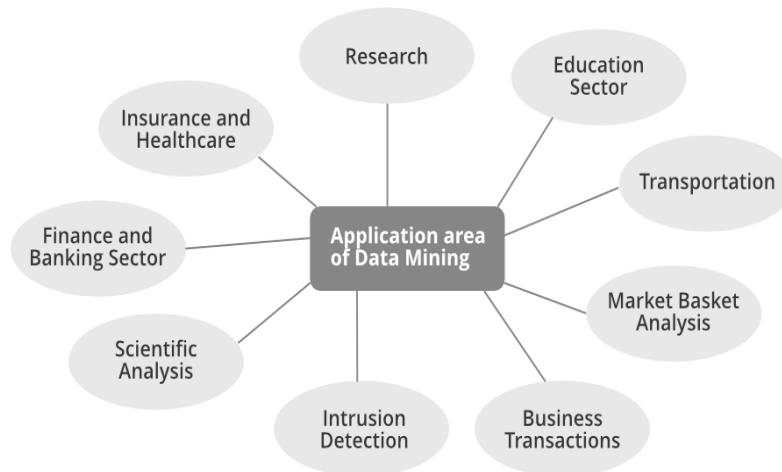
Data Mining Applications

Data mining may be a comparatively new technology that has not absolutely matured. Despite this, there are a variety of industries that are already victimized by it on an everyday basis. A number of these organizations embrace retail stores, hospitals, banks, and insurance firms. This technology is popular in several businesses as a result of it permits them to be told additional regarding their customers and create sensible

selling choices. Here is summary of business issues and solutions found victimization data processing technology.

Here is the list of areas where data mining is widely used –

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific Applications



III. CONCLUSION

Data mining has importance relating to finding the patterns, foretelling, discovery of data etc., in numerous business domains. data processing techniques and algorithms like classification, clump etc., helps to find the patterns to make your mind up upon the longer term trends in businesses to grow. data processing has wide application domain nearly in each trade wherever the information is generated that’s why data processing is taken into account one in every of the foremost necessary frontiers in info and knowledge systems and one in every of the foremost promising knowledge base developments in data Technology. Given the industrial interest in data processing, it’s hardly stunning that variety of package tools have appeared on the market. Some ar general tools, like powerful applied mathematics knowledge exploration systems, whereas others primarily get to place the capability for extracting data from knowledge within the hands of the domain knowledgeable instead of knowledgeable knowledge analyst.

IV. REFERENCES

- [1] Bharati M. Ramageri / Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305
- [2] <https://www.allbusiness.com/Technology/computer-software-data-management/633425-1.html>,last retrieved on 15th Aug 2010.
- [3] Hindawi Computational Intelligence and Neuroscience Volume 2022, Article ID 6439315, 11 pages <https://doi.org/10.1155/2022/6439315>
- [4] International Journal of Information Technology & Decision Making Vol. 5, No. 4 (2006) 597–604
- [5] www.tutorialspoint.com/data_mining/dm_applications_trends.htm
- [6] Data Mining: Statistics and More?, D. Hand, American Statistician, 52(2):112-118.
- [7] www.ijraset.com Volume 5 Issue VI, June 2017 IC Value: 45.98 ISSN: 2321-9653
- [8] Han and Kamber, “Data mining concepts and techniques”, 2nd edition(2010)
- [9] International Journal of Computer Science and Information Security (IJCSIS) Vol. 14, No. 11, November 2016.
- [10] Ling, Charles X., and Chenghui Li. "Data Mining for Direct Marketing: Problems and Solutions." KDD. Vol. 98. 1998.