

# Classification of Dataset Using Business Intelligence for IT Support Services

Ms.Arti Ugale<sup>1</sup>, Prof.S.D.Kamble<sup>2</sup>, Ms.Pratibha Mishra<sup>3</sup>

M.Tech. Computer Sci. & Engg., G.H.Raisoni Institute of Engineering & Technology for Women, Nagpur, M.S, India<sup>1</sup>

Prof, Department of Computer Technology., in yeshwantrao chavan college of Engineering, Nagpur, M.S., India<sup>2</sup>

Prof. Department of Computer Sci. & Engg.,in G.H.Raisoni Institute of Engineering & Technology for Women, Nagpur, M.S, India<sup>3</sup>

arti.ugale@gmail.com<sup>1</sup>

**Abstract-** Business intelligence is a set of theories, methodologies, architectures, and technologies that transform raw data into meaningful and useful information for business purposes. BI can handle enormous amounts of unstructured data to help identify, develop and create new opportunities. BI, in simple words, makes interpreting voluminous data friendly. Making use of new opportunities and implementing an effective strategy can provide a competitive market advantage and long-term stability. In this project we have targeted a techniques, approaches and totally different area of the analysis that are useful and marked because the necessary field of information mining Technologies. As we tend to area unit aware that a lot of mnc's and enormous organizations area unit operated in numerous places of the various countries. Every place of operation could generate massive volumes of information. Company call manufacturers need to access from all such sources and take strategic choices. The data warehouse is employed within the vital business worth by rising the effectiveness of social control decision-making. In associate degree unsure and extremely competitive business atmosphere, the worth of strategic information systems like these area unit simply recognized but in today's business atmosphere, potency or speed is not the sole key for aggressiveness. This kind of giant quantity of data area unit offered within the variety of tera- to peta-bytes that has drastically modified within the areas of science and engineering. To analyze, manage and create a choice on a quantity of information we like techniques referred to as the information mining which can reworking in several fields. This paper imparts additional variety of applications of the information mining and additionally focuses scope of the information mining which can useful within the more analysis in business intelligence.

**Index Terms-** *Clustering; distributed algorithms; Business Intelligence; classification.*

## 1. INTRODUCTION

Data mining is "the machine-driven method of discovering antecedently unknown helpful patterns among structured information." The information warehouse could be a excellent atmosphere to conduct data processing exercises alongside on-line analytical processing. Data processing is that the extraction of hidden prophetic data from giant information bases and it's a strong new technology with nice potential to assist firms specialize in the foremost vital data in their data warehouses through classification and cluster. Business Intelligence could be a idea of applying a group of technologies to convert information into meaning data. Bismuth ways embody data retrieval, data processing, applied math analysis yet as information visual image. Giant amounts of knowledge of information originating completely different in several numerous formats and from different sources may be consolidated and regenerate to key business knowledge. Presents a general read on however information square measure remodeled to business intelligence. The method involves each business consultants and technical consultants. It

converts an outsized scale of information to meaning outcomes therefore on offer decision-making support to finish users.

## 2. LITERATURE REVIEW

In paper [1], In this paper, we presented a two-stage method for creating accurate classifiers for DNA sequences with interesting and comprehensible features. Data mining is the extraction of hidden predictive information from large databases and it is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

In paper [2], different feature selection techniques for classification: Big Data concerns large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data is now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences.

In paper [3], this paper presents an application of business intelligence (BI) for electricity

management systems in the context of the Smart Grid domain. This distribution is the basis for performing on an upper level all the business processes for managing the energy demand and other customer services in a Smart Grid. We study here the problem of secure mining of association rules in horizontally partitioned databases. In that setting, there are several sites (or players) that hold homogeneous databases, i.e., databases that share the same schema but hold information on different entities.

In paper [4] this paper proposed a data mining methodology called business intelligence driven data mining. It combines the knowledge driven data mining and method driven data mining and fills the gap between business intelligence knowledge and existence various data mining methods in e-Business. It setup a four layer frame layer. outlier detection is the data mining task whose goal is to isolate the observations which are considerably dissimilar from the remaining data. This task has practical applications in several domains such as fraud detection, intrusion detection, data cleaning, medical diagnosis, and many others.

In paper [5] this paper presents an application of business intelligence for electricity management systems in the context of the Smart Grid domain. This distribution is the basis for performing on an upper level all the business processes for managing the energy demand and other customer services.

In paper[6] proposed a data mining methodology called business intelligence driven data mining. It combines the knowledge driven data mining and method driven data mining and fills the gap between business intelligence knowledge and existence various data mining methods in e-Business.

### **3. BUSINESS INTELLIGENCE**

Business Intelligence techniques applied to the database of Web-based Educational Systems could help instructors and other educational experts to generate statistics, analytical models, and uncover meaningful patterns from these huge volumes of data. In this article, a framework for applying business intelligence in e-learning environments has been proposed, which increased both flexibility and performance of e-learning environments. Hence, on one hand, the proposed environment enables educational technologists to identify, analyze and monitor relevant aspects of instruction, such as different style, paths, and strategies of learning. On the other hand, such parameters may be used to adapt the learning process to each individual learner and improve the performance of the learning process.

In short, business intelligence helps companies to gain a comprehensive and integrated

view of their business and facilitate better and more effective decision-making and other benefits such as having access to the summarized and distributed relevant information on time. Moreover, companies are provided with a framework capable of introducing and measuring business key performance indicators while analyzing its process and understanding their behavior. In order to carry out these tasks, Business Intelligence uses a wide range of techniques and technologies.

### **4. DATA FEATURE SELECTION**

A “feature” or “attribute” or “variable” refers to an aspect of the data. Features could be discrete, continuous or nominal. Feature selection was a technique of choosing a subset of relevant features, among all the features, that showed the best classification accuracy. The purpose of feature selection was to determine which features are the most relevant to the current classification task and the intention of dimensionality reduction was to improve the classification accuracy performance. It was characterized by three namely relevant, irrelevant and redundant. The feature selection process consists of four components namely feature subset generation, subset evaluation, stopping criteria and results evaluation. The report had highlighted that the quality of the data was first and foremost. If information was irrelevant or redundant or noisy and unreliable then the information would make the knowledge discovery is very difficult

### **5. CLASSIFICATION**

Classification may refer to categorization, the process in which ideas and objects are recognized, differentiated, and understood. In many classification applications, Support Vector Machines (SVMs) have proven to be highly performing and easy to handle classifiers with very good generalization abilities. A large classification problem can be split into mainly easy and only a few hard subproblems. On standard benchmark datasets, this approach achieved great speedups while suffering only slightly in terms of classification accuracy and generalization ability. In addition, DM and BI can give users the ability to spot patterns by putting the data in a visual form. They can further enhance the usefulness of the information by enabling models to identify or confirm relationships, and providing the tools to the user to drill-down and focus on particular areas of interest. If the system is used regularly, comprehensive and timely information can be

utilized to spot technical, organizational, and behavioural problems within the entity in time and with sufficient detail to correct the problem

#### **Distributed Data warehouse Algorithm:**

In the field of data management, data classification as a part of Information Lifecycle Management process can be defined as a tool for categorization of data to enable/help organization to effectively answer following questions:

- What data types are available?
- Where are certain data located?
- What access levels are implemented?

It is worth to observe that several mining algorithms deal with distributed data set by computing local models which are aggregated in a general model as a final step in the supervisor node. This algorithm is different, since it computes the true global model through iterations where only selected global data and all the local data are involved. Data warehouses rule square measure designed to assist you analyze information. for instance, to find out a lot of regarding your company's sales information, you'll build a warehouse that concentrates on sales.

- Knowledge Layer: This layer is on prime interfacing with end-users. It's referred as a data or business decision-making support. This layer harvests business intelligence from patterns that square measure derived from business information
- Data Layer: This layer is liable for providing information supply of data discovery. The info supply has been preprocessed, which suggests remodeling the raw information to clean data.

#### **Clustering methods:**

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Clustering analysis is employed to explore the classification for big dataset and capital of Australia distance is

generalized in order that it will method the info with categorical attributes. The hypothetical analysis shows that the training method and therefore the classifying method of the planned classification technique has nearly linear time quality, that makes the strategy leads to smart measurability and applicable to massive dataset. The experimental results demonstrate that our technique is effective and practicable, and has high prediction accuracy.

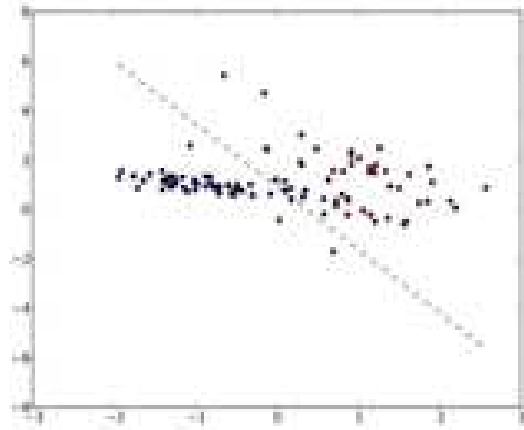
#### **Content based indexing algorithm:**

In this when an algorithm is develop on dataset for index formation and more accuracy. By content based indexing algorithm exact data find out to process the dataset by indexing the clusters. These methodologies are concerned with efficient storage and retrieval of records. The current technology of text-based indexing and retrieval implemented for relational databases does not provide practical solutions for this problem of managing huge multimedia repositories. Most of the commercially available multimedia indexing and search systems index the media based on keyword annotations and use standard text based indexing and retrieval mechanisms to store and retrieve multimedia data. There are often many limitations with this method of keywords based indexing and retrieval especially in the context of multimedia databases. First, it is often difficult to describe with human languages the content of a multimedia object, for example an image having complicated texture patterns. Second, manual annotation of text phrases for a large database is prohibitively laborious in terms of time and effort. Third, since users may have different interests in the same multimedia object, it is difficult to describe it with a complete set of key words. Finally, even if all relevant object characteristics are annotated, difficulty may still arise due to the use of different indexing languages or vocabularies by different users. In content-based retrieval, manual annotation of visual media is avoided and indexing and retrieval is instead performed on the basis of media content itself. There have been extensive studies on the design of automatic *content-based indexing and retrieval* (CBIR) systems.

#### **Support vector machine**

SVM are supervised learning methods used for classification and regression tasks. A classification task usually involves training and test sets which consist of data instances. Each instance in the training set contains one target value and several attributes. The goal of a classifier is to produce a model able to predict target values of data instances in the testing set, for which only the attributes are known. Generally are capable of delivering higher

performance in terms of classification accuracy than the other data classification algorithms.



Key elements in the implementation of SVM are the techniques of mathematical programming. The method, called *Clustering-Based SVM (CB-SVM)*, is specifically designed for handling very large data sets. When the size of the data set is large, SVMs tend to perform worse with training from the entire data than training from a fine quality of samples of the data set.

## 6. CONCLUSION

In this paper, Business intelligence is useful to obtain some guided data mining methods by identifying the related services. The core idea and information behind this architecture is to design a generic system that is flexible enough to suit to different requirement of the business. The business intelligence data of large enterprises can help to generate a very powerful knowledge base. We investigate the issue of classifying large dataset to mixed attributes data and present a constrained clustering algorithm. We regard the clustering results as classification.

## 7. REFERENCES

- [1] Wendy Ashlock, *Student Member, IEEE*, and Suprakash Datta, "Evolved Features for DNA Sequence Classification and Their Fitness Landscapes" *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, VOL. 17, NO. 2, APRIL 2013
- [2] S.C. Punitha, R. Jayasree, Dr. M. Punithavalli, "Partition Document Clustering using Ontology Approach" 2013 International Conference on

Computer Communication and Informatics (ICCCI - 2013), Jan. 04– 06, 2013, Coimbatore, INDIA

- [3] Xindong Wu<sup>1,2</sup>, Xingquan Zhu<sup>3</sup>, Gong-Qing Wu<sup>2</sup>, Wei Ding<sup>4</sup> "Data mining with big data" *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*

- [4] Angelina Espinoza, *Member, IEEE*, Yoseba Peña, *Senior Member, IEEE*, Juan Carlos Nieves, Mariano Ortega, *Member, IEEE*, Aitor Peña, and Daniel Rodríguez "Supporting Business Workflows in Smart Grids: An Intelligent Nodes-Based Approach" *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, 3 AUGUST 2013

- [5] Tamir Tassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*

- [6] Yang Hang, Simon Fong, "framework of business intelligence driven data mining for e-Business" 2010 fifth international joint conference

- [7] Fabrizio Angiulli, *Senior Member, IEEE*, Stefano Basta, Stefano Lodi, and Claudio Sartori, Distributed Strategies for Mining Outliers in Large Data Sets, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 25, NO. 7, JULY 2013

- [8] Mohammad Hassan Falakmasir, Shahrouz Moaven, Hassan Abolhassani, Jafar Habibi, "Business Intelligence in E-Learning"