

## Survey on Data Mining and its Area of Application

Sushma pal\*, Aafreen Jama\*\*

\*Department of computer Science & Engineering Al-Falah School of Engineering and Technology,  
Dhauj -121004(Haryana), India

\*\*Assistant Professor, Department of Computer Science & Engineering Al-Falah School of  
Engineering and Technology, Dhauj -121004(Haryana), India

**ABSTRACT:-** In this Paper we have focused on different area of the research in which data mining play very important role. This includes data mining techniques and function which are used in real life to resolve many problems. In this review paper we have discuss about some field like medical ,education, banking, marketing in which we use data mining techniques and algorithm for organization profit.

**Keyword:-** Data mining, data mining techniques, data mining algorithm, functions, application areas of data mining.

### I. INTRODUCTION

Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. It can be also defined as the process of data selection and exploration and building models using vast data stores to uncover previously unknown patterns.

Data mining, as we use the term, is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. The goal of data mining is to allow a corporation to improve its marketing, sales, and customer support operations through a better understanding of its customers. Data mining comes in two categories directed and undirected. Directed data mining attempts to explain or categorize some particular target field. Undirected data mining attempts to find patterns or similarities among groups of records without the use of a particular target field or collection of predefined classes.

### II. DATA MINING TECHNIQUES AND ALGORITHMS

Data mining algorithms specify a variety of problems that can be modeled and solved. Data mining functions fall generally into two categories:

1. Supervised Learning
2. Unsupervised Learning

Concepts of supervised and unsupervised learning are derived from the science of machine learning, which has been called a sub-area of artificial intelligence. Artificial intelligence means the implementation and study of systems that exhibit autonomous intelligence or behavior of their own. Machine learning deals with techniques that enable devices to learn from their own performance and modify their own functioning. Data mining applies machine learning concepts to data.

#### 1. Supervised Learning

Supervised learning is also known as directed learning. The learning process is directed by a previously known dependent attribute or target or facts. Directed data mining explain the behavior of the target as a function of a set of independent attributes or predictors. The building of a supervised model involves training, a process whereby the software analyzes many cases where the target value is already known. In the training process, the model "learns" the logic for making the prediction. For example, a model that seeks to identify the customers who are likely to respond to a promotion must be trained by analyzing the characteristics of many customers who are known to have responded or not responded to a promotion in the past.

Supervised Data Mining Algorithms: Table (1) describes the data mining algorithms for supervised functions.

TABLE 1

Algorithm	Function	Explanation
<b>Decision Tree</b>	Classification: Classification consists of examining the features of a newly presented object and assigning it to one of a predefined set of classes. The objects to be classified are generally represented by records in a database table or a file, and the act of classification consists of adding a new column with a class code of some kind.	A decision tree model consists of a set of rules for dividing a large heterogeneous population into smaller, more homogeneous groups with respect to a particular target variable. Decision trees extract predictive information in the form of human-understandable rules. The rules are if-then-else expressions.
<b>Naive Bayes</b>	Classification	Naive Bayes makes predictions using Bayes' Theorem, which derives the probability of a prediction from the underlying evidence, as observed in the data.

## 2. Unsupervised Learning

Unsupervised learning is non-directed. There is no distinction between dependent and independent attributes. There is no previously-known result to guide the algorithm in building the model. Unsupervised learning can be used for descriptive purposes. It can also be used to make predictions.

**Unsupervised Data Mining Algorithms:** Table (2) describes the unsupervised data mining algorithms.

TABLE 2

Algorithm	Function	Explanation
k-Means	Clustering: Clustering is the task of segmenting a heterogeneous population into a number of more homogeneous subgroups or clusters. In clustering, there are no predefined classes and no examples. The records are grouped together on the basis of self-similarity. It is up to the user to determine what meaning, if any, to attach to the resulting clusters.	K-Means is a distance-based clustering algorithm that partitions the data into a predetermined number of clusters. Each cluster has a centroid (center of gravity). Cases (individuals within the population) that are in a cluster are close to the centroid.
Apriori	Association: In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together. Based on this data businesses can have corresponding marketing campaign to sell more products to make more profit.	Apriori performs market basket analysis by discovering co-occurring items (frequent item sets) within a set. Apriori finds rules with support greater than a specified minimum support and confidence greater than a specified minimum confidence. For example Find the items that tend to be purchased together and specify their relationship.

### **III. APPLICATION AREA OF DATA MINING**

Data mining is not new—it has been used intensively and extensively by financial institutions, for credit scoring and fraud detection; marketers, for direct marketing and cross-selling or up-selling; retailers, for market segmentation and store layout and manufacturers, for quality control and maintenance scheduling. In Medical Science, data mining is becoming increasingly popular, if not increasingly essential. Several factors have motivated the use of data mining applications in healthcare.

Some areas we discuss here for data mining are:

- Banking
- Medical
- Marketing
- Education

#### **Data mining in banking**

Data mining is a technique used to extract vital information from existing large amount of data and enable better decision-making for the banking. They use data warehousing to combine various data from databases into an acceptable format so that the data can be meaningful. The data is then analyzed and the information that is captured is used by the organization to support decision-making. Data Mining techniques are very useful to the banking sector for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval which is used for fraud prevention, fraud detection in real time, providing segment based products, analysis of the customers, transaction patterns over time for better retention and relationship, risk management and marketing.

#### **A. Customer Retention in Banking Sector**

Data mining can help in targeting ‘new’ customers for products and services and in discovering a customer’s previous purchasing patterns so that the bank will be able to retain existing customers by offering incentives that are individually tailored to each customer’s needs. Churn in the banking sector is a major problem today. Losing the customers can be very expensive as it costs to acquire a new customer. Predictive data mining techniques are useful to convert the meaningful data into knowledge.

To improve customer retention, three steps are needed: (i) measurement of customer retention (ii) Identification of root causes of defection and related key service issues and the (iii) development of corrective action to improve retention. Measurement of existing customer retention rates is the first significant step in the task of improving loyalty. This involves measuring retention rates and profitability analysis by segment.

#### **B. Automatic Credit Approval**

Fraud is a significant problem in banking sector. Detecting and preventing fraud is difficult, because fraudsters develop new schemes all the time, and the schemes grow more and more sophisticated to elude easy detection. Bank Fraud is a federal crime in many countries, defined as planning to obtain property or money from any federally insured financial institution. It is sometimes considered a white collar crime.

Automatic credit approval is the most significant process in the banking sector and financial institutions. Fraud can be prevented by making a good decision for the credit approval using the classification models based on decision trees, Support Vector Machine (SVM) and Logistic Regression Techniques. It prevents the fraud which is going to happen.

#### **C. Marketing**

Bank analysts can also analyze the past trends, determine the present demand and forecast the customer behavior of various products and services in order to grab more business opportunities and anticipate behavior patterns. Data mining technique also helps to identify profitable customers from non-profitable ones. Another major area of development in banking is Cross selling i.e banks make an attractive offer to its customer by asking them to buy additional product or service.

#### **D. Risk Management**

Data mining technique helps to distinguish borrowers who repay loans promptly from those who don't. It also helps to predict when the borrower is at default, whether providing loan to a particular customer will result in bad loans etc. Bank executives by using Data mining technique can also analyze the behavior and

reliability of the customers while selling credit cards too. It also helps to analyze whether the customer will make prompt or delay payment if the credit cards are sold to them.

### **Data mining in Medical for Health care**

There is vast potential for data mining applications in healthcare. Generally, these can be grouped as the evaluation of treatment effectiveness; management of healthcare; customer relationship management; and detection of fraud and abuse.

#### **A. Treatment effectiveness**

Data mining applications can be developed to evaluate the effectiveness of medical treatment. By comparing and contrasting causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective. For example, the outcomes of patient groups treated with different drug regimens for the same disease or condition can be compared to determine which treatments work best and are most cost-effective. Other data mining applications related to treatments include associating the various side-effects of treatment, collating common symptoms to aid diagnosis, determining the most effective drug compounds for treating sub-populations that respond differently from the mainstream population to certain drugs, and determining proactive steps that can reduce the risk of affliction.

#### **B. Healthcare management**

To aid healthcare management, data mining applications can be developed to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims.

#### **C. Fraud and abuse**

Data mining applications that attempt to detect fraud and abuse often establish norms and then identify unusual or abnormal patterns of claims by physicians, laboratories, clinics, or others. Among other things, these applications can highlight inappropriate prescriptions or referrals and fraudulent insurance and medical claims. For example, the Utah Bureau of Medicaid Fraud has mined the mass of data generated by millions of prescriptions, operations and treatment courses to identify unusual patterns and uncover fraud.

## **IV. DATA MINING IN MARKETING FOR MARKET ANALYSIS**

Marketing research is a process of collecting and using information for marketing decision making and plays an essential role in marketing management. Tools for supporting individual phases of marketing research, especially collection and analysis of information can be successfully facilitated by increased use of databases and data mining techniques. As a part of a Marketing Information System such tools provide decision makers with a continuous flow of information relevant to their area of responsibility.

In the area of marketing research are commonly used traditional statistical methods. Our goal is to try modern approaches of artificial intelligence tools on data from the marketing research which deals with consumer behavior in the food market.

The issue of consumer behavior falls into the field of marketing. Into issue of consumer behavior fall Categories of recognition and understanding of how consumers think, feel, evaluate, and choose among different alternatives, how consumers are influenced by their surroundings, how they act during the

Decision-making and purchasing, how is their behavior limited by their knowledge or ability to process information, what motivates them and how they differ in their decision-making in different ways depending on the importance or product interest.

Three basic methods are used, classification with the aid of Multi-layer Perceptron neural network with Back-propagation algorithm, classification with the aid of Bayesian Networks and classification with the aid of Decision Tree. Finally, applicability of these algorithms is compared. These algorithms are applied over the data from a survey about consumer behavior in the food market.

## **V. DATA MINING IN EDUCATION**

Data mining is a powerful tool for academic intervention. Through data mining, a university could, for example, predict with 85 percent accuracy which students will or will not graduate. The university could use this information to concentrate academic assistance on those students most at risk.

In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: Classification, categorization, estimation, and visualization. Classification identifies associations and clusters, and separates subjects under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay." Estimation includes predictive functions or likelihood and deals with continuous outcome variables, such as GPA and salary level. Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. Visualization is used primarily to depict three-dimensional geographic locations of mathematical coordinates.

Higher education institutions can use classification, for example, for a comprehensive analysis of student characteristics, or use estimation to predict the likelihood of a variety of outcomes, such as transferability, persistence, retention, and course success.

## **VI. SUPERVISED AND UNSUPERVISED MODELING**

Classification and estimation use either unsupervised or supervised modeling techniques. Unsupervised data mining is used for situations in which particular groupings or patterns are unknown. In student course databases, for example, little is known about which courses are usually taken as a group, or which course types are associated with which student types. Unsupervised data mining is often used first to study patterns and search for previously hidden patterns, in order to understand, classify, typify, and code the objects of study before applying theories.

Supervised data mining, however, is used with records that have a known outcome. A graduation database, for example, contains records of students who completed their studies, as well as of those who dropped out. Supervised data mining is used to study the academic behavior of both groups, with the intention of linking behavior patterns to academic histories and other recorded information.

This so-called "machine learning" uses artificial intelligence to induct rules and delineate patterns that analysts can apply to new data. Once a model performs well, the analyst can feed in another student group, such as new students, and the model applies the learned information to the new group to predict the likelihood of graduation. All of these steps are automated to produce accurate estimations quickly, saving time and resources compared to conventional behavior prediction methods.

## **VII. CONCLUSION**

Data mining is a technique used to extract vital information from existing huge amount of data and enable better decision-making for the banking and retail industries. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support decision-making. Data Mining techniques are very useful to the banking sector for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval which is used for fraud prevention, fraud detection in real time, . Data mining techniques have been used to uncover hidden patterns and predict future trends and behaviors in financial markets ,Education Market and Medical Sector. The competitive advantages achieved by data mining include increased revenue, reduced cost, and much improved marketplace responsiveness and awareness. This paper therefore recommends various organizations to use data mining techniques in future to resolve complex problems.

## **REFERENCES**

- [1]. K. Chitra, B.Subashini, Customer Retention in Banking Sector using Predictive Data Mining Technique, International Conference on Information Technology, Alzaytoonah University, Amman, Jordan, [www.zuj.edu.jo/conferences/icit11/paperlist/Papers/](http://www.zuj.edu.jo/conferences/icit11/paperlist/Papers/)
- [2]. K. Chitra, B.Subashini, Automatic Credit Approval using Classification Method, International Journal of Scientific & Engineering Research (IJSER), Volume 4, Issue 7, July-2013 2027 ISSN 2229-5518.
- [3]. K. Chitra, B.Subashini, Fraud Detection in the Banking Sector, Proceedings of National Level Seminar on Globalization and its Emerging Trends, December 2012.
- [4]. K. Chitra, B.Subashini, An Efficient Algorithm for Detecting Credit Card Frauds, Proceedings of State Level Seminar on Emerging Trends in Banking Industry, March 2013.

- [5]. Petra Henniker, Andreas Maier, Alex Nippe, Markus Tresch, Douglas Weers, and Peter Zemp, Data Mining at a major bank: Lessons from a large marketing application <http://homepage.sunrise.ch/homepage/pzemp/info/pkdd98.pdf>
- [6]. Michal Meltzer, Using Data Mining on the road to be successful part III, [http://www.dmreview.com/editorial/newsletter\\_article.cfm?nl=bireport&articleId=1011392&issue=20082](http://www.dmreview.com/editorial/newsletter_article.cfm?nl=bireport&articleId=1011392&issue=20082), October 2004.
- [7]. Fuchs, Gabriel and Zwahlen, Martin, What's so special about insurance anyway?, published in DM Review Magazine, [http://www.dmreview.com/article\\_sub.cfm?articleId=7157](http://www.dmreview.com/article_sub.cfm?articleId=7157), August 2003.
- [8]. Kincade, K. (1998). Data mining: digging for healthcare gold. *Insurance & Technology*, 23(2), IM2-IM7.
- [9]. Milley, A. (2000). Healthcare and data mining. *Health Management Technology*.
- [10]. Christy, T. (1997). Analytical tools help health firms fight fraud. *Insurance & Technology*.
- [11]. Gillespie, G. (2000). There's gold in them thar' databases. *Health Data Management*.
- [12]. Kolar, H.R. (2001). Caring for healthcare. *Health Management Technology*
- [13]. Ehsan Hajizadeh\*, Hamed Davari Ardakani and Jamal Shahrabi Application of data mining techniques in stock markets:A survey.
- [14]. Yue, X., Wu, Y., Wang, Y. L., & Chu, C. (2007).
- [15]. Oxford Concise English Dictionary, 11th Edition, Oxford University Press, 2009.
- [16]. Phua, C., Lee, V., Smith, K. & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research, *Artificial Intelligence Review* (2005) 1–14.
- [17]. Wang, J., Liao, Y., Tsai, T. & Hung, G. (2006). Technology-based financial frauds in Taiwan: issue and approaches, *IEEE Conference on: Systems, Man and Cyberspace Oct (2006)* 1120–1124.