4.1 Introduction

Advances in information technologies and the increased digitization of business have led to an explosive growth in the amount of structured and unstructured data collected and stored in databases and other electronic repositories. Much of this data comes from operational business software [e.g., finance/accounting applications, Enterprise Resource planning (ERP), Customer Relationship Management (CRM), workflow and document management systems, surveillance and monitoring systems, and Web log] and is often archived into vast data warehouses to become part of corporate memory. The result of this massive accumulation of data is that organizations have become data-rich yet still knowledge-poor. What can be learned from these mountains of data to improve decisions? How can an organization leverage its massive data warehouses for strategic advantage? A large number of methods with roots in statistics, informational retrieval and machine learning have been developed to address the issue of knowledge extraction from large data sets. The term data mining (DM) refers to this collection of methods. These methods have broad applications; they have been successfully applied in areas as diverse as market-basket analysis of scanner data, customer relationship management, direct marketing, fraud detection, personalization and recommendation systems, risk management and credit scoring.

Data mining is currently in a state of growth and it needs further improvements to attain the development. More products are being developed, more businesses are incorporating the efforts of data mining into their decision making processes. Most of the successful business decisions are made from reliable data source and their validation through the application of tools and techniques. Most of the literature on data mining focuses on its benefits and burdens in making business decisions.
4.2 Data Mining Concept:

Generally, data mining (sometimes called knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

(Fayyad, et al;1996:12) defines the data mining methods as tools for searching databases with special algorithms to identify general patterns which can be used in the classification of the individual observations and making predictions thereof. According to (Weiss&Indurkhya; 1998) data mining is the search for valuable information in large volumes of data. According to (Hand;1998:112) “data mining is the process of secondary analysis of large databases aimed at finding unsuspected relationships which are of interest or value to the database owners.” Data Mining has been referred as a statistical process of analyzing data stored in a warehouse. According to the Gartner Group, “Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques.

Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases (Peter,etal;1998:5).
Data mining is predicted to be one of the most revolutionary developments of the next decade (Rachel; 2001). Thus the entire process of applying a computer-based methodology, including new techniques, for discovering knowledge from data is called data mining.

Data mining is part of a group of concepts or techniques related to business intelligence, or e-business intelligence. Data mining involves obtaining information from a variety of sources that is stored in a data warehouse. This information becomes the input for various applications that uncover relationships and trends related to customers and processes. Online analytical processing (OLAP) allows a user to view data from many different angles to uncover correlations and relationships (Roberts-Witt; 2001, 5). These results are then used by managers and others to make better decisions. The emphasis is on data sharing where the web allows various types of information to be accessible to the masses. Managers, customers, suppliers and partners can ask the data warehouse questions about various aspects of the business through query and reporting applications. The figure (2) illustration below provides a graphic view of the data mining concept.
4.3 Purpose and Uses of Data Mining

The purpose of data mining is to identify patterns in order to make predictions from information contained in databases. It allows the user to be proactive in identifying and predicting trends with that information. Common uses of data mining in government include knowledge discovery, fraud detection, analysis of research, decision support, and website personalization. The most common government uses of data mining as identified by GAO include:

1. Improving service or performance
2. Detecting fraud, waste, and abuse
3. Analyzing scientific and research information
4- Managing human resources
5- Detecting criminal activities or patterns
6- Analyzing intelligence and detecting terrorist activities (GAO;2004).

State government data mining efforts include programs to ensure that the proper beneficiaries of state benefits programs receive the correct amount of benefits. Such uses can save states substantial amounts of money that otherwise would be erroneously paid out in the form of state benefits (NASACT;2004). Moreover, in a recent report, GAO found that twenty one states are using data mining software to look for unusual patterns in claims, provider, and beneficiary information stored in data warehouses in order to identify potential provider abuse.

The need to understand large complex information rich data sets is common to virtually all fields of business, science and engineering. In the business world, corporate and customer data are becoming recognized as a strategic asset. The ability to extract useful knowledge hidden in these data and to act on that knowledge is becoming increasingly important in today's competitive world.

Data mining is an iterative process within which progress is defined by discovery, through either automatic or manual methods. Data mining is most useful in an exploratory analysis in which there are no predetermined notions about what will constitute an interesting outcome. Data mining is the search for new, valuable, and nontrivial information in large volumes of data. It is a cooperative effort of humans and computers. Best results are achieved by balancing the knowledge of human experts in describing problems and goals with the search capabilities of computers.
Some other applications of Data Mining include: 1) Simulating and optimizing supply chain flows, reducing inventory and stock-outs, 2) Identifying customers with the greatest profit potential, 3) Identifying the price that will maximize yield or profit, 4) Selecting the best employees for tasks or jobs, 5) Detecting and minimizing quality problems, 6) Proving a better understanding of the drivers of financial performance including nonfinancial factors, 7) Improving the quality, efficacy and safety of products and services (Davenport;2006:105).

From the intelligence perspective, the National Research Council ranked data mining technology with antibiotics, vaccines, imaging and other technologies in the fight against terrorism. Text mining, video mining, audio phone mining and e-mail mining could all become important in the area of homeland defense (Roberts-Witt; 2002: 6).

Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among internal factors such as price, product positioning, or staff skills, and external factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data.

With data mining, a retailer could use point-of-sale records of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop products and promotions to appeal to specific customer segments.
There are four paradigms of science: 1) Theory, 2) Experimentation, 3) Computation and simulation, and 4) Data mining. The next scientific revolution involves using the fourth paradigm, deep-data-mining tools to solve the world’s problems in astronomy, oceanography, healthcare, water management, and climate change (Hey; 2010: 59).

(DM) or Knowledge Discovery in Databases (KDD) has been defined (Fayyad, et al; 1996) as the automatic discovery of previously unknown patterns or relationships in large and complex datasets. Most DM algorithms have been drawn from the areas of Statistics and Machine Learning adapted to induce knowledge from data contained within a database. The main objective of DM is to use the discovered knowledge for the purposes of explaining current behavior, predicting future outcomes, or providing support for business decision. The DM techniques used in business-oriented applications are also known as Business Intelligence (BI). BI is a general term to mean all processes, techniques, and tools that gather and analyze data for the purpose of supporting enterprise users to make better decisions (E. Awad & H. Ghaziri; 2004).

Therefore Data mining refers to using a variety of techniques to identify nuggets of information or decision-making knowledge in bodies of data, and extracting these in such a way that they can be put to use in the areas such as decision support, prediction, forecasting and estimation. The data is often voluminous, but as it stands of low value as no direct use can be made of it; it is the hidden information in the data that is useful.

**4.4 Classification of Data Mining Systems**

Data mining is an interdisciplinary field, the confluence of a set of disciplines, including database systems, statistics, machine learning,
visualization, and information science (J. Han & M. Kamber, 2006). Moreover, depending on the data mining approach used, techniques from other disciplines may be applied, such as neural networks, fuzzy and/or rough set theory, knowledge representation, inductive logic programming, or high-performance computing. Depending on the kinds of data to be mined or on the given data mining application, the data mining system may also integrate techniques from spatial data analysis, information retrieval, pattern recognition, image analysis, signal processing, computer graphics, Web technology, economics, business, bioinformatics, or psychology.

Because of the diversity of disciplines contributing to data mining, data mining research is expected to generate a large variety of data mining systems. Therefore, it is necessary to provide a clear classification of data mining systems, which may help potential users distinguish between such systems and identify those that best match their needs. Data mining systems can be categorized according to various criteria, as shown in Figure (3):

Figure (3) Data mining as a confluence of multiple disciplines
4.5 Data Mining Application:

Data mining aims to discover hidden knowledge, unknown patterns, and new rules from large databases that are potentially useful and ultimately understandable for making crucial decisions. It applies data analysis and knowledge discovery techniques under acceptable computational efficiency limitations, and produces a particular enumeration of patterns over the data (Fayyad, G. Piatetsky-Shapiro, & P. Smyth; 1996). The insights obtained via a higher level of understanding of data can help iteratively improve business practice. Nowadays, data mining software vendors are integrating fundamental data mining capabilities into database engines, so that users can execute data mining tasks in parallel inside the database, which reduces response time. Based on the type of knowledge that is mined, data mining can be mainly classified into the following categories (Han & Kamber; 2001):

1- Association rule mining uncovers interesting correlation patterns among a large set of data items by showing attribute-value conditions that occur together frequently. A typical example is market basket analysis, which analyzes purchasing habits of customers by finding associations between different items in customers’ shopping baskets.

2- Classification and prediction is the process of identifying a set of common features and models that describe and distinguish data classes or concepts. The models are used to predict the class of objects whose class label is unknown. A bank, for example, may classify a loan application as either a fraud or a potential business using models based on characteristics of the applicant. A large number of classification models have been developed for predicting future trends of stock market indices and foreign exchange rates.

3- Clustering analysis segments a large set of data into subsets or clusters. Each cluster is a collection of data objects that are similar to one
another within the same cluster but dissimilar to objects in other clusters. In other words, objects are clustered based on the principle of maximizing the intra-class similarity while minimizing the inter-class similarity. For example, clustering techniques can be used to identify stable dependencies for risk management and investment management.

4- Sequential pattern and time-series mining looks for patterns where one event (or value) leads to another later event (or value). One example is that after the inflation rate increases, the stock market is likely to go down.

The knowledge to be mined is closely related to a target application and the original data. Therefore, data mining should be considered along with several other issues rather than an isolated task: First, data mining needs to take ultimate applications into account. For example, credit card fraud detection and stock market prediction may require different data mining techniques. Second, data mining is dependent upon the features of data. For example, if the data are of time series, data mining techniques should reflect the features of time sequence.

Third, data mining should take advantage of domain models. In finance, there are many well-developed models that provide insight into attributes that are important for specific applications. Many applications combine data mining techniques with various finance and accounting models (e.g., Capital Asset Pricing Model).

4.6 Data Mining Tasks:

Data mining involves six common classes of tasks (fayyad, et al; 1996)

1- Anomaly detection: (Outlier/change/deviation detection) – The identification of unusual data records, that might be interesting or data errors and require further investigation.
2- Association rule learning (Dependency modeling) – Searches for relationships between variables. For example, a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.

3- Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.

4- Classification – is the task of generalizing known structure to apply to new data. Regression – Attempts to find a function which models the data with the least error.

5- Summarization – providing a more compact representation of the data set, including visualization and report generation.

In practice, the two primary goals of data mining tend to be prediction and description. Prediction involves using some variables or fields in the data set to predict unknown or future values of other variables of interest. Description, on the other hand, focuses on finding patterns describing the data that can be interpreted by humans. Therefore, it is possible to put data-mining activities into one of two categories:

1) Predictive data mining, which produces the model of the system described by the given data set, or

2) Descriptive data mining, which produces new, nontrivial information based on the available data set.

On the predictive, the goal of data mining is to produce a model, expressed as an executable code, which can be used to perform classification, prediction, estimation, or other similar tasks. On the other,
The goal is to gain an understanding of the analyzed system by uncovering patterns and relationships in large data sets. The relative importance of prediction and description for particular data-mining applications can vary considerably. So the goals of prediction and description are achieved by using data-mining techniques.

4.7 Data Mining Techniques

Among a variety of data mining techniques that have been used in finance, we mainly focus on introducing five commonly used techniques, namely neural networks, genetic algorithms, statistical inference, rule induction, and data visualization:

1. **Neural Networks**:

   Artificial neural networks are computer models built to emulate the human pattern recognition function through a similar parallel processing structure of multiple inputs. A neural network consists of a set of fundamental processing elements (also called neurons) that are distributed in a few hierarchical layers. Most neural networks contain three types of layers: input, hidden, and output. The figure (4) shows that:

   Figure (4): Neural connections & Neural network layers.

Adapted from DR. YASHPAL SINGH, ALOK SINGH CHAUHAN, 2009, p.38 Journal of Theoretical and Applied Information Technology, India.
After each neuron in a hidden layer receives the inputs from all of the neurons in a layer ahead of it (typically an input layer), the values are added through applied weights and converted to an output value by an activation function (e.g., the Sigmoid function). Then, the output is passed to all of the neurons in the next layer, providing a feed forward path to the output layer. The weights between two neurons in two adjacent layers are adjusted through an iterative training process while training samples are presented to the network. They are used to store captured knowledge and make it available for future use. Characterized by the pattern of connections between neurons, the method of determining weights on the connections, and the node activation function, a neural network is designed to capture causal relationships between dependent and independent variables in a given data set. Neural networks offer a class of tools that can approximate financial patterns to a satisfactory degree of accuracy.

Although artificial neural networks could include a wide variety of types, the most commonly used are feed-forward error back-propagation type neural networks. In these networks, the individual elements (“neurons”) are organized into layers in such a way that output signals from the neurons of a given layer are passed to all of the neurons of the next layer. Thus, the flow of neural activations goes in one direction only, layer-by-layer (feed-forward). Errors made by the neural network are then used to adjust all the network weights by moving back through the network (error back-propagation). The smallest number of layers is two, namely the input and output layers. More layers, called hidden layers, could be added between the input and the output layer. The function of the hidden layers is to increase the computational power of the neural nets. Artificial Neural Network (ANN) have been proven to be universal approximators assuming that sufficient hidden layer neurons are provided and assuming
that the activation function is bounded and non-constant (Ian H. Witten & etal; 2011).

Neural networks weights are tuned to fulfill a required mapping of inputs to the outputs using training algorithms. The common training algorithm for the feed-forward nets is called “error back-propagation”. This is a supervised type of training, where the desired outputs are provided to the ANN during the course of training along with the inputs. The provided input-output couplings are known as training pairs, and the set of given training pairs is called the training set.

2. **Genetic Algorithms**:  
The basic idea of genetic algorithms is that given a problem, the genetic pool of a specific population potentially contains the solution, or a better solution. Based on genetic and evolutionary principles, the genetic algorithm repeatedly modifies a population of artificial structures through the application of initialization, selection, crossover, and mutation operators in order to obtain an evolved solution. It starts with a set of solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population. This is motivated by a hope that the new population will be better than the old one.

3. **Statistical Inference**:  
Statistics provides a solid theoretical foundation for the problem of data analysis. Through hypothesis validation and/or exploratory data analysis, statistical techniques give asymptotic results that can be used to describe the likelihood in large samples. The basic statistical exploratory methods include such techniques as examining distribution of variables, reviewing large correlation matrices for coefficients that meet certain thresholds, and examining multidimensional frequency tables. Multivariate
exploratory techniques designed specifically for identifying patterns in multivariate data sets include cluster analysis, factor analysis, discriminant function analysis, multidimensional scaling, log-linear analysis, canonical correlation, stepwise linear and nonlinear regression, time-series analysis, and classification trees. Among them, discriminant analysis, factor analysis, principle component analysis, and regression models have been frequently used to identify either influential variables in financial problems or relationships between different variables and financial markets.

4. **Rule Induction:**
Rule induction models belong to the logical, pattern distillation based approaches of data mining. Based on data sets, these techniques produce a set of if-then rules to represent significant patterns and create prediction models. Such models are fully transparent and provide complete explanations of their predictions. One commonly used and well-known type of rule induction is the family of algorithms that produce decision trees.

A decision tree, which is usually constructed using a training data set, consists of hierarchically organized sets of rules. It is a simple recursive structure for representing a decision procedure in which a new instance is classified into one of the predefined classes. In decision trees, instances are represented as feature vectors containing a list of attribute value pairs. Each internal node represents a decision attribute-value test. Each branch represents an outcome of the test, and each leaf node denotes a decision class.

5. **Data Visualization**
Data are difficult to interpret due to its overwhelming size and complexity. In order to achieve effective data mining, it is important to include people in the data exploration process and combine the
flexibility, creativity, and general knowledge of people with the enormous storage capacity and computational power of today’s computers. Data visualization is the process of analyzing and converting data into graphics, thus taking advantage of human visual systems. This technique allows decision makers and analysts to gain insight into the data, draw conclusions, and directly interact with the data. It is proven to be of high value in exploratory data analysis, especially useful when little is known about the data and exploration goals are vague (Dwinnell; 2002, 56).

4.8 Data Mining And Knowledge Discovery Processes

Data mining is the search for relationships and global patterns that exist in large databases but are hidden among the vast amount of data, such as a relationship between patient data and their medical diagnosis. These relationships represent valuable knowledge about the database and the objects in the database and, if the database is a faithful mirror, of the real world registered by the database.

Basically data mining is concerned with the analysis of data and the use of software techniques for finding patterns and regularities in sets of data. It is the computer which is responsible for finding the patterns by identifying the underlying rules and features in the data. The idea is that it is possible to strike gold in unexpected places as the data mining software extracts patterns not previously discernable or so obvious that no-one has noticed them before.

Data mining analysis tends to work from the data up and the best techniques are those developed with an orientation towards large volumes of data, making use of as much of the collected data as possible to arrive at reliable conclusions and decisions. The analysis process starts with a set of data, uses a methodology to develop an optimal
representation of the structure of the data during which time knowledge is acquired. Once knowledge has been acquired this can be extended to larger sets of data working on the assumption that the larger data set has a structure similar to the sample data. Again this is analogous to a mining operation where large amounts of low grade materials are sifted through in order to find something of value. The following diagram summarizes the some of the stages/processes identified in data mining and knowledge discovery by (Fayyad & Evangelos Simoudis), two of leading exponents of this area, as shown in figure (5) below:

Figure (5): Data mining in KDD

adapted from Fayyad, G. Piatetsky-Shapiro,& P. Smyth;1996
The phases depicted start with the raw data and finish with the extracted knowledge which was acquired as a result of the following stages:

**Selection**

Selecting or segmenting the data according to some criteria e.g. all those people who own a car, in this way subsets of the data can be determined.

**Pre-processing**

Before data mining algorithms can be used, a target data set must be assembled. As data mining can only uncover patterns actually present in the data, the target dataset must be large enough to contain these patterns while remaining concise enough to be mined within an acceptable time limit. A common source for data is a data mart or data warehouse. Pre-processing is essential to analyze the multivariate datasets before data mining. The target set is then cleaned. Data cleaning removes the observations containing noise and those with missing data.

**Transformation**

The data is not merely transferred across but transformed in that overlays may added such as the demographic overlays commonly used in market research. The data is made useable and navigable.

**Data mining**

The final step of knowledge discovery from data is to verify that the patterns produced by the data mining algorithms occur in the wider data set. Not all patterns found by the data mining algorithms are necessarily valid. It is common for the data mining algorithms to find patterns in the training set which are not present in the general data set. This is called
over fitting. To overcome this, the evaluation uses a test set of data on which the data mining algorithm was not trained. The learned patterns are applied to this test set and the resulting output is compared to the desired output. This stage is concerned with the extraction of patterns from the data. A pattern can be defined as given a set of facts.

**Interpretation and evaluation**

The patterns identified by the system are interpreted into knowledge which can then be used to support human decision-making e.g. prediction and classification tasks, summarizing the contents of a database or explaining observed phenomena. If the learned patterns do not meet the desired standards, then it is necessary to re-evaluate and change the pre-processing and data mining steps. If the learned patterns do meet the desired standards, then the final step is to interpret the learned patterns and turn them into knowledge. So Competitive advantage requires abilities. Abilities are built through knowledge. Knowledge comes from data. The process of extracting knowledge from data is called Data Mining.

### 4.9 Data Mining in Finance

Specifics of data mining in finance are coming from the need to:

- Forecast multidimensional time series with high level of noise;
- Accommodate specific efficiency criteria (e.g., the maximum of trading profit) in addition to prediction accuracy make coordinated multire solution forecast (minutes, days, weeks, months, and years);
- Incorporate a stream of text signals as input data for forecasting models.
- Be able to explain the forecast and the forecasting model (“black box” models have limited interest and future for significant investment decisions);
- Be able to benefit from very subtle patterns with a short life time; and incorporate the impact of market players on market regularities.

Data mining creates tools which can be useful for discovering subtle short-term conditional patterns and trends in wide range of financial data. Forecasting stock market, currency exchange rate, bank bankruptcies, understanding and managing financial risk, trading futures, credit rating, loan management, bank customer profiling, and money laundering analyses are core financial tasks for data mining (Nakhaeizadeh et. al.; 2002). Its own niche in financial modeling. Similarly to other computational methods almost every data mining method and technique has been used in financial modeling. An incomplete list includes a variety of linear and non-linear models, multi-layer neural networks (Kingdon; 1997), k-means and hierarchical clustering; k-nearest neighbors, decision tree analysis, regression (logistic regression; general multiple regression), ARIMA, principal component analysis, and Bayesian learning. Less traditional methods used include rough sets (Shen & Loh; 2004), relational data mining methods (deterministic inductive logic programming and newer probabilistic methods (Muggleton; 2002), support vector machine, independent component analysis, Markov models and hidden Markov models.

Bootstrapping and other evaluation techniques have been extensively used for improving data mining results. Specifics of financial time series analyses with ARIMA, neural networks, relational methods, support vector machines and traditional technical analysis is discussed in (Murphy; 1999). In fact, the only realistic approach proven to be successful is providing comparisons between different methods showing
their strengths and weaknesses relative to problem characteristics conceptually and leaving for user the selection of the method that likely fits the specific user problem circumstances. In essence this means clear understanding that data mining in general, and in finance specifically, is still more art than hard science. For instance, understanding the power of first-order If-Then rules over the decision trees can significantly change and improve data mining design. User’s actual experiments with data provide a real judgment of data mining success in finance. In comparison with other fields such as geology or medicine, where test of the forecast is expensive, difficult, and even dangerous, a trading forecast can be tested next day in essence without cost and capital risk involved in real trading.

Attribute-based learning methods such as neural networks and decision trees dominate in financial applications of data mining. These methods are relatively simple, efficient, and can handle noisy data. However, these methods have two serious drawbacks: a limited ability to represent background knowledge and the lack of complex relations. Relational data mining techniques that include Inductive Logic Programming (ILP) (Muggleton; 1999) intend to overcome these limitations. Previously these methods have been relatively computationally inefficient and had rather limited facilities for handling numerical data (Bratko&Muggleton; 1995). Currently these methods are enhanced in both aspects and are especially actively used in bioinformatics.

I believe that now is the time for applying these methods to financial analyses more intensively especially to those analyses that deal with probabilistic relational reasoning. Various publications have estimated the use of data mining methods like hybrid architectures of neural networks with genetic algorithms, chaos theory, and fuzzy logic in finance.
4.10 Data Mining Technology and Decision Support System:

With the increase of economic globalization and evolution of information technology, financial data are being generated and accumulated at an unprecedented rate. It is used to keep track of companies’ business performance, monitor market changes, and support financial decision-making. Nonetheless, the rapidly growing volume of data has far exceeded our ability to analyze them manually.

There is a critical need for automated approaches to effective and efficient utilization of massive financial data to support companies and individuals in strategic planning and investment decision-making. Data mining is able to uncover hidden patterns and predict future trends and behaviors in financial markets. It creates opportunities for companies to make proactive and knowledge-driven decisions in order to gain a competitive advantage. Data mining has been applied to a number of financial applications, including development of trading models, investment selection, loan assessment, portfolio optimization, fraud detection, bankruptcy prediction, real-estate assessment, and so on. The competitive advantages achieved by data mining include increased revenue, reduced cost, and much improved marketplace responsiveness and awareness.

Rapid advances in information and technologies along with the availability of large-scale scientific and business data repositories or database management technologies, combined with breakthroughs in computing technologies, computational methods and processing speeds, have opened the floodgates to data dictated models and pattern matching (Fayyad & Uthurusamy, 2002:29). The use of sophisticated and computationally intensive analytical methods are expected to become even more commonplace with recent research breakthroughs in
computational methods and their commercialization by leading vendors (Grossman et al., 2002:60).

Scientists and engineers have developed innovative methodologies for extracting correlations and associations, dimensionality reduction, clustering or classification, regression and predictive modeling, tools based on expert systems and case based reasoning, as well as decision support systems for batch or real-time analysis. They have utilized tools from areas like traditional statistics, signal processing and artificial intelligence as well as emerging fields like data mining, machine learning, operations research, systems analysis and nonlinear dynamics (Ganguly & et al.; 2002a).

Innovative models and newly discovered patterns in complex, nonlinear and stochastic systems, encompassing the natural and human environments, have demonstrated the effectiveness of these approaches. However, applications that can utilize these tools in the context of scientific databases in a scalable fashion have only begun to emerge (Ganguly & et al.; 2002b:16).

Business solution providers and IT vendors, on the other hand, have focused primarily on scalability, process automation and workflows, and the ability to combine results from relatively simple analytics with judgments from human experts, for example e-business applications. in the areas of supply chain planning, financial analysis and business forecasting, traditionally rely on decision support systems with embedded data mining, operations research and OLAP technologies, business intelligence (BI) and reporting tools as well as an easy to use GUI (graphical user interface like the researcher use it in the practical side) and extensible business workflows (Geoffrion & Krishnan; 2003). These applications can be custom built by utilizing software tools, or available as prepackaged e-business application suites from large vendors like SAP.
and Oracle (that oil companies in India uses). For a scientist or an engineer, as well as for a business manager or management scientist, DMT and DSS are tools used for developing domain-specific applications.

These applications might combine knowledge about the specific scientific or business domain with data dictated or decision making tools like DSS and DMT. Within the context of these applications, DMT and DSS can aid in the discovery of novel patterns, development of predictive or descriptive models, mitigation of natural or man-made hazards, preservation of civil societies and infrastructures, improvement in the quality and span of life as well as in economic prosperity and well being, and development of natural and built environments in a sustainable fashion. Disparate applications utilizing DMT and DSS tools tend to have interesting similarities.

Business forecasting, planning and decision support applications (Shim & et al., 2002, 117) usually need to read data from a variety of sources like on-line transactional processing (OLTP) systems, historical data warehouses and data marts, syndicated data vendors, legacy systems, or public domain sources like the Internet, as well as in the form of real-time or incremental data entry from external or internal collaborators, expert consultants, planners, decision makers or executives. Data from disparate sources are usually mapped to a predefined common data model, and incorporated through extraction transformation and loading (ETL) tools. End users are provided GUI based access to define application contexts and settings, structure business workflows and planning cycles and format data models for visualization, judgmental updates or analytical and predictive modeling. The parameters of the embedded data mining models might be preset, calculated dynamically based on data or user inputs, or specified by a power user. The results of
the data mining models can be automatically utilized for optimization and recommendation systems and/or can be used to serve as baselines for planners and decision makers.

The power of information technologies has been utilized to acquire, manage, store, retrieve and represent data in information repositories, and to share, report, process, collaborate on and move data in scientific and business applications.

Database management and data warehousing technologies have matured significantly over the years. Tools for building custom and packaged applications, including but not limited to workflow technologies, web servers and GUI-based data entry and viewing forms, are steadily maturing. There is a clear and present need to exploit the available data and technologies to develop the next generation of scientific and business applications, which can combine data-dictated methods with domain specific knowledge. Analytical information technologies, which include DMT and DSS, are particularly suited for these tasks. These technologies can facilitate both automated (data-dictated) and human expert driven knowledge discovery and predictive analytics, and can also be made to utilize the results of models and simulations that are based on process business insights. If DMT and DSS were to be defined broadly, a broad statement can perhaps be made that while business applications have scalable but straightforward DMT embedded within DSS, scientific applications have utilized advanced DMT, but focused less on scalability and DSS. Multidisciplinary research and development efforts are needed in the future for maximal utilization of analytical information technologies in the context of these applications.
4.11 Data Mining Technology In The Financial Decision Support System

The current data mining technology in the Financial Decision Support System are
(Ma Xiao Hu. & Li Gaojin, 2010; 383):

1. Financial Analysis

Financial analysis is an important component of financial management, including enterprise solvency analysis, capacity analysis of business operations, corporate profitability analysis, and corporate development capacity analysis. Financial Analysis System can use data mining classification techniques, forecasting techniques, according to companies in the past, the present financial data for further processing, collation, analysis and evaluation, in predicting future financial position at the same time from which to obtain useful information for policy makers.

2. Financial Forecast

Financial forecast system is an important part of FDSS, whose function is divided into two aspects: on the one hand, to make use of existing financial data to forecast the company's future financial condition and successful operating; On the other hand, to make use of expert experience and expertise to forecast a financial topic. The main contents of financial forecast include sales forecasts, profit forecasts, cost forecasts, financial forecasts, financial indicators forecasts, and so on. We can forecast the corporate future financial performance according to established corporate financial data through the technique of regression and neural network, and thus judge the possibilities of the corporate occurring financial crisis in the future.
3. Funding Decision
Funding is the process about a corporate when, how, access to what the scale of funding. Corporate financing decisions include the number of fund-raising decisions, financing decisions and debt repayment decision. Generally speaking, companies should first consider funding equity financing, that is, equity financing; and then consider debt financing, which aims to minimize financial risk. We can make use of the technique of classification and clustering in data mining to determine reasonable financial programs according to the needs of decision-making information output about enterprise management.

4. Investment Decisions
Corporate investment decisions include long-term investment decision-making within the enterprise, joint venture investment decisions and portfolio investment decisions. The problem of investment decision-making is the more complex issues of that corporate management, its decision-making problems are generally divided into semi structured or unstructured problems. We can make use of the technique of forecast and Relevance to confirm the projection of investment for the investment opportunities, investment scale, and investment modes. We can achieve maximum efficiency of funds according to the system output of the most valuable projection decision information which is selected in a number of projects. Users DM tool OLAP tool Model approach libraries and their management system Figure1 the structure of the financial decision support system Data warehouse Database management system Knowledge libraries and their management system Assistant decision information Human computer interaction of problem-solving, analysis for interactive system
5. Cost Decision
Costs decision involved in various fields of the corporate decision-making for marketing, production, operation and capital operation. There is the cost of decision-making in various economic activities which lead to costs and expense.
Corporate cost decision include: inventory cost decision, production costs decision, the cost of capital decision, cost of sales decision and service costs of the decision, etc. In which the cost of sales and service costs of the decision-making have many unstructured factors, including promotional costs, advertising costs, sales services fees. The degree of these unstructured costs decision increase, which make it more complicated to determine the decision project. These require the use of data mining in the technique of time series analysis, correlation analysis and other techniques to analyze and forecast according to historical data, and analysis and forecasting of sales and service cost analysis.

6. Dividend Allocation Decisions
Dividend distribution is that the company distributes dividends to its shareholders. Whether dividend decisions are reasonable or not, will have a significant impact on the company's development and the interests of shareholders. Dividend allocation decisions including dividend repayment policy, dividend payout ratio and dividend payout forms of decision-making and so on. As the dividend distribution policy is subject to legal, investment, financing and surplus stability, the interests of shareholders, as well as the stock market, the impact of external factors, many of whose decisions are semi-structured and unstructured problems, we can make use of classification technique in data mining techniques to provide support.
7. Inventory Decision

Inventory decision-making mainly refers to the decision making materials and finished products decision, namely, to determine a rational economic order quantity and when the order is the most opportune time, and strive to make the stock on the total cost spent to reach a minimum. Uncertainty of Sales makes inventory decision-making to become a risk decision-making. It needs past experience save, the analysis of historical statistical data and research data of user input, using decision tree method in data mining techniques to help decision makers to determine the scope of demand variable and their probability, and provide the best reference data.

As can be seen from the above review, there is a multitude of forecasting algorithms available, which offer a range of settings that can permit tuning of the technique. This leads to a general problem of finding the best forecasting technique and the best set of parameters for the chosen technique that result in the most accurate forecast. This process is difficult because the more trial and error executed, the higher the probability of finding a solution that does very well on the given data, but that does not forecast well into the future. Even with the use of a testing set, the same problem can occur if performance on the testing set is evaluated many times. The more choices we have in forecasting techniques and parameters, the more problematic this situation becomes. The complexity of this problem can be reduced by relying on a class of algorithms that are called “universal approximators”, which can, by definition, approximate any function, to an arbitrary accuracy. Using such universal approximators, any required function between past and future data can be learned, thus making exiting forecasting techniques a subset of the functions that the universal approximator can learn is artificial intelligence techniques, such as Artificial Neural Networks.
(ANN) are universal approximators and can be used to learn any function. Of specific interest to fourth hypothesis of my study are Artificial Neural Networks, ANN are frequently used to predict time series (Giles & et al; 2001). Because ANN perform back-propagation of error through time that permits the Neural Network to learn patterns through to an arbitrary depth in the time series. This means that even though a time window of data is provided as the input dimension to the ANN, it can match pattern through time that extends further than the provided current time window, because it has recurrent connections.