Adapted Framework for Data Mining Technique to Improve Decision Support System in an Uncertain Situation

Ahmed Bahgat El Seddawy[1], Dr. Ayman Khedr[2] and Prof. Dr. Turky Sultan[3]

[1]Department of Business Information System, Arab Academy for Science and Technology, Egypt
AhmedBahgat2@hotmail.com
[2]Department of Information system ,Helwan University, Egypt
Aymankhedr@gmail.com
[3]Department of Information System, Helwan University, Egypt
TurkySultan@hotmail.com

Abstract

Decision Support System (DSS) is equivalent synonym as management information systems (MIS). Most of imported data are being used in solutions like data mining (DM). Decision supporting systems include also decisions made upon individual data from external sources, management feeling, and various other data sources not included in business intelligence. Successfully supporting managerial decision-making is critically dependent upon the availability of integrated, high quality information organized and presented in a timely and easily understood manner. Data mining have emerged to meet this need. They serve as an integrated repository for internal and external data-intelligence critical to understanding and evaluating the business within its environmental context. With the addition of models, analytic tools, and user interfaces, they have the potential to provide actionable information that supports effective problem and opportunity identification, critical decision-making, and strategy formulation, implementation, and evaluation. The proposed system will support top level management to make a good decision in any time under any uncertain environment.

Key words:

DSS, DM, MIS, CLUSTERING, CLASSIFICATION, ASSOCIATION RULE, K-MEAN, OLAP, MATLAB

1. INTRODUCTION

Decision Support System (DSS) is equivalent synonym as management information systems (MIS). Most of imported data are being used in solutions like data mining (DM). Decision supporting systems include also decisions made upon individual data from external sources, management feeling, and various other data sources not included in business intelligence. Successfully supporting managerial decision-making is critically dependent upon the availability of integrated, high quality information organized and presented in a timely and easily understood

DOI : 10.5121/ijdkp.2012.2301

manner. Data mining have emerged to meet this need. They serve as an integrated repository for internal and external data-intelligence critical to understanding and evaluating the business within its environmental context. With the addition of models, analytic tools, and user interfaces, they have the potential to provide actionable information that supports effective problem and opportunity identification, critical decision-making, and strategy formulation, implementation, and evaluation. The proposed system will support top level management to make a good decision in any time under any uncertain environment [4]. This study aim to investigate the adoption process of decision making under uncertain situations or highly risk environments effecting in decision of investing stoke cash of bank. This applied for two types of usage investment - direct or indirect - or credit and any sector of investment will be highly or moderate or low risk. And select which one of this sectors risk 'rejected' or un-risk 'accepted' all that under uncertain environments such as; political, economical, marketing, operational, internal policies and natural crises, all that using the contribution of this study enhancing k-mean algorithm to improve the results and comparing results between original algorithm and enhanced algorithm. The paper is divided into four sections; section two is a background and related work it is divided into two parts, part one is about DSS, part two is about DM. Section three presents the proposed Investing Data Mining System 'IDMS. Section four presents conclusion and finally section five present future works2. Tables, Figures and Equations

2. BACKGROUND AND RELATED WORK

2.1 Decision Support System (DSS)

DSS includes a body of knowledge that describes some aspects of the decision maker's world that specifies how to accomplish various tasks, that indicates what conclusions are valid in different circumstances [4].The expected benefits of DSS that discovered are higher decision quality, improved communication, cost reduction, increased productivity, time savings, improved customer satisfaction and improved employee satisfaction. DSS is a computer-based system consisting of three main interacting components:

- **A language system**: a mechanism to provide communication between the user and other components of the DSS.
- **A knowledge system**: A repository of problem domain knowledge embodied in DSS as either data or procedures.
- **A problem processing system**: a link between the other two components, containing one or more of the general problem manipulation capabilities required for decision-making.

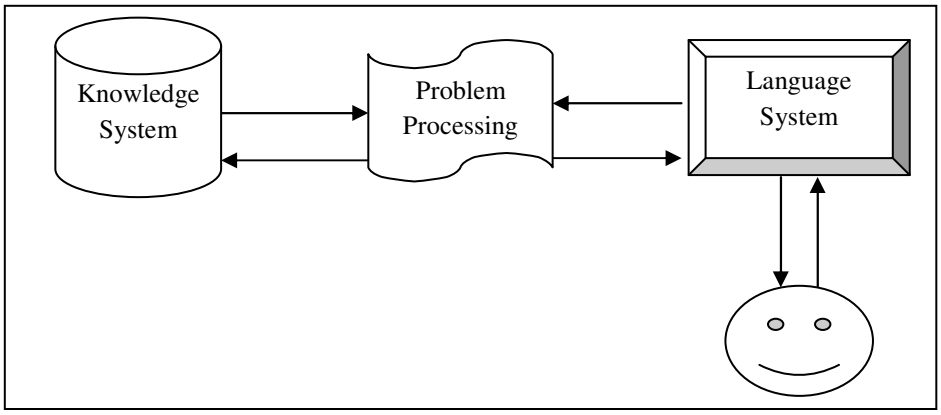

Fig 1: DSS Main Components

After surveying multiple decision support systems, it is concluded that decision support systems are categorized into the following [5]:

- **File drawer systems**: This category of DSS provides access to data items.
- **Data analysis systems**: Those support the manipulation of data by computerized tools tailored to a specific task or by more general tools and operators.
- **Analytical information systems**: Those provide access to a series of decision-oriented databases.
- **Accounting and financial models**: those calculate the consequences of possible actions.
- **Representational models**: those estimate the consequences of actions based on simulation models that include relationships that are causal as well as accounting definitions.
- **Optimization models**: those provide guidelines for actions by generating an optimal solution consistent with a series of constraints.
- **Suggestion models**: those perform the logical processing leading to a specific suggested decision or a fairly structured or well understood task.

This section describes the approaches and techniques mostly used when developing data warehousing systems that data warehousing approaches such as; Online Analytical Processing 'OLAP', Data Mining 'DM' and Artifial Intelligence 'AI'. Data Warehousing Techniques such as; Star Schema, Snowflake Schema and A Star Schema or a Snowflake Schema

2.2 Data Mining Techniques (DM)

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information [10]. DM techniques are the result of a long process of research and product development [10]. The evolution of DM [6] is shown in table 1.

Table 1: The evolution by DM [6]

Evolutionary Step	Business Question	Enabling Technologies	Product Providers
Data Collection (1960s)	"What was my total revenue in the last five years?"	Computers, tapes, disks	IBM, CDC
Data Access (1980s)	"What were unit sales in New England last March?"	Relational databases (RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft
Data Warehousing & Decision Support (1990s)	"What were unit sales in New England last March? Drill down to Boston."	On-line analytic processing (OLAP), multidimensional databases, data warehouses	Pilot, Comshare, Arbor, Cognos, Microstrategy
Data Mining (Emerging Today)	"What's likely to happen to Boston unit sales next month? Why?"	Advanced algorithms, multiprocessor computers, massive databases	Pilot, Lockheed, IBM, SGI, numerous startups (nascent industry)

There are several processes for applying DM:

1. Definition of the business objective and expected operational environment.
2. Data selection is required to identify meaningful sample of data.
3. Data transformation that involves data representation in an appropriate format for mining algorithm.

4. Selection and implementation of data mining algorithm depends on the mining objective.
5. Analysis of the discovered outcomes is needed to formulate business outcomes.
6. Representing valuable business outcomes.

Data mining consists of five major elements; to extract, to transform, and to load transaction data onto the data warehouse system, to store and manage the data in a multidimensional database system, to provide data access to business analysts and information technology professionals, Analyze the data by application software, and finally to present the data in a useful format, such as a graph or table.

DM techniques usually fall into two categories, predictive or descriptive. Predictive DM uses historical data to infer something about future events. Predictive mining tasks use data to build a model to make predictions on unseen future events. Descriptive DM aims to find patterns in the data that provide some information about internal hidden relationships.

Descriptive mining tasks characterize the general properties of the data and represent it in a meaningful way. Figure2 shows the classification of DM techniques.

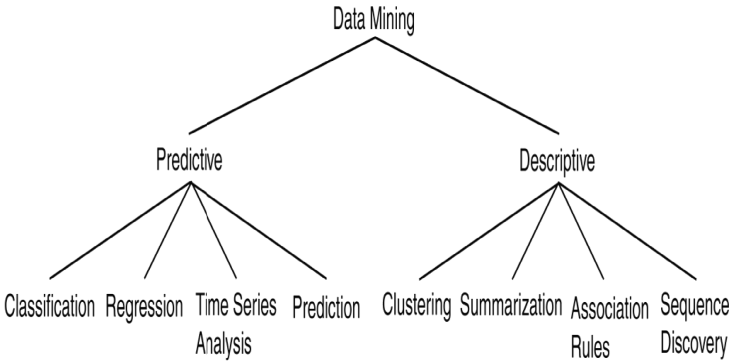

Fig 2: DM Techniques [5]

Association Rule is used to discover relationships between attribute sets for a given input pattern. [6] Define sequence discovery as "a sequential technique is a given set of sequences find the complete set of frequent subsequences". Clustering is "the process of organizing objects into groups whose members are alike in some way" [7]. A cluster is therefore a collection of objects which are "similar" among them and are "dissimilar" to the objects belonging to another cluster. So, it deals with finding the internal structure in a collection of data, as shown in figure 3.

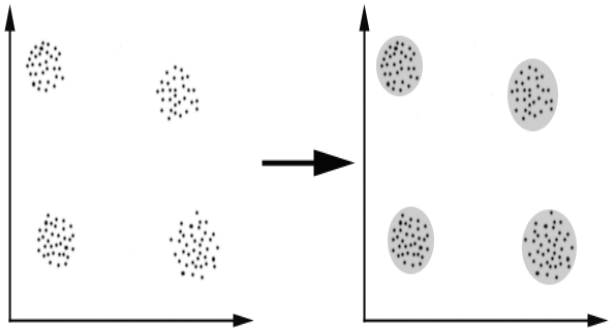

Fig 3: Simple graphical for clustering data [7]

[8] Define that “Clustering involves identifying a finite set of categories or segments ‘clusters’ to describe the data according to a certain metric". [9] Define that “Clustering enables to find specific discriminative factors or attributes for the studied data. Each member of a cluster should be very similar to other members in its cluster and very dissimilar to other clusters. When a new data is introduced, it is classified into the most similar cluster". Several researchers classified clustering algorithms differently.

Some classifies clusters as mutually exclusive, hierarchical or overlapping. Others classifies cluster into hierarchal and partitional. The most common classification is shown in figure 3. Techniques for creating clusters include partitioning methods as in k-means algorithm, and hierarchical methods as in decision trees, and density-based methods.

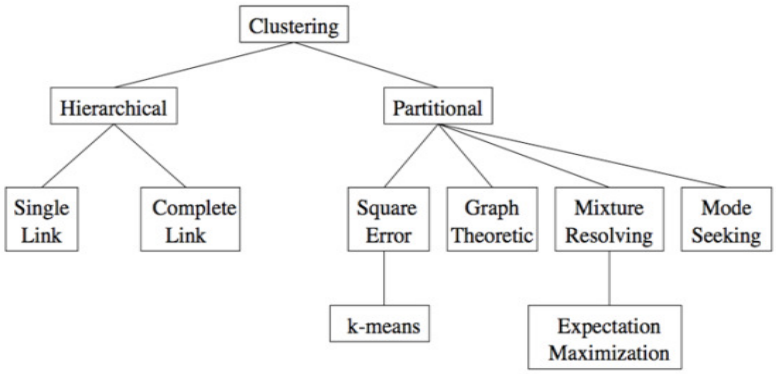

Fig 4: Clustering methods classifications for Moses Charikar [10]

3. THE PROPOSED 'IDMS'

Investment Data Mining System ‘IDMS’ aims to build a data mining system for investment in the banking sector. IDMS consists of several components; data gathering, preparing data to discover knowledge, data preprocessing, using data mining techniques in sequences steps start with

classification data, clustering data especially using K-mean algorithm and enhanced K-mean algorithm to set which best result and then set and run association rules to solve problem, post processing and finally get result and visualize result to create best decision to take a good decision for investment under uncertain situations. IDMS Shown in figure 5.

Hardware for applying the IDMC system is a personal computer configurations with this Processor 3.2, Hard Disk 160 gaga, Ram 2 G and Monitor 17 Inch. Operating system is windows XP services pack 3. Several software tools have been used. The first is Microsoft Excel sheets 2007 and has been used for analysis and filtering data. Mat-lab version 6.5 has been used in data preprocessing and data classification. The last software is the WEKA which is a collection of java tools for DM written by staff at the University Of Waikato, New Zealand.

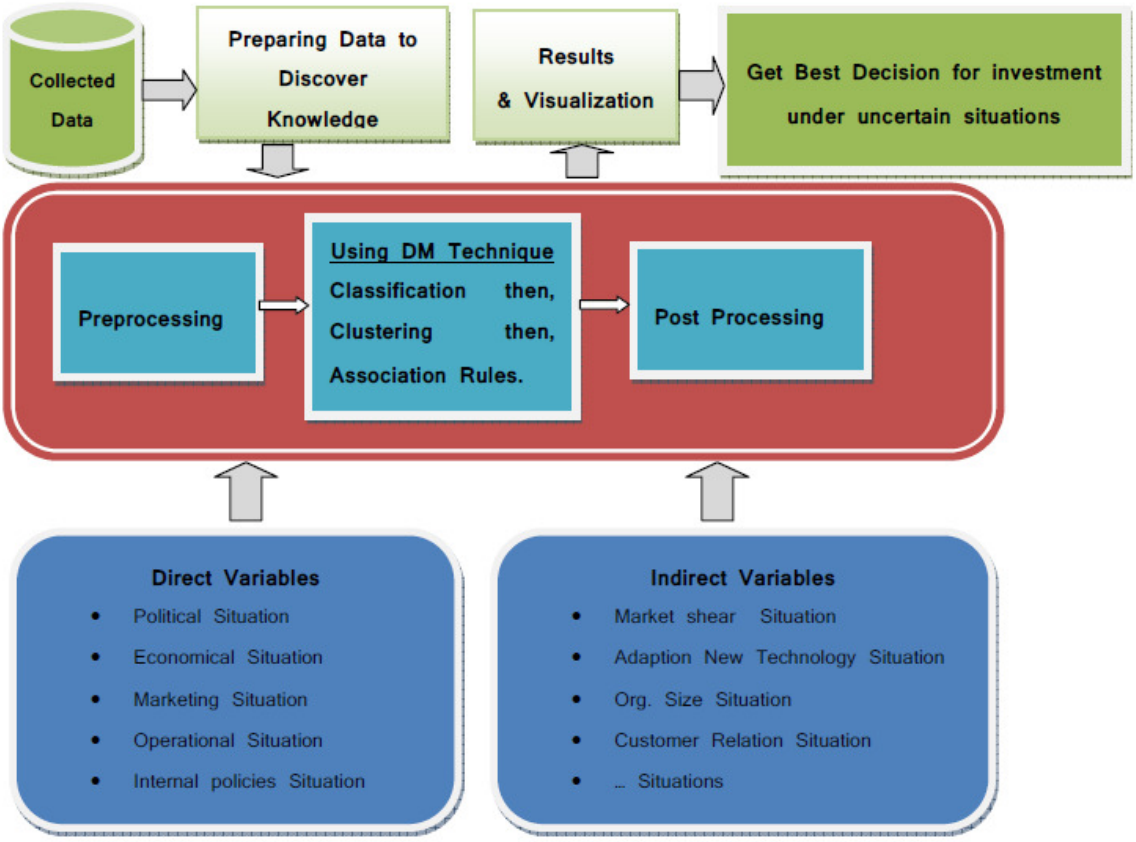

Fig 5: Proposed IDMS

4. CONCLUSIONS

This paper represents a contribution of applying DM and DSS for banking sector especially in investment which has been rarely addressed before. IDMS is a new proposed system which is simple, straightforward with low computation needs. The proposed preprocessing component is an aggregation of several known steps. The post processing component is an optional one that eases the interpretation of the investment results. The banking is planning a set of actions in accordance of IDMS outcomes for decision making in investment sector. The investment

department in the banking is starting to analyze the approached investment sector, to introduce a good decision under uncertain situation.

5. FUTURE WORK

In next step of this study implementing this proposed approach and creating full steps of preprocessing using enhanced k-mean algorithm by real data from bank as a case study, to give us a best result and support high level of management with a good decision. After implementing this case study and proposed system will get the main factors which effect in decision under uncertain

ACKNOWLEDGMENT

I want to express my deepest gratitude for my professor's supervisors Prof. Dr. Turky Sultan and Dr. Ayman Khedr for their major help and support through all the phases of research and development.

REFERENCES

[1] A. Hunter and S. Parsons, "A review of uncertainty handling formalisms", Applications of Uncertainty Formalisms, LNAI 1455, pp.8-37. Springer -Verlag, 1998.

[2] E. Hernandez and J. Recasens, "A general framework for induction of decision trees under uncertainty", Modelling with Words, LNAI 2873, pp.26–43, Springer-Verlag, 2003.

[3] M. S. Chen, J. Han, and P. S. Yu. IEEE Trans Knowledge and Data Engineering Data mining. An overview from a database perspective, 8:866-883, 1996.

[4] U. Fayyad, G. Piatetsky-Shapiro and W. J. Frawley. AAAI/MIT, Press definition of KDD at KDD96. Knowledge Discovery in Databases, 1991.

[5] Gartner. Evolution of data mining, Gartner Group Advanced Technologies and Applications Research Note, 2/1/95.

[6] International Conferences on Knowledge Discovery in Databases and Data Mining (KDD'95-98), 1995-1998.

[7] R.J. Miller and Y. Yang. Association rules over interval data. SIGMOD'97, 452-461, Tucson, Arizona, 1997.

[8] Zaki, M.J., SPADE An Efficient Algorithm for Mining Frequent Sequences Machine Learning, 42(1) 31-60, 2001.

[9] Osmar R. Zaïane. "Principles of Knowledge Discovery in Databases - Chapter 8 Data Clustering". & Shantanu Godbole data mining Data mining Workshop 9th November 2003.

[10] T.Imielinski and H. Mannila. Communications of ACM. A database perspective on knowledge discovery, 39:58-64, 1996.

[11] BIRCH Zhang, T., Ramakrishnan, R., and Livny, M. SIGMOD '96. BIRCH an efficient data clustering method for very large databases. 1996.

[12] Pascal Poncelet, Florent Masseglia and Maguelonne Teisseire (Editors). Information Science Reference. Data Mining Patterns New Methods and Applications, ISBN 978 1599041629, October 2007.

[13] Thearling K, Exchange Applications White Paper, Inc. increasing customer value by integrating data mining and campaign management software, 1998.

[14] Noah Gans, Spring. Service Operations Management, Vol. 5, No. 2, 2003.

[15] Joun Mack. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. An Efficient k-Means Clustering Algorithm, Analysis and Implementation, VOL. 24, NO. 7, JULY 2002.

[16] Andrew Moore and Brian T. Luke. Tutorial Slides, K-means and Hierarchical Clustering and K-Means Clustering, Slide 15, 2003.

[17] E. Turban, J. E. Aronson, T. Liang, and R. Sharda, Decision Support and Business Intelligence Systems, eighth edition. Prentice Hall, 2007.

Author

Ahmed B. El Seddawy, M.S. degrees in Information System from Arab Academy for Science and Technology in 2009. He now with AASTMT Egypt Teacher Assisting BIS Department.