

A Proposed Data Mining Technique to Improve Decision Support System in an Uncertain Situation

Ahmed Bahgat El Seddawy¹, Prof. Dr. Turkey Sultan², Dr. Ayman Khedr³

¹Department of Business Information System, Arab Academy for Science and Technology, Egypt

^{2,3}Department of Information System, Helwan University, Egypt

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.

Keywords:- Dss, Dm, Mis, Clustering, Classification, Association Rule, K-Mean, Olap, Matlab.

I. 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 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 works. Tables, Figures and Equations.

II. BACKGROUND AND RELATED WORK

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 Artificial Intelligence ‘AI’. And in this paper will using DM as approach and technique.

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]. There are several processes for applying DM:

Definition of the business objective and expected operational environment.

1. Data selection is required to identify meaningful sample of data.
2. Data transformation that involves data representation in an appropriate format for mining algorithm.
3. Selection and implementation of data mining algorithm depends on the mining objective.
4. Analysis of the discovered outcomes is needed to formulate business outcomes.
5. Representing valuable business outcomes.

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.

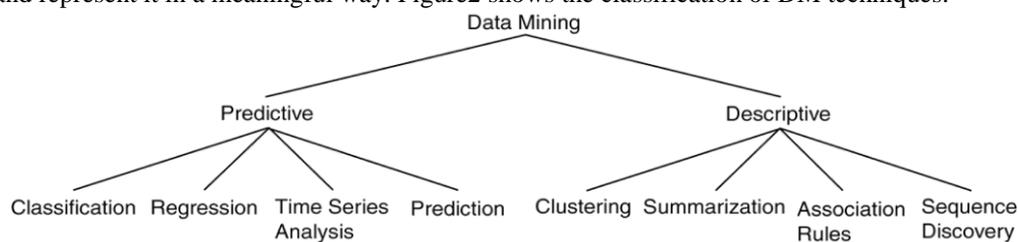


Fig 2: DM Techniques [5]

1. 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 3. 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.

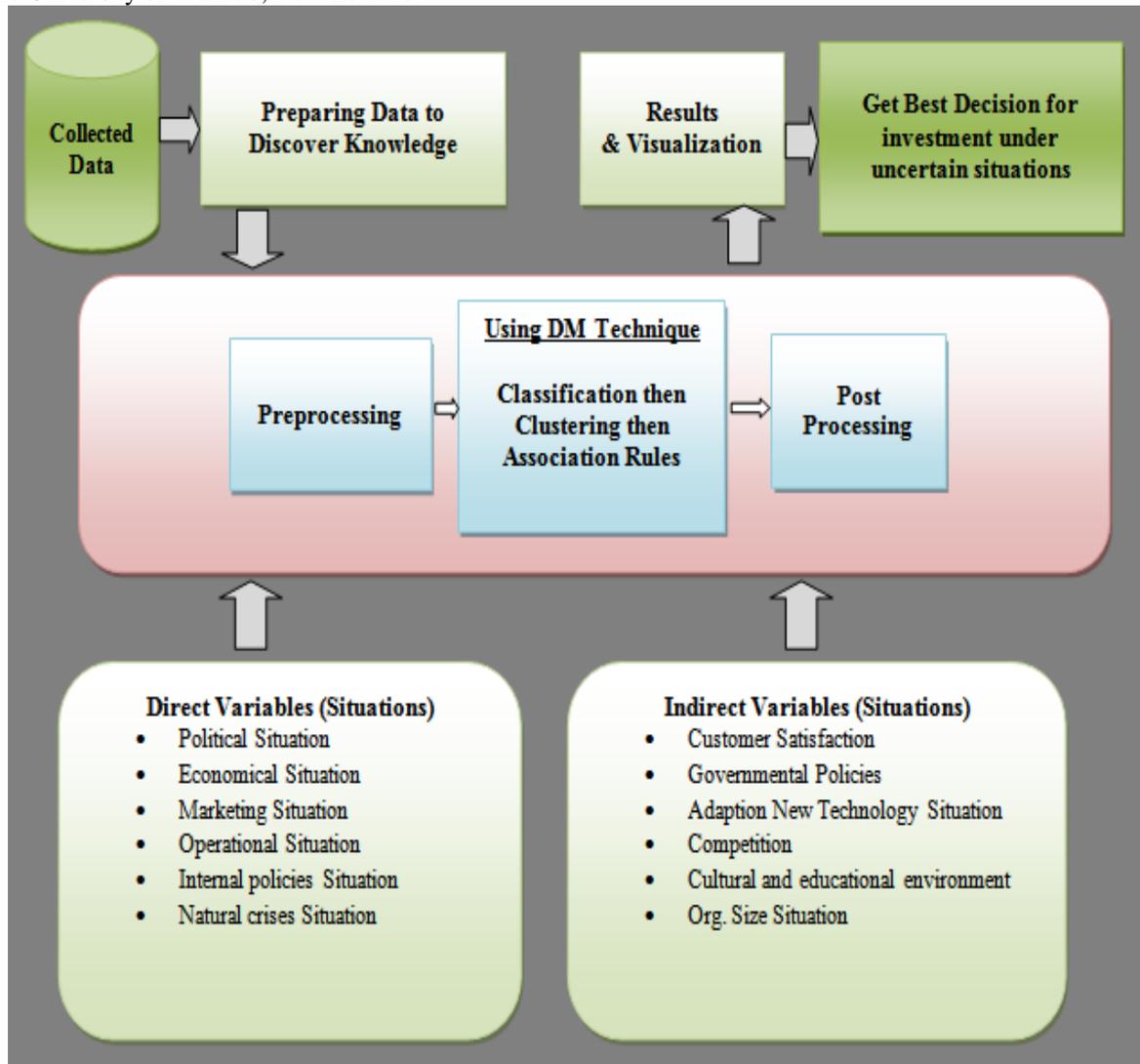


Fig 3: Proposed IDMS

2. IDMS Implementation

IDMS execution done via several techniques started with clustering technique using enhanced k-mean, classification technique using ID3 algorithm and association rules technique using apriori algorithm, this description shown in figure 4.

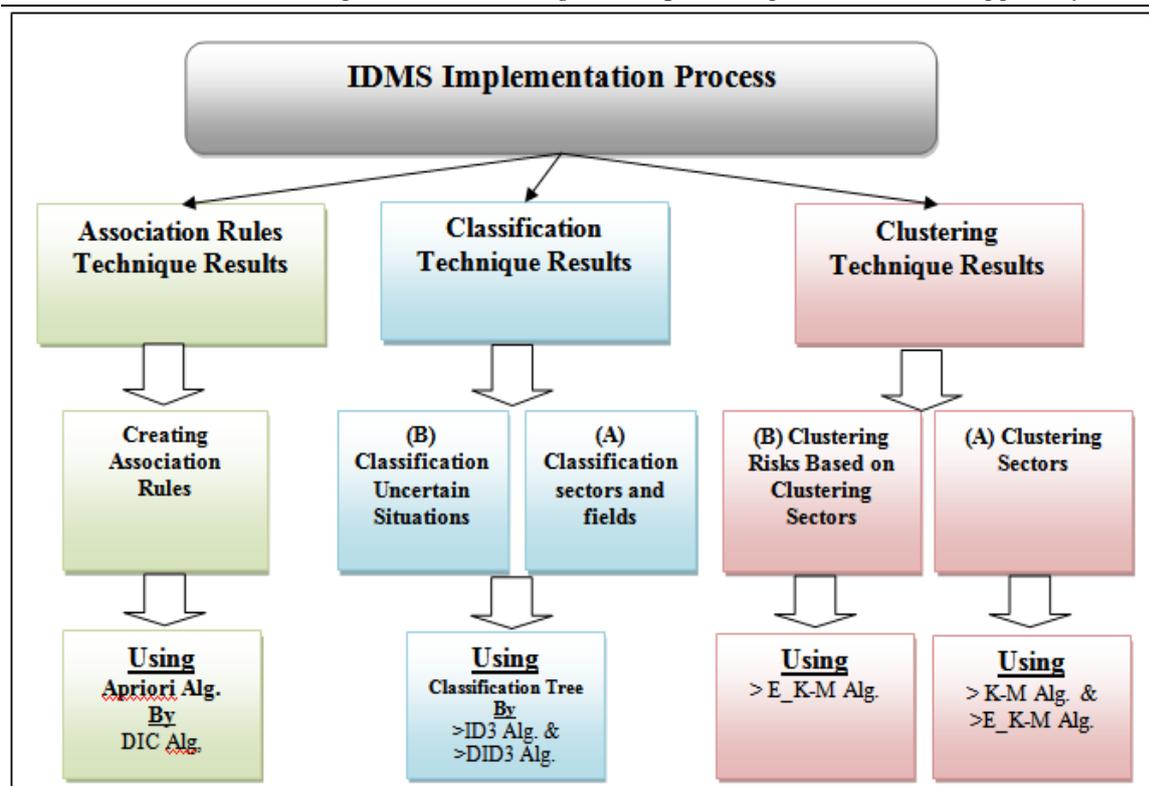


Fig: 4 IDMS Implementation Process

III. CONCLUSIONS

This thesis applies DM in the banking sector investment to support decision-making under conditions of uncertain which rarely been addressed before. Investment decision support system in the area of data mining "IDMS" aims to build data mining system to invest in the banking sector. 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 investment department is planning a set of actions in accordance with IDMS outcomes replacing to manual and expert findings. IDMS uses the most popular algorithms in clustering techniques for getting the best result for the organization to be able to find the best decision for investors. IDMS enhances K-m algorithm to give the best solutions for results. IDMS consists of several components, data collection, data preparation for knowledge discovery processing data, using the techniques of data mining in the sequence of steps starting with data classification, and data collection especially using an algorithm K-m and promote algorithm K-m to set the best result, which then set and run association rules to solve the problem, after processing and finally get a result and as a result create a better perception of the decision to make a good decision to invest in under uncertain conditions.

Proposed new IDMS system is simplicity, and clarity to the owners needs in conditions of uncertainty. It works on the post processing is the best optional ease the interpretation of the results of the investment. The investment destinations for planning a series of actions and in accordance with the results expected from IDMS to replace the results of the expert hand. The proposed system uses the most popular algorithms in exploration and harvesting techniques for the data to get the best result for the organization in order to be able to find the best decision for investors. The proposed system enhances K-m algorithm to provide the best solutions in order to achieve the best results.

The proposed system depends on improving the performance of decision-making using DM under conditions of uncertainty, as well as support the investor to know how much risk and the ability to invest in any sewers under any of the sectors, and this by using a range of different styles algorithms different as well improve K-m algorithm so as to extract best results and clearest in the proposed system.

The proposed system is working to remedy decision within the framework of the work in the field of banking and investment sector in particular and state of research done on a private investment bank in Egypt. Respect to the proposed system to work in the direct investment and determine the degree of risk its own decision. This philosophy of thesis proposes a modified version of the well known k-m clustering algorithm. The modified algorithm maintains all important characteristic features of the basic k-m and at the same time eliminates the possibility of generation of empty clusters. It has been shown that the present algorithm is

semantically equivalent to the serial k-m algorithm. A detailed comparison of this new algorithm with the basic k-m has been reported. Experimental results show that the proposed clustering scheme is able to solve the empty cluster problem and apply algorithm with center point to classify all data with effective size. There are more common features between them such as;

- Both support investment department with good ideas to create good decision.
- Both have more popularity, flexibility, applicability and efficiency in clustering data especially in the investment sector.
- Both algorithms use a large number and small number of clusters.
- The K-Means algorithm and the E_K-M algorithm can both be used to find natural clusters within given data based upon varying input parameters.
- The K-Means algorithms typically join to a solution very quickly as opposed to other clustering algorithms.

The Scientific Results of Research:

Scientific findings extracted yet clear improvement km within the proposed system as follows:

- Improve the performance of algorithms km provides many advantages of high interest in the application of DM as especially in the area of banks under the investment sector.
- Results extracted after improving the performance km accurate and correct distribution and ease of presentation graphics to decision makers.
- Results extracted after improving the performance of the km-quality data also show through the proposed system using utility programs
- Optimization process allowed ease of application of the proposed system in more than one area and various sectors.
- These improvement works to increase time in the implementation but not significant in light of ensuring a high degree of accuracy and quality of results in decision-making.

The Management Results of Research

Results that help those administrative decisions extracted within the proposed system as follows:

- Through these results, it is clear decision-making easier to identify areas investment most use.
- Through these results, it is clear easing of decision-making to determine the sectors investment most use within each domain.
- Through these results, it is clear ease of decision-making to determine the degree of importance of yielding investments is different depending on the severity of the (high - medium - low).
- Process data as graphics graphic illustration allows easy decision-making to a user of the proposed system.
- Support sector investment is a good idea to create a good decision.
- Proposed system increases the ease, flexibility, and efficiency in the application of data collection, especially in the investment sector.

IV. RECOMMENDATIONS

There are several recommendations should be followed in this study should be analyzed before using IDMS, such as; external factors, internal factors and technological factors.

V. FUTURE WORK

In the short-term, the results of data mining will be in profitable, if applying several additions for IDMS, such as, create expert system using 'BIS' in banking sector to support decision making for investment in any sector to help IDMS for future. Expert system had useful additions. This addition will help to recommend investor to another alternative for investing specially under uncertain situations happen in Egypt.

ACKNOWLEDGMENT

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

REFERENCE

- [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. Zaiane. "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 Masegla 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.



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.