Knowledge-Driven Decision Support System Based on Knowledge Warehouse and Data Mining for Market Management

Banaz Anwer Qader[1] and Dr. Murtadha M. Hamad[2]

[1] University of Anbar/ Iraq

Received: 11 December 2012 Accepted: 1 January 2013 Published: 15 January 2013

Abstract

Over the recent years, a great interest has appeared in studying " knowledge warehouse (KW) , decision support system (DSS) , data mining (DM) , and knowledge discovery process in database (KDD)" , taking into consideration that each of these fields is related to and influenced by the others. In order to manage an enterprise resources, there is a necessary need to build a DSS which helps the manager or the decision maker in the decision making and managing processes. While the primary goal of a (KW) is to provide the decision-maker with an intelligent analysis platform that enhances all phases of the knowledge management process , the (KDD) process should be applied to discover knowledge and build (KW) , where (DM) technique is considered the most important step in the process (KDD).So in this paper , we merged the concepts of data warehouse (DW) and knowledge warehouse (KW) proposing and building a system of kind (knowledge-driven DSS) which depends on KW for managing (storing and retrieving) the knowledge for the benefit of the process of decision making and management of the market resources (items), where this study includes an application on a DW of marketing building resources (items).This study uses the data mining technique specifically its functionality (Association Rules Mining) in the knowledge discovery process and building KW. Eventually, the designed system was constructed and executed by using (C version 2008) which is a visual and object oriented programming language. Good system results (knowledge) were obtained in a very little time taking two minutes approximately. This proves the efficiency of the proposed algorithms and our knowledgedriven DSS system in the supporting the market manager or decision maker to take accurate and right decisions for managing the market items in a perfect way.

Index terms— decision support system, data mining, knowledge discovery, knowledge warehouse, data warehouse.

1 Introduction

ecision Support Systems (DSS) increasingly become more critical to the daily operation of organizations [1]. Decision Support System (DSS) is an equivalent synonym as management information systems (MIS). Most of imported data are used in solutions like data mining (DM). Successfully supporting managerial decision -making is critically dependent upon the availability of integrated, high quality informationorganized and presented in a timely and easily understood manner [2]. Since the mid-1980s, data warehouses have been developed and deployed as an integral part of a modern decision support environment [1]. Therefore Data Warehouse provides an infrastructure that enables businesses to extract, cleanse, and store vast amounts of corporate data from operational systems for efficient and accurate responses to user queries [3]. Data Warehouse (DW) is one of the solutions for decision-making process in a business organization. But it only stores data for managerial

purpose and it has no intelligent mechanism for decision making. This raises the issue of knowledge storage in organization for high capability decision support [4]. knowledge in the form of procedures, best practices, business rules, expert knowledge, facts within context, and processed data can be stored in logical structures accessible by computers. The logical structures in the knowledge warehouse to store knowledge are analogous to the system of tables that implement data storage in the data warehouse. Knowledge is applied through a layered representation that is readable by both humans and machines this representation is also a system executable that is portable and can be run on a computer to help make decisions and take actions [5]. The enterprise-wide information delivery systems provided in a data warehouse can be leveraged and extended to create a knowledge warehouse (KW). A framework of knowledge warehouse is introduced, which is enhanced form of data warehouse to provide a platform/ infrastructure to capture, refine and store consistent and adequate knowledge along with data to improve decision making in an organization [4]. The primary goal of a (KW) is to provide the decision-maker with an intelligent analysis platform that enhances all phases of the knowledge management process. Knowledge Warehouse (KW) architecture will not only facilitate the capturing and coding of knowledge but also enhance the retrieval and sharing of knowledge across the organization [3]. In order to understand, analyze, and eventually make use of a huge amount of data, Enterprises use mining technologies to search vast amounts of data for vital insight and knowledge. Mining tools such as data mining, text mining, and web mining are used to find hidden knowledge in large databases or the Internet [6]. Data Mining (DM) is the process of identifying interesting patterns from large databases. Data mining has been popularly treated as a synonyms of knowledge discovery in databases, alt hough some researchers view data mining as an essential step of knowledge discovery **???**].

In this paper, mining tools are automate software tools used to achieve decision making process by finding hidden relations (rules), and predicting future events from vast amounts of data.

2 II.

The Knowledge-Driven Dss A knowledge-driven DSS provides specialized problem solving expertise stored as facts, rules, procedures, or in similar structures. It suggest or recommend actions to managers [8].

A KD-DSS is a knowledge driven decision support system, which has problem solving expertise. The KD-DSS can give suggestions or recommendations based on several criteria's. These systems require human-computer interaction. Advanced analytical tools like data mining can be integrated with the KD-DSS to find hidden patterns. Knowledge Driven DSS is also called as Intelligent Decision Support methods, and it is analogues to the knowledge warehouse strategy work. We choose KD-DSS model, because it has capacity to self-learn, identify associations between the data, and perform heuristic operations, if required. These abilities turn the DSS system into intelligent, increase the capacity of problem solving and improve suggestion accuracy. It is important to mention that the Knowledge representation play key role in KD-DSS. Well-defined knowledge representations include rule-based systems, semantic web and frame systems. A rule-based system contains rules in the database [9].

3 III.

4 Knowledge Warehouse

knowledge warehouse (KW) can be thought of as an "information repository". The knowledge warehouse consists of knowledge components (KCs) that are defined as the smallest level in which knowledge can be decomposed. Knowledge components (objects) are cataloged and stored in the knowledge warehouse for reuse by reporting, documentation, execution the knowledge or query and reassembling which are accomplished and organized by instructional designers or technical writers. The idea of knowledge warehouse is similar to that of data warehouse. As in the data warehouse, the knowledge warehouse also provides answers for ad-hoc queries, and knowledge in the knowledge warehouse can reside in several physical places [10].

A knowledge warehouse (KW) is the component of an enterprise's knowledge management system. The knowledge warehouse is the technology to organize and store knowledge. The knowledge warehouse also has logical structures like Computer programs and databases to store knowledge that are analogous to the system of tables that implement data storage in the data warehouse [5]. The primary goal of a KW is to provide the knowledge worker with an intelligent analysis platform that enhances all phases of the knowledge management process [3] [1]. Like the DW, the KW may be viewed as subject oriented, integrated, time-variant, and supportive of management's decision making processes. But unlike the DW, it is a combination of volatile and nonvolatile objects and components, and, of course, it stores not only data, but also information and knowledge [11].

The KW can also evolve over time by enhancing the knowledge it contains [3]. Knowledge warehouse provides the infrastructure needed to capture, cleanse, store, organize, leverage, and disseminate not only data and information but also knowledge [4].

IV.

5 Knowledge Discovery Process

Knowledge discovery in databases (KDD) is a rapidly growing field, whose development is driven by strong research interests as well as urgent practical, social, and economical needs. The term KDD is used to denote

the overall process of turning low-level data into high-level knowledge. A simple definition of KDD is as follows: Knowledge discovery in databases is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [12].

Knowledge Discovery has also been defined as the 'non-trivial extraction of implicit, previously unknown and potentially useful information from data'. It is a process of which data mining plays an important role to extract knowledge from huge database (data warehouse) [13]. Data mining is the core part of the knowledge discovery in database (KDD) process as shown in the figure (1). The KDD process may consist of the following steps: 1) data integration, 2) data selection and data pre-processing , 3) data mining as it will be explained in section 5; 4) interpretation & assimilation. Where data comes in, possibly from many sources. It is integrated and placed in some common data store like data warehouse. Part of it is then selected and pre-processed into a standard format. This 'prepared data' is then passed to a data mining algorithm which produces an output in the form of rules or some other kind of 'patterns'. These are then interpreted to give new and potentially useful knowledge. Although the data mining algorithms are central to knowledge discovery, they are not the whole story. The pre-processing of the data and the interpretation of the results are both of great importance [13].

V.

6 Data Mining Technique

Data mining (DM) is one of the most important techniques that are used to discover required knowledge for intended enterprise.

Data mining derives its name from the similarities between searching for valuable information in a large database and mining rocks for a vein of valuable ore. Since mining for gold in rocks is usually called "gold mining" and not "rock mining", thus by analogy, data mining should have been called "knowledge mining" instead [14]. Data mining is the knowledge discovery process by analyzing the large volumes of data from various perspectives and summarizing it into useful information [15].

Data mining is the process of discovering interesting knowledge, such as patterns , associations, changes, anomalies, and significant structures from large amount of data stored in databases, data warehouse, or other information repositories **??**16]. Data mining refers to discover useful, previously unknown knowledge by analyzing large and complex" data sets. Data mining is defined as the extraction of patterns or models from observed data [12].

Data Mining, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and knowledge discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process [14].

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, transforms data into actionable results **??**17].Other similar terms referring to data mining are: data dredging, knowledge extraction and pattern discovery [14].

7 VI. The Proposed and Designed System

In this paper, we proposed a knowledge-driven DSS and it consists of several phases as shown in the figure (2). These phases are: Access file) in our system to be ready for importing in to the C# environment for other data pre-processing techniques like resolving inconsistency and reduction.

In our proposed system, integration step led to emerging duplicated records (transactions) and inconsistent attributes which are processed in the data pre-processing phase by applying proposed algorithms of reduction and consistency techniques that are (Removing Duplication (Reduction) Algorithm) and (Resolving Inconsistency Algorithm). The cleaned and prepared data from pre-processing phase are loaded into the data warehouse (DW) which is a wide data store of the market that contains historical data and complete information about building items and has capability of modifying its data and ready for processing phase. In order to mine vast amounts of data in the data warehouse for discovering knowledge, part of the data should be selected and customized in the Data Selection phase, where we use the concept of data mart to select and customize the data for processing phase depending on the technique used for knowledge discovery. In Data Selection phase the set of items is selected for Data Mining and as input of the proposed (Index-based Apriori Algorithm) because the used technique is Data Mining and specifically the Association functionality. In the discovering knowledge phase, we use Data Mining and apply its Association functionality. The selected set of items is entered to the proposed algorithm (Index-based Apriori) for mining association rules. The number of mining association rules are different based on specified and entered min. count threshold for generating supported itemsets and min. confidence threshold for generating interesting association rules. The market manager to be able of taking decisions and managing the market resources, these rules must be interpreted for discovering knowledge to support the process of decision making.

In the Association Rules Interpretation phase, we proposed and used an algorithm named (Association Rules Interpretation Algorithm) applying a simple statistical method which is represented by substituting and counting the items in the antecedent and consequent of the association rules. The results of this system represent the

discovered knowledge which are the predicted ratios of items sales for the next year. The results and visualization phase which we explain and discuss in the next section, visualizes the results graphically using Line Chart tool to provide the decision maker or the market manager with conceptual values (knowledge) supporting him in managing the market easily and in a perfect way.

8 VII.

9 Implementation and Results

The proposed and designed system has been executed by using (C# programming language). So the implementation of the system is performed on phases.

The system includes several interfaces to execute it easily and to support the manager or decision maker in the process of decision making.

The discovered knowledge in our system refers to the predicted ratios of sales for the items during a specified month in the next year based on statistic analysis applied on items' sales through the previous years stored in our marketing Data Warehouse (DW). The visualized results that have been illustrated clearly in figures (4), (**??**), (**??**) are for "January" of the next year. It is important to mention that these predicted ratios of items sales are being different by differing the min. count threshold and min. confidence threshold of the Index-based Apriori and according to the chosen specified month. Therefore, we executed our system and got various results (ratios) for "January" using three different min. count thresholds (2100, 2150, 2200) and three min. confidence thresholds (60% , 80% , 100%) as illustrated below in figures (4), (**??**), (6). We used three various min. count thresholds (2100 , 2150 , 2200) each of which with three various (lower, mid, higher) min. confidence thresholds which are (60% , 80% , 100%), (70% , 80% , 90%) , and (50% , 90% , 100%) for executing the system to generate various number of supported itemsets and interesting association rules and getting various ratios of items sales , as shown below in table (1).

Table (1) shows various system results for "January" of the next year according to different min. count thresholds for supported itemsets and min. confidence thresholds for interested association rules.

10 Conclusions

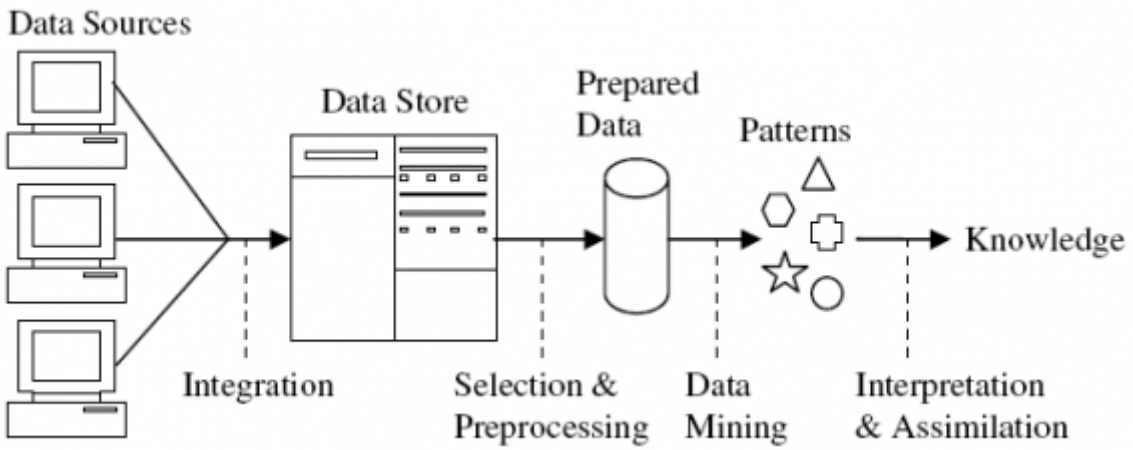
After the implementation of our DSS system and through the execution of the Index-based Apriori algorithm for association rules mining, and Association Rules Interpretation algorithm and from obtained results, we concluded the following: 1. Through the execution of our system, it is become explicit that the knowledge warehouse (KW) is smaller, more accurate and more close-fitting than the data warehouse (DW) because the knowledge that has been stored in the (KW) in the form of rules or patterns or any other forms is discovered and gained from large amount of data stored in the (DW). 2. The accuracy of discovered knowledge depends on the specified and used thresholds in the Indexbased Apriori algorithm. The knowledge accuracy increases by decreasing the min. count threshold and min. confidence threshold, because using lower thresholds increases the number of supported itemsets and interested association rules which lead to get more accurate knowledge and support the manager or decision maker to take accurate decisions. 3. Reducing number of itemsets will reduce the number of generating association rules and lead to gain low quality knowledge. 4. Reducing number of generating itemsets and association rules will lead to shorten run time and will reduce the used space in memory. In order to reduce the used space in memory and shorten run time without reducing the number of itemsets and association rules, we have used indexing method for fast access through applying the proposed Index-based Apriori algorithm.

[1] [2]

[1] © 2013 Global Journals Inc. (US)

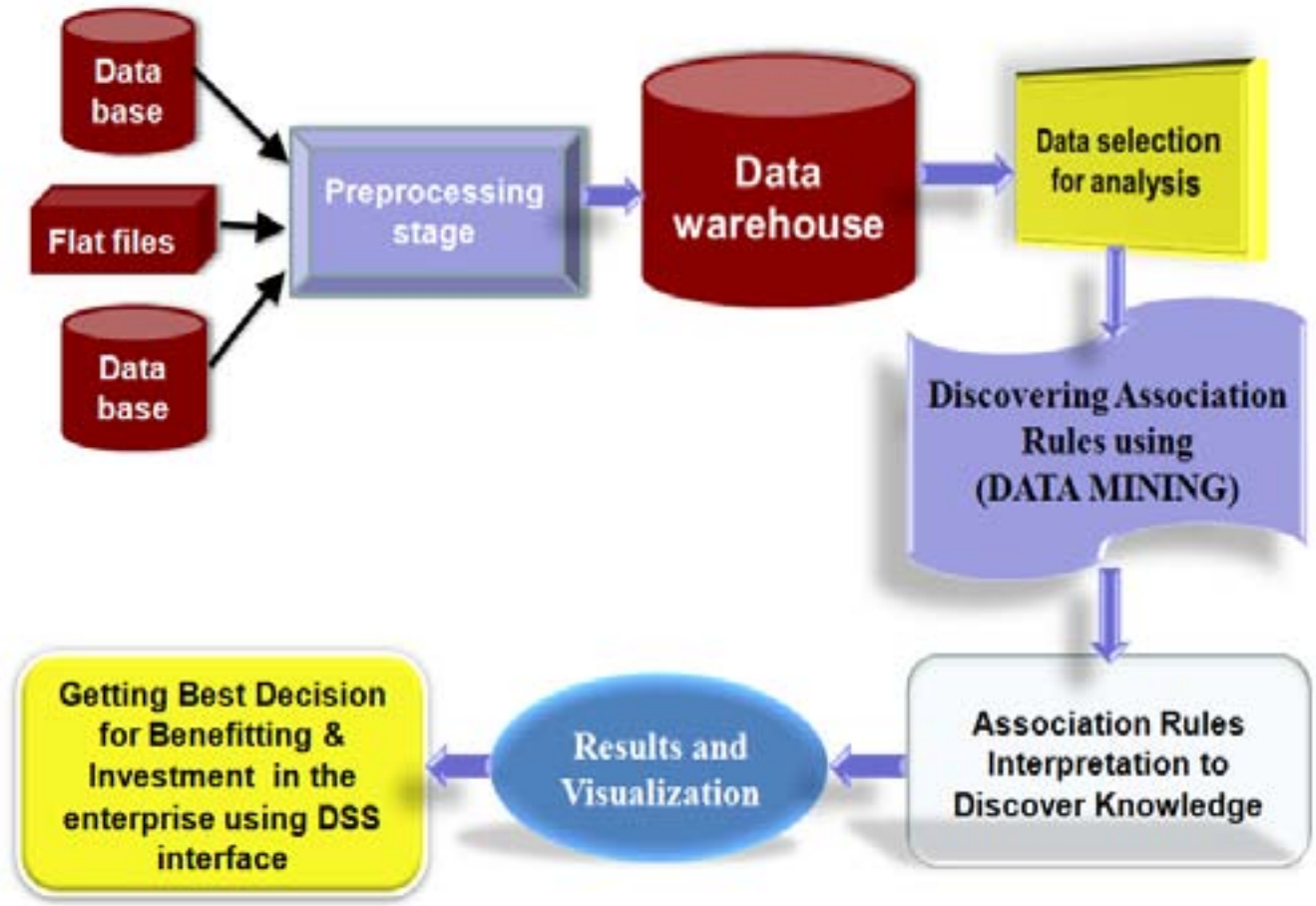
[2] © 2013 Global Journals Inc. (US) Knowledge-Driven Decision Support System Based on Knowledge Warehouse and Data Mining for Market

Figure 1: D



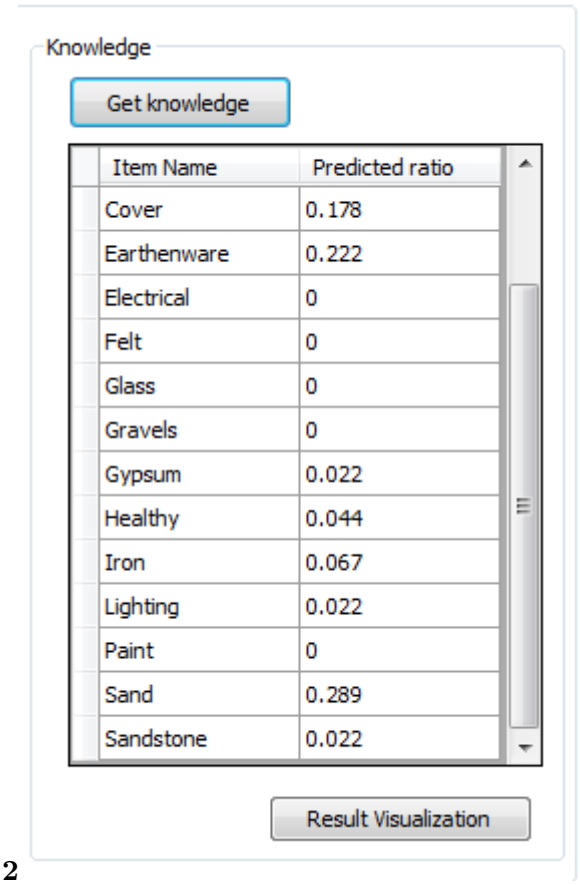
1

Figure 2: Figure 1 :


Data base
Flat files
Data base
Preprocessing stage
Data warehouse
Data selection for analysis
Discovering Association Rules using (DATA MINING)
Association Rules Interpretation to Discover Knowledge
Results and Visualization
Getting Best Decision for Benefitting & Investment in the enterprise using DSS interface

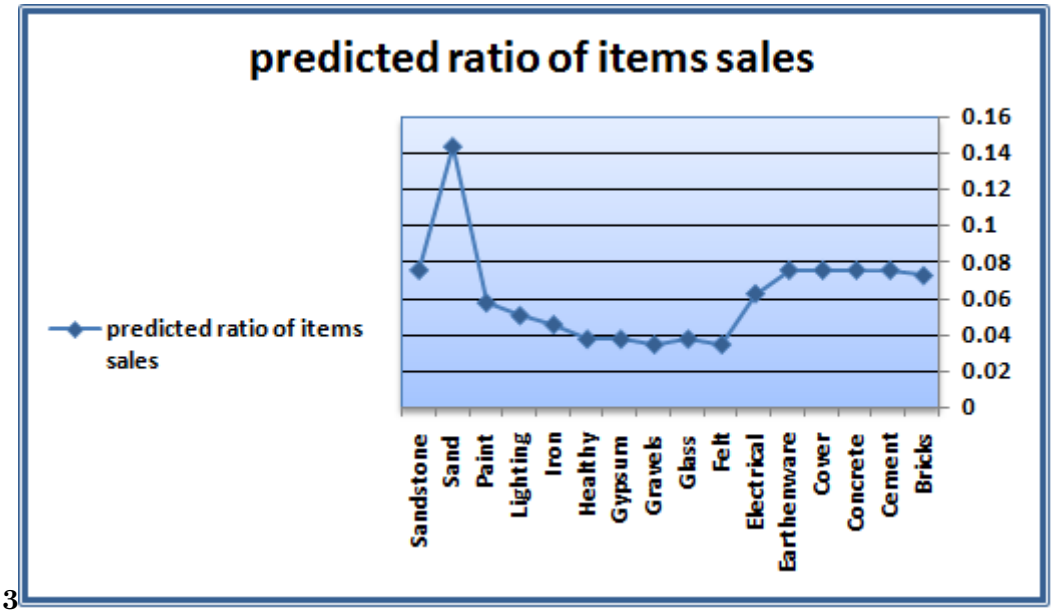
1

Figure 3: 1 .



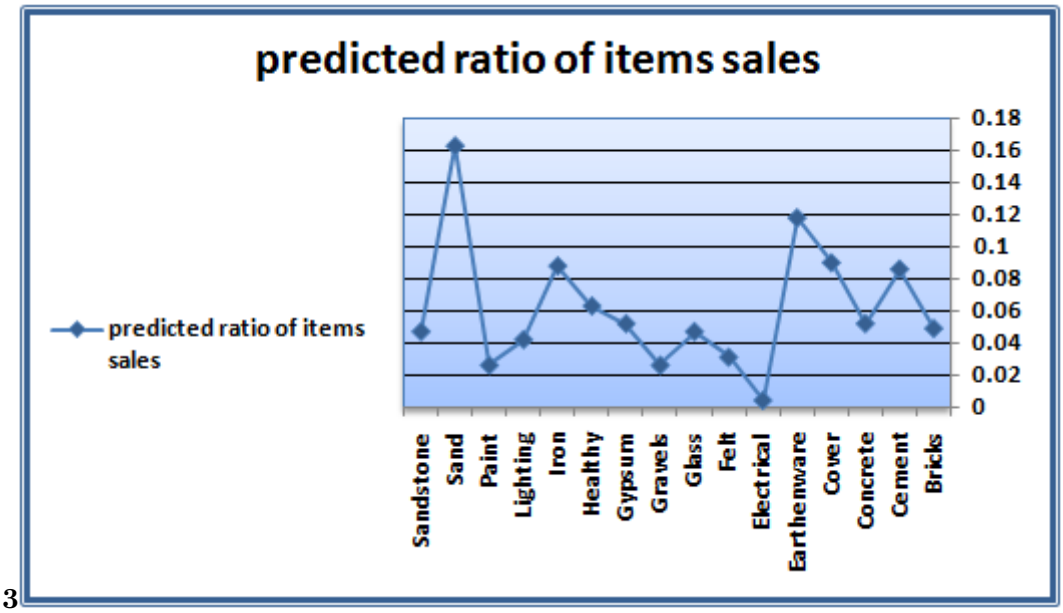
2

Figure 4: Figure 2 :



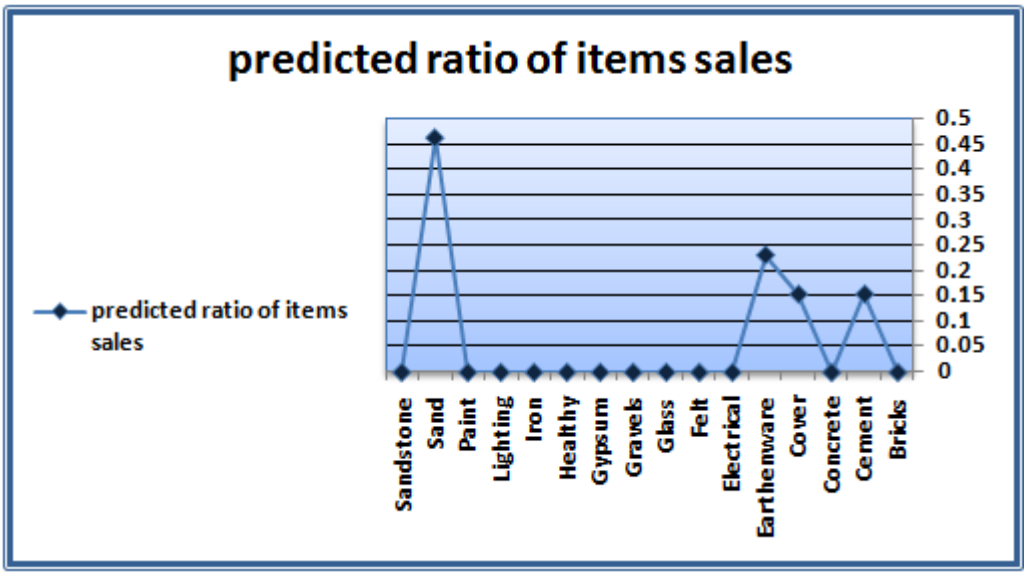
3

Figure 5: Figure (3)



3

Figure 6: Figure 3 :



456

Figure 7: Figure 4 :Figure 5 :Figure 6 :

1

2013
ear
Y
Volume XIII Issue X Version I
)
(
Confidence Thresholds for Rules for "January"
Global Journal of Management and Business Research

Figure 8: Table 1 :

[Goebel and Gruenwald (1999)] 'A Survey of Data Mining and Knowledge Discovery Software Tools'. Michael Goebel , Le Gruenwald . *SIGKDD Explorations. Copyright © 1999 ACM SIGKDD*, June 1999. 2009. 1 p. .

[Ahmed Bahgat El Seddawy1, Dr. Ayman Khedr2 and Prof. Dr (2012)] 'Adapted Framework for Data Mining Technique to Improve Decision Support System in an Uncertain Situation'. *International Journal of Data Mining & Knowledge Management Process (IJDKP)* Ahmed Bahgat El Seddawy1, Dr. Ayman Khedr2 and Prof. Dr (ed.) May 2012. 2 (3) p. . (Turky Sultan)

[Suresh et al. ()] 'An XML Based Knowledge-Driven Decision Support System For Design Pattern Selection'. S Suresh , Prof M M Naidu , S Asha Kiran . *International Journal of Research in Engineering and Technology* 2277 -4378. 2012. IJRET. 1 (3) .

[Rupali and Gupta] 'Data Mining: Techniques, Applications and Issues'. Gaurav Rupali , Gupta . *International Journal of Advanced Research in Computer Science*

[Abdul-Aziz Rashid Al-Azmi (2013)] 'DATA, TEXT, AND WEB MINING FOR BUSINESS INTELLIGENCE: A SURVEY'. Abdul-Aziz Rashid Al-Azmi . *International Journal of Data Mining & Knowledge Management Process (IJDKP)* March 2013. 3 (2) . Kuwait University

[Fu (1997)] Fu . *Data Mining : Tasks , Techniques , And Applications*, Oct/Nov 1997. IEEE. 16. (Issue 4 Pages 18 -20)

[Firestone ()] *Knowledge Base Management Systems and The Knowledge Warehouse: A (Strawman)*, Joseph M Firestone , PhD . http://www.dkms.com,eisai@home.com 2009. Executive Information Systems, Inc. p. . (Executive Information Systems)

[Mir Sajjad Hussain Talpur et al. ()] 'Knowledge Warehouse Framework'. Hina Mir Sajjad Hussain Talpur , Sher Muhammad Shafi Chandio , Hira Chandio , Sajjad Talpur . *International Journal of Engineering Innovation & Research* 2277 -5668. 2012. 1 (3) p. .

[Hamid et al. ()] 'Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing'. R Hamid , David M Nemati , Lakshmi S Steiger , Richard T Iyer , Herschel . http://www.elsevier.com/locate/dsw *Decision Support Systems* 2002. 33 p. .

[Hamid et al. ()] *Knowledge Warehouse: An Architectural Integration of Knowledge Management, Decision Support, Data Mining and Data Warehousing*, R Hamid , David M Nemati , Lakshmi S Steiger , Richard T Iyer , Herschel . 2009. University of North Carolina at Greensboro

[Bramer and Book ()] *Principles of Data Mining*, Max Bramer , Book . 2007. Springer-Verlag London Limited. (Printed on acid-free paper)

[Cmput690 et al. ()] *Principles of Knowledge Discovery in Databases"\Chapter I: Introduction to Data Mining*, Cmput690 , R Book ; © Osmar , Zaïane . 1999.

[Power et al. ()] *Reflections on the Past and Future of Decision Support Systems: Perspective of Eleven Pioneers "\chapter*, Daniel J Power , Frada Burstein , Ramesh Sharda . 2011. Springer Science+Business Media, LLC.

[Yacci (1999)] *The Knowledge Warehouse: Reusing Knowledge Components*, Michael Yacci . September 1999. 2008. 12 p. .

[Dymond et al. ()] 'The Knowledge Warehouse: The Next Step Beyond the Data Warehouse'. Anthony Dymond , Dymond , Associates , Llc , Concord , Ca . *Data Warehousing and Enterprise Solutions \SUGI 27 \Paper*, 2008. p. .