

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 8, Issue. 3, March 2019, pg.240 – 248

Design and Implementation of a Prototype Data Mining Agent System

By

Ali Abdulhussein Mohammed

Supervised by: **Dr. Sefer Kurnaz**

Altinbaş University, Turkey

ABSTRACT: *An analytical, descriptive and perhaps instinctive methodology regarding sourcing out ways to improve business opportunities and technical know-how is incorporated within the scope of Data Mining. Its designation relates to how an event deals with obtaining unprocessed information from vast databases by deeply searching the resourceful highland areas for further processing to qualify valuableness. Its applications are typically employed in facilitating the operations of marketing, sales and services of business though it works more with stable datasets that are kept in the business's threshold of sensitive and discreet information. The essence of its strategies in the modern world includes automated prediction of trends and behavior as well as abstracting initially unknown patterns. Grave utterances concerning the supposedly efficient data mining techniques have been made significant over the years since its discovery. There are still positive and negative remarks about these said techniques claimed to improve business functionalities. The former which may involve how increased processing speeds, lowered costs for storage and better software packages tend to make data mining more economical. And as for the latter, involving how it is deemed unethical to use the processes of data mining as it violates innocent people's privacy. Data mining can be quite an enigma for the inadequately learned individuals and firms due to its cognitive requirements that appear to be highly critical and extensive upon application.*

KEYWORDS: *DATA MINING METHODS, DATA MINING STRATEGIES, DATA MINING PROCESS, APPLICATIONS OF DATA MINING, IMPORTANCE OF DATA MINING.*

1. Introduction

The paradigms of how data mining has been useful in the business industry are tremendously increasing from time to time. When one engages in an activity wherein accurate measures to marketing protocols and securing of its business's processes are integrated, it tends to cover an extreme wideness of consistent effectiveness as to how operations are carried out. Its versatility has been used in multiple business fields like engineering, communication, transportation, medicine, architecture, and so on.

The elaborate mechanism behind the concept of data mining has aided many research and cognitive schemes about human resource management and its extensiveness. This data can serve as a useful medium for knowledge management initiatives that may further initiate data mining strategies that bring about increment in an organization's assets. It merely becomes logical to assume why individuals put in tactical measures to search and deduce methods that involve large data creation and management in order to acquire potential patterns, relations and associations.

Employing the fundamental processes from data mining can be relatively efficient in customer-relationship management. A system that deals with how the organization manages its interaction with customers or patrons by using proper data analyses about the history and subsequent transactions made with respect to the business concerned. Data Mining techniques, in this case, enhances the software that contains a wide range of applications designed to aid companies manage several business protocols which may majorly cover the access business information, customer data and customer interaction.

Despite the relative importance and applications that are customarily portrayed by data mining, the system has not been adequately adopted by many large companies, especially those involved with database communities. A prominent explanation behind this flaw is based on the nature of its process being difficult and laborious. The conventions that regard the applications of data mining can only be efficient with quality results when great efforts and multidisciplinary skills are invoked optimally.

As a result of the high expectations required from its typical methodologies, properly trained experts are to be accountable for making an iterative, multistep procedure to prepare and consistently access the databases which involve the selection of a suitable algorithm to extract all useful data. It is also by instituting new methods for smarter data analyses that mining of relevant processed data can be achieved. Just choosing the apt mining algorithm alone needs may need more effort and time from its users so the idea of enhancing updates for the analytical activities become essential.

For the ability of one to use the right principles for employing data mining techniques at large, problems that arise from automation and its related flexibility must be corrected appropriately. Mohammed J. Zaki and Limsoon Wong considered this problematic challenge among the most predominant challenges for data mining after making numerous observations from the scientific and technological aspects of the application. One reason is that the institution and usage of initial episteme into the concept of data mining algorithms can be usually tricky and arduous, even for a computer scientist.

However, professional agents from agent-based systems have been deemed increasingly prevalent in the computer-technology industries over the past few years. This is due to their modularity, flexibility and general versatility in handling a broad range of issues. Some of the popularly known problems that have been constantly treated from time include data analysis and data filtering, workflow management, condition monitoring, information brokering, alarm generation, gaming and simulation.

Extensive cognition about data mining supported by agents that tend to maximize the epistemic strategies behind the processes of this mining of data by providing simplified analyses of information systemization gotten from the patterns of massive data accumulation

The first person known to unravel and experience the agent-based data mining concepts was Winston Davies in the year, 1994. It was through his initiation of its relative applications that others began to create and develop further integrations that work between the two fields.

An act of combining data mining methods and agents' protocols to carry out specific activities invariably assist individuals in their respective ventures by strictly elaborating on integral models within a very

rational period. In essence, this facilitates thorough findings of information that may seem constantly discreet in a data store and then attain updates on general information and significant trend changes.

2. Literature Survey

- In some previous research and surveys, the need to constantly retrieve, analyze, brief, control, and visualize relevant information is employed in the internet and social media. This was revealed and defined constructive for data mining due to the vast social media platforms and applications that are further classified into web blogs, social network sites, forums, podcasts, media platforms, rating/review communities, bookmarking sites, and avatar-based VR media. (Zeng et al. 2010)
- The means required to monitor the operation of data mining over these media can be cumbersome due to the multitudinous networking sites as well as the large amounts, dynamics, and complexity of the media data. A more critical reference is given to the set of data and metadata which have not been addressed systematically in data-text mining literature. (Zeng et al. 2010)
- Data Mining techniques are applied in a wide range of web domains wherein huge amounts of data are identified among unknown or hidden patterns of information upon availability. Accordingly, this indicates a designation that data mining methods used in the World Wide Web are typically referred to as web mining. Likewise, those used in text are called text mining while those used in libraries are called bibliomining. (N. Girija and S.K. Srivasta 2006)
- The analysis involved the application of data mining is usually categorized into Classical Statistics, Artificial Intelligence (AI), and Machine Learning. Classical statistics is related to the study of data and its relationships as well as dealing with numeric data in large databases. The artificial intelligence brings about solving the statistical problems which may involve generic algorithms, fuzzy logic, and neural computing while the machine learning is based on combining advanced statistical methods and AI heuristics for data analyses and knowledge discovery. (Girija and Srivasta 2006, David J. Hand 1998, Kononenko and Kukar 2007)
- Figure 1 shows a normal data mining process that depicts an interactive sequence of steps that is initiated by raw data integrations obtained from various sources and then subjected to thorough analyses of duplicated or inconsistent data. The analyzed data are then classified into apt formats for subsequent mining processes. Filtration and aggregation techniques are then incorporated for the extraction of summarized data. Knowledge becomes achieved as the user attains patterns of more detailed information concerning data mining. (Han et al 2011)

2.1. RADICAL COGNITION AND PROCESSES OF DATA MINING

2.1.1. INTRODUCTION TO THE SCOPE AND CONCEPT

The term, data mining can be related to layman's language as the digging or excavation of valuables for productive and perhaps lucrative means. It makes perfect sense since the activities of data mining can be defined as mining deep in search for raw data which may appear in different forms in order for the "miner" to attain patterns about the initially unprocessed data and acquire knowledge from that pattern for the following application methods.

The processes concerned with data mining enable the user to initially sort out large data sets then identify patterns and relationships that can be established to carry out data analysis and provide solutions to

various specific issues in line with its concept of performance. In essence, productivity comes into play when analytical services portray capabilities in creating or developing a series of data mining solutions. This also includes the analyzing and predicting results of customer purchase control to monitor potential buyers and enhance lucrativeness.

From one perspective, one can regard the retrieval of data as a procedure to get new natural resources that offer relevant items that are to be processed and stored as an important tool for future references and subsequent applications. This is one apparent reason as to why the foundation of the fourth industrial revolution greatly relies on the scope and concept of data and connectivity.

Presence of unidentified data within a system merely assumes grave importance to one dealing with mining of data. To employ the data mining strategies appropriately, the fundamentals to its process must be acquainted as well as having sufficient knowledge about its application techniques.

The scope of data mining can also be described as Knowledge Discovery in Data which concerns finding huge storage of data and unravelling hidden patterns and trends that may involve complex analyses. Consequently, it employs multidisciplinary skills and correct measures to perform efficiently. The basic processes that are covered by data mining principles are discussed below.

- **Accumulation and Preparation of Data**
Every starting data mining process involves gathering all achieved data and properly organizing them. Doing this makes further findings of information by the user much easier and accessible.
- **Classification**
This merely deals with the application of the right methodologies for various models in order to mark out the needed parameters to values that are as accurate and precise as possible. It is a task that depicts and discerns data classes and concepts for different models constructed. This principle brings about identification of data set, deducing of new categories, and providing updates on data trends.
- **Cluster Analysis**
This is defined as a statistical classification approach for discovering whether all data that have been accounted are in different groups thereby making quantitative comparisons of combined characteristics. In essence, situations where different objects that tend to exhibit similar characteristics are separated and categorized in a single cluster by the virtue of automation.
- **Evaluation Paradigms**
The methods that are invoked chiefly depend on the analysis capacity to cover a broad range of organizational information and needs with the most rational verdict to be made. Before this can be achieved properly, the training and classification phases must have been employed.
 - The training procedure has to do with designing of classification models via algorithms. The algorithms which vary in the application are used to create a classifier by subjecting models to learning schemes that will help improve its prediction of optimal outcomes.
 - The classification phase here involves the use of well-learned models to predict class labels. It is performed accurately when the originally designed models are rendered by test data in order to estimate the maximize the accuracy of classification principles.

2.1.2. USES AND APPLICATION OF DATA MINING STRATEGIES

Data mining is a task that is popularly claimed to be versatile and efficient in many fields. It is primarily used in areas that are in constant search of information to develop stronger customer focus with CRM integration, unauthorized detection as well as analytical statistics.

This typical activity “digs out” raw data from transactions, consumer price rates, preferences, impact on sales and general satisfaction of services rendered and then processes them adequately to deduce strategic predictions and decisions for the user. The commonest places to witness the applications of data mining mainly include the following:

- **Banking and Finance**

In modern technology, great electronic data repositories are currently being maintained by banks and other financial institution which results in why it becoming strategically important for the banking sector. Conventionally, data mining helps a bank in acquiring hidden patterns and determining unknown relationship from the data which is why technological innovations have enabled banking and finance industries to convey effective acquisition and delivery channels. Equally acknowledging the importance of customer relationship management, the banking industries tend to facilitate their strategies with data mining in order to achieve informative ways to build long-lasting relationships with customers and increase their profits and revenue as well. This is ultimately exhibited due to the circumstances that the financial industries may face regarding how to retain patrons at the lowest cost. CRM focuses on shifting from customer acquisition to customer retention in a very flexible manner when the ethics of data mining are applied sufficiently. Also, fraudulent activities become invariable minimized when data analysis is incorporated into financial activities. Detecting fraud within the system via traditional means of data analysis requires a great deal of complex and time-consuming investigations. However, the use of data mining processes does not just make fraud detection easier to monitor but gives a predictive systemization of controlling further unethical activities.

- **Sales and Marketing**

Data mining is known to be used in various marketing business industries to mainly predict and discover new trends in the system. It helps many marketing professionals to develop and improve their understanding of customer behavior and related interactions with the business. As a result, the users get a better understanding that enables them to target marketing campaigns more aptly and work in line with what and how the sales and marketing schemes can favor the business. Data mining is typically applied in the market-basket analysis in order to provide processed data on what transaction processes occur at subsequent periods. This includes what were bought as well as the sequence of purchases that were made on a timely basis. Merely doing this helps the sales and marketing sector to promote their most profitable products and deduce a similar pattern as to how other products can be enhanced likewise. The retail companies in this sector make use of data mining methods equally search and identify patterns that are used to regulate new and trending information about the regulation of their products among customers.

- **Health Care**

The application of data mining strategies in health care or insurance schemes chiefly define how versatile the activity can be in various aspects of information. Since data mining has a purpose of analyzing wide range sets of data, these patterns gotten from the databases can determine or predict new health and safety measures to facilitate and increase health-care

conducts for every individual. There are some prominent enhancements that have been observed in recent years since the inception of data mining to the welfare of mankind.

- Measuring the effectiveness of treatments: This deals with the analogical methodology concerning symptoms, causes and remedies of diseases and how to find the ideal course of treatment can be measured effective over a specific time. The consistent application of data mining in this category aids higher standards of treatment methods by improving diagnosis and remedial processes at an easier and faster rate.
- Detecting abuse: It involves finding typical patterns that are further identified with respect to unethical activities like claims by clinics, medical practitioners and other related infringements or violations to the health-care customs.

- **Transportation Engineering**

Information-technology practices have made a huge contribution in carrying out many engineering disciplines. In addition to this remarkable offer, data mining has been equipped as a means of further facilitating information within the transportation industry. Its application is known to assist transportation systems and agents with problems related to the nature of the activity. Some of the issues reportedly claimed to be properly controlled as a result of the data mining integrations include road accidents, traffic management, pavement management, trafficking and general transportation tasks. In essence, data mining has helped society in predicting or determining the patterns of information that are associated with transportation management so as to monitor the conveyance of people and goods within the area. Companies invoke the processes of data mining to control the distribution schedules among warehouses and outlets as they make a consistent analysis of loading patterns from time to time. In a wide region, graphical information for data transportation is also accounted for when data mining protocols are induced to the system.

- **Telecommunication**

This is claimed to be one of the first industries to incorporate the process of data mining technology. It is also one of the pronounced sectors that deal with a vast amount of data that are continuously generated within the operations. To manage the tremendous data, telecommunication companies employ data mining to increase efficiency as they provide solutions to problems. These data include customer-information data, network data, and call-detail data which further expand into numerous quantities. Just as logic defines data as the base of telecommunication, the act of mining is applied to data operations by extracting and applying the outcomes to areas where needed. There are several significant reasons as to why data mining plays an imperative role in this industrial sector.

- A high acquaintance of customer profile is relatively crucial so data mining is used to study the profiles and behavior of customers in order to easily account for their usual wants from services.
- Beyond familiarizing customers in the business, working to retain them as they patronize the company is equally essential. Data mining tools tend to help study and research customers' database with respect to how to deal with each one, ideally.
- The detection of fraudulent activities is an integral aspect of security in every telecommunication industry which is why data mining measures are applied to analyze such an event.
- Products and services that are highly favored by customers can be deduced with data mining technology with optimal results. The analysis of these transactions gives the business information about popularly demanded goods and services.

2.2. THE CHALLENGES TO DATA MINING

Despite the evidently powerful tasks that data mining entails, there are still some critical downsides to the optimum efficiency of upon implementation of its objectives being description and prediction of data-information. These challenges are generally classified under the processes of data mining including its particular performance, strategies, and methods used in a system. Notwithstanding, most issues developed or associated with modern data mining technology can still curb to the most minimal effect that is attainable. Some of the commonest challenges faced by this activity are elaborated below.

- **Complexity**

In one saying, data mining is complex because of the intrinsic complexity of data. As the revolution of information and technology continues to expand exponentially, data mining algorithms tend to become more sophisticated. This can be acknowledged factually from the constant change and accumulation of data magnifies over time via various media like social networking. Consequently, the need to use extract and process data in accordance with the identification of relevance for multidisciplinary areas of information is deemed indispensable. The increasing power of I.T. does not dispense with the need the need to obtain its important information from its raw form that keeps accumulating but still aggravates the complexity of this task.

- **Distribution**

Unprocessed information is typically stored on different system databases within distributed virtual platforms such as the internet or software application. The act of using data mining processes to gather all data into centralized storage can be daunting as a result of the technicalities involved as well as organization for easier reference. Normally, users understand that distribution plays an essential role in the processes of data mining, especially the requirement of large data sources in repository and computation time. In order to make these data sets scalable, cumbersome efforts and mechanism cognition must be applied to properly distribute processed data among several networking sites on a regular basis.

- **Inconsistency and noisiness**

Data mining deals with a great volume of data extraction and identification for making good depictions and predictions for a specific system. Now, accumulating all these required data in such ample quantities is prone to containing inaccurate and unnecessary (i.e. inconsistency) data sets for future classification. Particularly, noisy data is a figurative terminology which defines data initially collected in huge amounts to possess additional meaningless information. It is also referred to as data that is corrupt and cannot be coherent or interpreted correctly. This property has been a major challenge to the efficiency of data mining because it can lead to a false sense of accuracy or conclusion upon analysis.

- **Data Visualization**

This deals with the process of transferring processed data in a way that can be simply deciphered by a viewer or interpreted by a system. The data involved in this process are conventionally in vast sizes and it is invariably meant to undergo the creation of approaches for depicting abstract information in a logical representation. Practically, the visualization of data mining or visual data mining brings about engaging in decision-making schemes that include selection, transformation and then the representation of data that facilitates data exploration and knowledge. However, this technique in data mining is very difficult to portray as the input data and output information has to experience complex mechanisms for optimum performance.

- **Privacy and Security**

A major concern related to security during the application of data mining processes is that the individuals are usually unaware of how their respectively mined information is used in a system platform. It is believed that the methodologies associated with data mining on a person's information should be made known to such an individual. This is because several people who have luckily become aware of data mining processes claimed that data mining practices are relatively intrusive. The typical nature of every data mining technique regards a search for initially unknown patterns in data wherein potential information of individuals are unravelled after being identified and analyzed thoroughly. In essence, data mining leads to serious problems of security and privacy due to its infamous ability to extract sensitive data and process into valuable information.

2.3. THE UPSIDES TO THIS ACTIVITY

Though data mining has some limitations that may tend to hinder its application in some specific business sectors that cannot afford to undergo the grave tasks and relative compromises, the advantages of mining still portray more captivating promises that include measures to reduce the challenges associated with it. Significant performances employed to boost and improve the processes of data mining in modern computer technology have been achieved in many cases.

- **Decision making**

This is a factor that is classified under epistemic business analysis which deals with knowledge gotten from available data sets with the aid of data mining tools or techniques. The intelligent decision methods are made at optimum through computer-based processes that are used for identifying and extracting relevant business data. Ideal decisions are almost invariably made when the right business intelligence schemes are applied. This tends to bring about an increment in organizational achievements, maximization of profit and security of patrons.

- **Predictions and Prospects**

Numerous manufacturing industries tend to generate and update production schedules in order to increase their productivity and reduce operational costs. As a result, high demand for deducing future outcomes of business affairs is needed as often as possible with high accuracy. The basis for data mining patterns that are used to create predictive information involves the production planning and scheduling of the organization's activities. Subsequent schedules and planning help a business to identify unforeseen resource conflicts that are further curbed to protect its system and enhance its desired goals.

- **Development of Products and Services**

Intrinsic business protocols that employ the technicalities of data mining include how industries normally devote a rational amount of their budget to products and services development. This is ultimately done in order for the enterprises to distinguish their goods and services that are demanded greatly with respect to other inferiorities. Data mining analysis provide informative patterns and representations as to why the specific products are popularly requested or the common services that are often preferred. The business then becomes acquainted with the nature of those particular trends and thus, develop similar or better means to facilitate the quality of other goods and services.

- **Acquisition and Retention of Customers**

Data mining is a process that allows its predictive analytical means to increase customer acquisition and retention. The conventions of data mining express the paradigms of business

intelligence by using data to ask questions and investigate hypothesis. Concerning the methods of frequent acquisition, the context of inquisition patterns can range widely in various intriguing customer acknowledgements and situations. This naturally makes avid buyers return to seek and discover more about whatever new information the business has to offer from time to time, hence depicting the law of customer retention stratagem.

- **Cost Reduction**

This is the most conspicuous advantages of data mining technology in every industrial sector. Data mining can be used to reduce costs in business exercises and attain better efficiency over a brief period and this makes companies to constantly increase revenue and attain competitive advantages. The major drive to data accumulation for reducing business expenses gives real-time information as opposed to traditional means. Some of these profitable data mining patterns are asserted to a company's supply chain which may typically include plan loads, optimize interaction and communication, strategize conveyance of services, and find opportunities to freight costs in the process.

3. CONCLUSION

The application of data mining processes in several industrial sectors depicts an emerging trend that has drawn the attention of casual information-technology practitioners and academics for the major purpose of understanding the patterns of behavior of associated users as well as patterns of information on resource usage throughout its methodology. The literature review explains references sources of data mining information to facilitate and ease interested individuals that seek further ramifications of this activity with respect to related paradigms. This thesis approaches theories and illustrations that engulf important information about the scope, concept, challenges and benefits to the processes of data mining by focusing on captious and faceted details and facts about the essence of the activity in contemporary business.

REFERENCES

- Han, Jiawei & Kamber, Micheline. (2006). Data Mining Concepts and Techniques (2nd Edition).
- Rajan, Pappu. (2015). Clustering Methods and Algorithms in Data Mining: Concepts and a Study. Journal of Computing Technology. 4. 8-11.
- Tomoyuki NANN0, Toshiaki FUJIKI, "Automatically Collecting, Monitoring, and Mining Japanese Weblogs", WWW2004, May 17–22, 2004.
- Ralph Gross, Alessandro AcquistiH. John Heinz III, "Information Revelation and Privacy in Online Social Networks (The Facebook case)", Pre-proceedings version. ACM Workshop on Privacy in the Electronic Society (WPES), 2005.
- Neha Mehta, Mamta Kathuria, Mahesh Singh, "Comparison of Conventional & Fuzzy Clustering Techniques: A Survey", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, April 2013.
- Mariam Adedoyin-Olowe, Mohamed Medhat Gaber and Frederic Stahl, "A Survey of Data Mining Techniques for Social Network Analysis", Journal of Data Mining & Digital Humanities, 2014.