

Data Mining and Decision Making Applied Model KhashmAlgerba Dam in Estrin Sudan

Dr. Alshafie Gafaar Mhmoud Mohmmed

Computer Science Department, *College of Science and human studies Al-Ghat,*
Majmaah University, Saudi Arabia

*Prof. Saife Eldin Fatoah Osman, Dean of Emirates Collage of Science &
Technology, Sudan*

Email: dr_shaf@yahoo.com

saifefatoah@hotmail.com

ABSTRACT: *The use of Technical Data Mining enables discovering and concentrating on the most important information of database in all fields. It also leads to future predictions, explore behaviors, and identify directions that ensure taking decisions effectively. This paper will shed some light on data mining, and the extent to which it contributes to taking decisions effectively.*

I. INTRODUCTION

The current era is distinguished with sophisticated technological revolution such as the Internet and Digital Economics. The flood of data and its rapid development suggests that it is difficult and even impossible for an analyzer to extract valuable information by using classical entrances to analyze primary data, especially with a large amount of data stored in databases and storage devices. There is thus an urgent need for developing storage tools, data analysis software, in addition to increasing extracting information and knowledge.

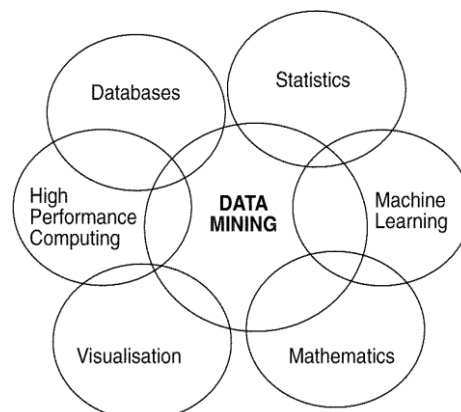


Figure1. Source: <http://blog.naseej.com>

Consequently, it appears that data Mining is able to extract information from large databases. This new modern technology is considered a breakthrough in information revolution. It helps companies and organizations discover and concentrate on the most important database. Mining Technology concentrates on making accurate future predictions and discover behaviors and aptitudes, which lead to taking appropriate decisions in a suitable time. This technology is however more complicated when used for discovering information in databases. It is related to a large extend to the operation of another development, which is very important in data storage. Most of the companies nowadays tend to discover information systematically and methodologically, which is considered as a backup that works best by activating effective systems to achieve distinctive competitions.

First: The concept of Data Base Mining:

The concept data Mining emerged in the mid of the nineteenth century in the United States of America. It includes both static and technology (Database, artificial intelligence & Machine learning). There are several yet interrelated definitions of data mining. One of these definitions is "the automated discovery of common pattern hidden database" (Cite). It also refers to the "procedures of intellect precise interactive analyses and series for runners of activities when using these procedures in decision taking and making appropriate activities on the benefit of the task that they are responsible for and the institution where they work" (Cite). Another similar definition is "it is the analysis of a large amount of data so as to find bases examples and models which used as guides for decision makers, and predict future behavior" (Cite). Moreover, it refers to "the analysis of a large size of data for the search of possible relation and summarize data in a new forms in order to be comprehensible and beneficial for its users" (Cite).

Based on the aforementioned definitions, one may realize that that Data Mining aims at extracting hidden information in databases. It helps institutions discover and focus on significant information found in large-scale databases. Such technology has become a research interest for many researchers to take advantage of. Furthermore, the rapid increase of data warehouses has made it necessary to explore effective alternative techniques for extracting information and knowledge from this crowded data, which are used for problem solving and decisions taking using modern computer applications. This form of data extraction is a new sophisticated technology for making computers act as human beings, perform as human workers. According to (Cite), it refers to "the idea of the discovering and mining data with sophisticated methods in problem solving and decision taking processes" (Cite). Moreover, this concept is considered as "a fundamental step of knowledge discovering by data baser" (Cite).

Second: Factors of Data Mining Revolution

Factors of data Mining revolution can be divided into two main factors:

1. Predictive Model.
2. Descriptive Model.

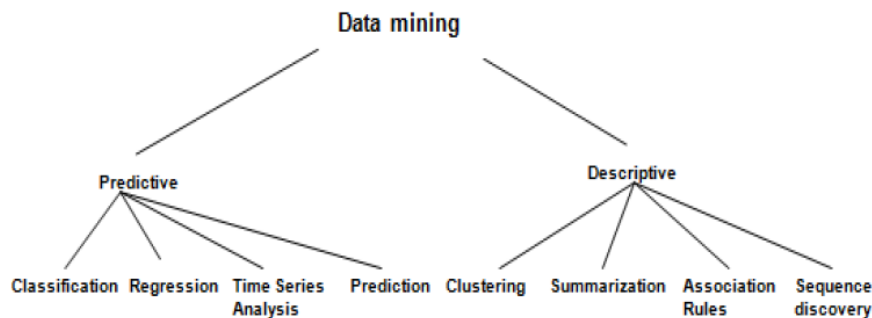


Figure 2

II. DATA MINING**Types and tasks of Data Mining Technology**

There are two models of Data Mining:

1. Predictive Models
2. Descriptive Models

The predictive Models help predict the value of some characteristics such as the predictions of water levels time series. Time series can be defined as a phenomenon value in a series of historical sequence days, months and years. The purpose of studying time series is to analyze a set of these series and the accompanying changes made by the phenomenon, which takes place through a period of time. It also aims to analyze the causes, results, and predictions depending on an idea that took place in the past and continued to the future. Another reason for studying time series is to make use of its analyses so as to predict the future.

To perform this analysis effectively, a growth scale, which interprets a mathematically, should be used so that the growth grade of time series, which is the purpose of the study, can be limited. Moreover, the general deviation of these series can be determined in addition to limiting its value that leads to the future.

Youssef (2002) points out that "a group of models are based on a hypothesis which proposes that the behavior of a future phenomenon depends entirely on its past behavior. Similarly, the prediction value also depends substantially on the phenomenon value in the past. For example, one of the widely recognized models is Jenkiz Box Model, which depends on a certain time series. However, interpretative variables, such as time series models and the falling models, are used for making perditions. These two models take the following form:

$$Z_t = f(x_t, \beta) + \delta$$

Z_t represents a variable used for predicting in the future.

x_t represents the independent model or interpretative model, which functions as a cursor of time (e.g. Trigonometric functions – polynomial functions).

β represents parameters used for estimations. These parameters should remain in downhill models and also in series time model analysis even though this hypothesis has been deleted.

δ represents the error of estimation.

The descriptive Method has been divided into two categories: -

- 1- Held models which permit the assembly of individuals, actions or products in clusters.
- 2- Creative Models which permit determination of relationships.

There are many tools for Mining in Data such as: -

1- Summarization: -

It refers to the method that transforms huge data into summarized matrices that provide a general description of the variables and their relations such as averages, totals, and statistical descriptions, which help generate the mean, median, mode, standard deviation. Although these standards provide large summaries, they often neglect details related to the consumer behavior.

2-Classification: -

Classification represents the interpretation or prediction according to individual specialization through other characteristics that can be undertaken through conducting classical statistics such as downhill and classification analyses or depending on modern methods such as link forces, depending on solutions, or neural networks. Examples of classification methods can be seen in the application of knowledge discovery, which includes directions of financial markets and electronic determination of important things (e.g. Database).

3-Prediction: -

It deals with estimation or classification processes regardless of how data have been classified based on future prediction methods. Prediction is used for determining future values of the dependent variables quantitatively. Some of the traditional tools used in prediction processes are downhill and distinctive analyses. However, there are more effective and accurate modern methods including but not limited to linking forces, decision tree, neural networking, and inherited algorithms.

4-Clustering

It classifies the population data into homogenous groups, representing individuals, families, or friends depending on the information included in the total variable. The purpose of this clustering is to develop marketing programs designed in an average of customers themselves. The flakiness assembly method is used in most cases by the aid of statistical flakiness method or by those of the decision tree, the neural networking, or internet algorithms.

5-Rule Analysis:-

Rule represented by the search for relations or links exist between different characteristics. Nonetheless, rule analysis refers to a group of methods used for linking purchase patterns with sector or through time (e.g. the use of the analysis used in market baskets to get benefit of the potential information of goods purchased by actual customers to predict the goods actually purchased if there is especial offer for them or if they have been informed by the goods).

6-Dection change and deviation: -

It depends on the most discovered variables of data through previous majors or standard values.

Fourth:-

Levels of Data Mining process:

Levels and steps of data Mining process can be summarized as follow:-

A-Business understanding is the first issue of knowledge discovery that deals with understanding the problems that face business. For example, how can the greatest beneficial of data mining be achieved which require a clear limited formula of business goals.

B-Data Understanding

The understanding of the nature of data is the most important factor in data mining processes and knowledge discovery for gaining a deep understanding of the given data. It refers to the help of designers to use tools Algorithms used in limited precise issues. This provides great opportunities for success in addition to raising the efficiency and qualification of knowledge discovery systems.

The most essential steps of understanding data process can be summarized as follow: -

1- Data collection: -

It deals with limiting sources of data used in the study included over general data such as taxes.

2- Data description: -

It focuses on the description and contents of files or charts.

3- Explanatory analysis of data:-

It uses the photography direct analysis method, which leads to the Primary analysis of data. This point is crucial, as it is based on the development of hypothesis related to the problem of the study.

Data preparation:-

It includes the following:

A. **Selection:** refers to the expected variables collection and sampling.

B. **Variable construction and transformation:** is an essential process for contextualizing new variables to construct effective models.

C. Data Integration: -

The collection of data in data Mining studies is considered here and database with multipurpose in only one database is also taken into account.

D. Data formation and designing: -

This step is concerned with the formation and order of data fields, as it is required in the model of data mining.

1-Model Building and validation: building a precise solution model takes place through the process of trial and error. In most cases, this process requires an aid from specialized individual in data mining so as to select and check different alternatives for achieving the best models that can be solve the problems under investigation.

5- Evaluation and Interpretation of findings.

When the model is contextualized and its validity and reliability are en/*sured, then the process of confirming data reliability, which fed by the model, is initiated. Once the outcome of this data is clear, then the expected findings can be compared with the actual data under study in order to reach the precise model.

Model Deployment:-

This stage includes developing an organizational model that helps take decisions. The valid model should then achieve users' satisfaction. However, the model can be selected through directed study or a simple sample model from a population sample.

Applied Frame

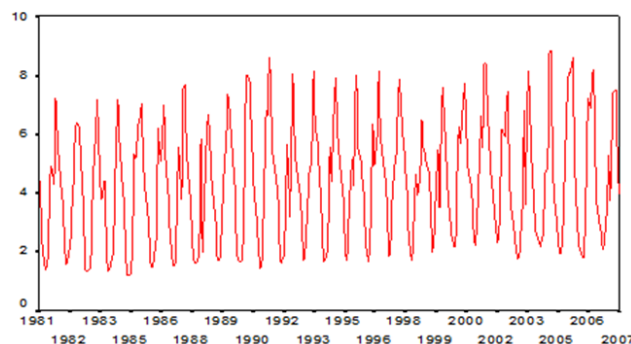


Figure3- A chart explains the downhill model series of a wasted water I a channel. explains the data time series for the wasted level in the channel.

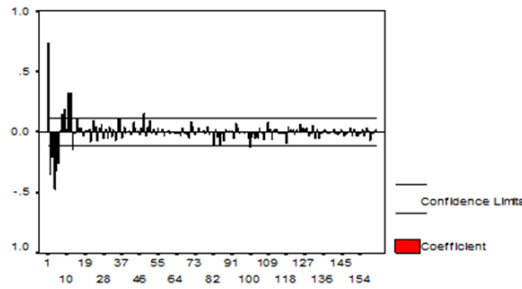


Figure 4: Based on the chart series above, it seems that there is no general direction. This means that the series was stable in the middle. Moreover, we noticed that there is no change in the series. The stability test of a wasted series in a channel.

Dis_ch	B	SEB	T-Ratio	Approx.prob
AR1	.7304185	.03788795	19.278387	.0000000
constant	4.3307376	.27596051	15.693324	.0000000

Residual Variance: 1.8231508

Chart 3 Explains the data time series for a level of a wasted channel

Figure 3

Based on the chart above, we can notice all the co-related efficiency located inside the granted period, which refers to the constant time series (Random data).

A model of downhill series of a wasted water in a channel

$$X_t = 4.3307 + 0.7304 X_{t-1}$$

Explains the Downhill Model from Selected class of a Series Wasted Water in a Channel.

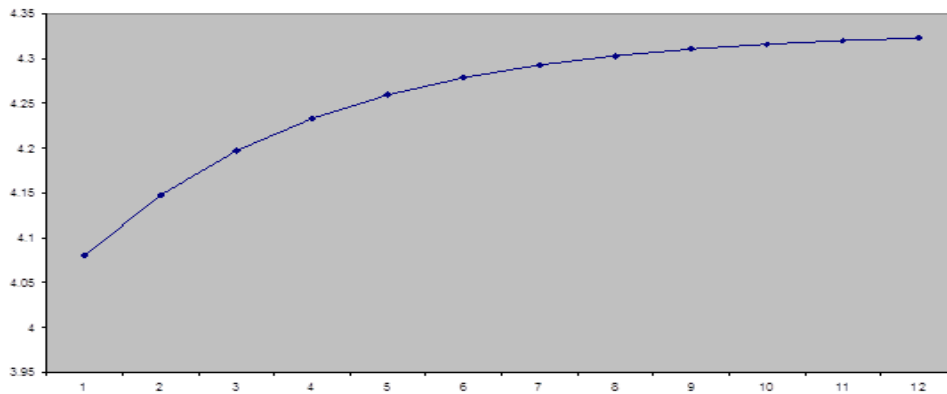


Figure 5

Explains the prediction of a channel wasted level.

Expected figures of a main wasted water in a channel

We notice that the expected figures had increased, which influences the irrigation of lands. Consequently, it reflects on the production of the scheme.

III. CONCLUSION

The mixture of information technology and statistical method leads to providing appropriate suitable prediction of future behaviors. Then, it seeks for suitable solutions for the problems before they occur. If it exists, the prediction of general and refresh in many domains using the technical of data mining, which forms a level of a whole knowledge discovery in database. In sum, it becomes one of the most important issues that make countries engage their institutions.

BIBLIOGRAPHY

- [1] Bashir A bbase .Digital Administration in Field and A applications . Emirates Centre of Studies and Research straggles .Abu Dubi .2005.Pages.84 .91-89.92.93.93-94.
- [2] Bazsalica M.Naim.p.Data Mining pour Le Web edEyrolles .Paris .2001.pages.01.68-69.66.68.

- [3] AbdAl-star Alali ,Abbass IBRAHIM Gedanhaly . GassanALamri. Introduction to KMWLEDGE Administration .DarrAlmsyra for publishing , printing and distribution. First Edition . Oman pages 177.100-102 .99-98.
- [4] Berry J.A.M Linoff.g.s Data Mining Techniques for Marketing Sales and Customer relationship Management 2nd Edition .Wiley Publishing INC.indianapolis.2014 .P .10