

# Training Students with Data Mining and GIS

Ali Taha Yaseen<sup>1\*</sup>, Zaki S. Tywofik<sup>2</sup>, Mohammed Q. Mohammed<sup>3</sup>

Department of Computer Science, Dijlah University College (DUC), Baghdad, Iraq  
University of Information Technology and Communications

\*Corresponding Author Email: [Ali.taha@duc.edu.iq](mailto:Ali.taha@duc.edu.iq)

## Abstract

Present in colleges to day the students have course of the data mining which is the procedure of mining knowledge from data to solve problems, gain knowledge and set expectations. Students in the course have acquired a great knowledge and understanding of the processes involved in extracting data through the experience gained in data extraction applications. The course has shown that students' education is appropriate and can be successful. In this research students apply their expertise on data taken from the Iraqi Health Ministry to persons infected with hepatitis in Iraq's cities to discover knowledge about the virus, and that the use of geographic information system (GIS) to display the result to extract the data in the map of Iraq, which give force for them work and that new way to combine data mining with GIS.

*Index Terms:* Data Mining, GIS, SOM, Classification

## 1. Introduction

Has revolutionized the ability to digitally store data in our community databases. Data that was filled with our files in the complete Cabinets can now be stored in a desktop on different forms of digital meters. The technological breakthrough has led to data storage in a huge amount of data today. The collection of information biological of medical are about patients lay in the filed bioinformatics. The field of bioinformatics relies heavily on work by experts in statistical methods and pattern recognition. Students come to bioinformatics from many fields, including mathematics, computer science, and linguistics.

Data is created faster than our ability to understand and use it. There may be hidden forms within this data that contain potentially useful information. The data tries to extract these patterns and detect this information. This field has become useful in the business world today and will prove successful operation of businesses for governments and other organizations in the future there greater data more important it is to obtain valuable data. Data extraction, the key to obtaining this valuable information [2].

Data mining is highly useful in the following domains first extract data . Data extraction helps students know and analysis and determine the there work which helps maximize profits. Data extraction can be used at any time for data analysis for example health and human services students need geographic information system (GIS) technological skills to analyze the geographic elements of hepatitis and human services science. GIS technology has progressed to a high level of sophistication and, therefore, is relevant to nearly every emphasis of the hepatitis. The rapid evolution of GIS as an analytical tool has created specific applications for multiple health-related fields from environmental and global health

to hospital management.

Many applications of data mining successes clearly demonstrate that data an important and useful tool to invest in the future. There is a growing demand for IT professionals trained to extract data. It good way to use GIS ( A Geographic Information System) feature to display, store, retrieve, manipulate information to produce mat with its associated attributes. The students can use GIS basic mapping terminology how use to produce maps and its information.

## Literature Review

Since the beginning of the last century, preachers and sociologists have identified social and economic differences in cities by observing residential patterns. Poverty maps in Charles Booth in London are a classic effort to map this social scene. Booth, who worked between 1886 and 1903, classified the streets of London using seven categories: the rich, the rich, the comfortable, the mixed, the poor, the very poor, the wicked, and the semi-criminals

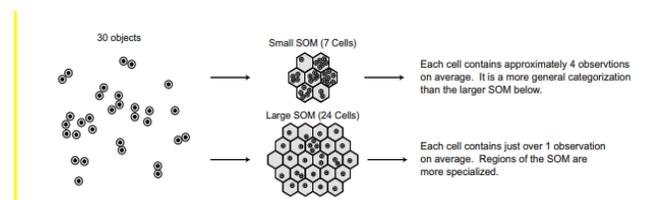


Fig. 1: Self-Organizing Map Size

Skopin and Hagelman (2003, 2005) used large networks to explore the demographic "clues" of different regions of Texas. In this context, a large network separates similar areas into unique areas on

the distinctive map. In this work, SOM is trained in 30 years of census data. Their large map allows them to examine how the characteristics of census tracts change over time by observing how individual tracts move around the output area over time. Medium networks are a compromise. They allow areas with clearly defined characteristics to be formed on the map, but general data on these areas can be issued because they contain a reasonable number of census areas.

### 2.1. Assessments of Faculty and Student Needs in Data Information

Literacy As well as electronic research, data literacy is not new in itself, but brings together experiences, parts of research methods, information and other information based on reading and writing, and a computerized approach to provide perspectives and more complete and shared techniques for future researchers. As electronic research encourages researchers from a variety of disciplines to collaborate to promote scientific knowledge [1] - [3], the discipline and desk staff must work together to identify the skill sets that students should know and develop best practices to transfer these skills to students. In order to extract samples, trends and interesting information from the data, each participant can help greatly in the educational process to better understand the learning process and focus on discovering, discovering and interpreting the educational phenomenon. [14]–[23]. As shown in fig. 2.

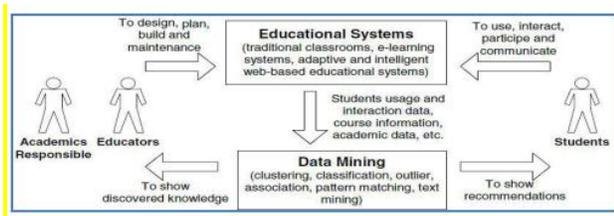


Fig. 2: The cycle of Applying Data Mining in Educational Institutions

Fig. 3 shows the method used to classify relevant attribute items based on different levels of abstraction and different thresholds. The process for each variable is iterative. The number of iteration depends on the minimum confidence threshold defined by the user and the actual degree of association mined from the data. Each further iteration provides information that is more specific. Algorithms can be developed by using Avenue scripts to carry out the iterative process [21].

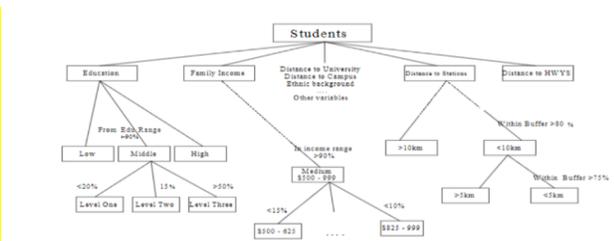


Fig. 3: Generalization and classification of relevant data items into different levels.

### 3. Development of the Course

Students in college of computer science or university in order to apply data extraction use computer and get a strong result rapidly in the computer provided [22]

The University's Computer programs are used to input information to plot the portfolios and their locations to determine the presence of disease hepatitis virus. In addition, the computer can be used to process the data on the data associated with the maps and the complex equations and calculate the results,

Computer helps to sort data in a way that is useful for extracting data. Data repositories contain large of data from many locations. These data are organized to analyze and give information about disease [23] as opposed to traditional database use University student, through the course of directed studies focus on using and its application [24] on map.

The Goddard Institute written by C ++ is developed. Their jobs are growing rapidly. Researchers can input environmental data in the GIS to obtain the result of spatial analysis that can show how widely distributed data are released from the hepatitis virus geographically. Besides, data mining can automatically mine hidden knowledge and analyze / extract knowledge from raw data, respectively. If it can be put into use with GIS, the hidden meanings or rules embedded in the environmental data can then be more thoroughly and accurately detected.

### 4. Application of the Course

Stimulate the study of data extraction, students investigate the data extraction of their choice and submitted their findings to the chapter. The class was supposed to be 12 students and made presentations in groups of two or three. Topics were data storage, data mining in the field of e-mining text, software data extraction [5], predictive modeling. This approach at the beginning of the chapter requires students to look at extracting data in newspapers and on the Internet and gain a better understanding of the breadth and depth of the field. Students quickly learn that data mining is an important new addition to computer science that is applied in many places.

The underlying factors and concepts of data mining were then studied. Lectures were given in accordance with chapters 1 to 5 of the Mining Mining data book. Compare the separation of classification, assemblies, aggregation, linear regression, and existing approaches for example to data mining. Students learn how to evaluate the results of data extraction algorithms and predict their performance. Also studied statistical methods such as cross-examination and pot strap. Weka proved to be a very useful educational tool. All of the algorithms listed above, and many more, are provided by WEKA. Students were able to implement these data extraction algorithms based on practice data sets supplying WEKA. In fact, vision algorithms work to analyze results and promote concepts that are learned by students.

During the remainder of the course, students applied in groups of two or three what they learned about data mining to a self-designed project. The focus of the The project was to implement the extraction of data on real data grouped in the form of a data warehouse and try to obtain new knowledge of the data. Project design industrial data collection models, students learn practical knowledge in data extraction through practical experience.

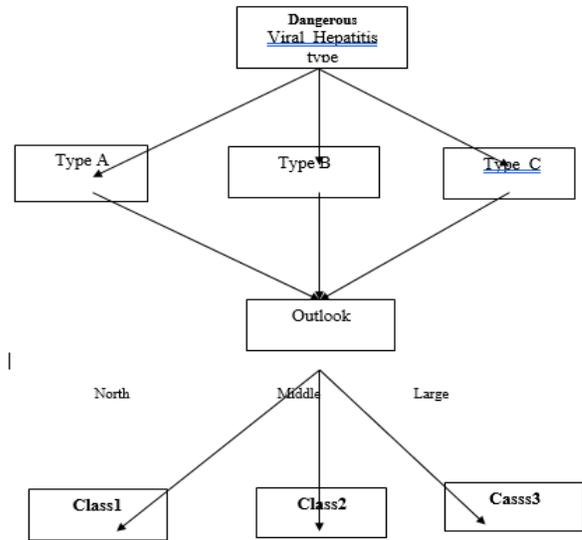
Contained in the store of data of the Iraqi Ministry of Health to all serious illnesses and disease such as regular viruses affecting humans in Iraq, and these data will focus on liver disease virus spread in all the portfolios of Iraq, and this statistical information has been taken for the year 2008 and these viruses, the data on the types of properties and will take Three types of viruses, which are c, b, a and c is the type of the most serious and the place of deployment, and these data are small for the data storage business.

Therefore, it is necessary to require access to their research suggested that students do their project on the design of their self -. Mwalimathm by the Iraqi Health Ministry and the Internet to the

knowledge of the characteristics of this type Alvarus and speed of deployment and the degree of heat and the nature of environmental and life cycle.

Where the students began to understand the data on the types of virus and liver, and formulated the questions began to take on hepatitis viruses and the creation of a relationship between these data and the effectiveness of the human disease or injury affecting the liver. Some students have taken a may ways, and chose to their own data and the design of algorithms for use in data mining. For example, a group of students to amend article The same group wanted to display them result in map graphics which give new application of them work that combine the data mining with GIS.

There is a relation between data geted from Iraqi health Ministry is, for example, that people infected with hepatitis virus differ archive also love to prepare them in the table No. 1 in the three types of the disease as there is a relationship between the spread of the virus and the different temperature and prevention of this disease. Table1 Alanchkas found that the incidence of the disease varies from one region to the other, where the liver infection in the northern region than other regions, as evident in Figure (a), Figure (b) and Figure (c) as well as increasing the incidence of the disease in areas with high population density, as is evident in the maps of Iraq where the numbers of liver infection [25].



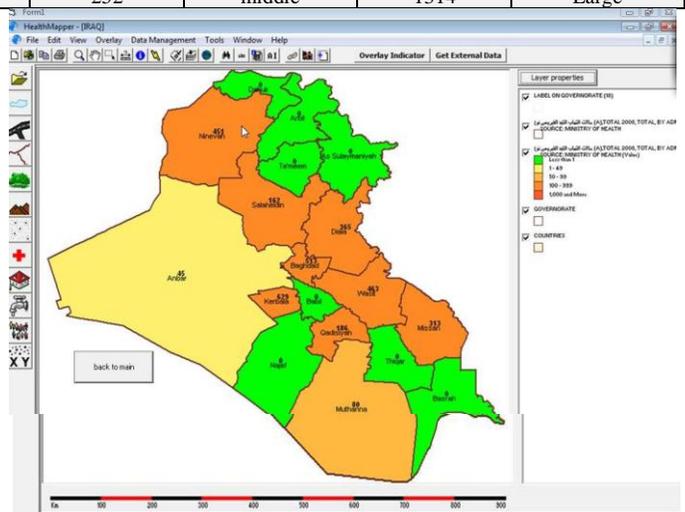
**Table 1:** Number of people infected with hepatitis virus in the city

City	Out Look	viral Hepatitis type A	Dangerous Class 1	viral Hepatitis type B	Dangerous Class 2	viral Hepatitis type C	Dangerous Class 2
Dahuk	North	0	Less	37	middle	436	Large
Arbl	North	0	Less	1	middle	231	Large
Mousal	North	451	Less	229	middle	288	Large
Karkok	North	0	Less	294	middle	1097	Large
Sulaymaniyah	North	0	Less	0	middle	0	Large
Salahedin	Middle	162	Less	195	middle	391	Large
Diala	Middle	365	Less	65	middle	398	Large
Baghdad	Middle	513	Less	1407	middle	3824	Large
Anbar	Middle	45	Less	111	middle	402	Large
Karbala	Middle	629	Less	88	middle	783	Large
Babil	Middle	0	0	315	middle	1542	Large
Wasit	Middle	463	Less	134	middle	397	Large
Qadisyah	South	186	Less	79	middle	262	Large
Missan	South	313	Less	54	middle	363	Large
Najf	Middle	0	Less	207	middle	843	Large
Thiqar	South	0	Les	95	middle	304	Large
Mutanna	South	80	Less	27	middle	1043	Large
Basra	South	0	Less	252	middle	1314	Large

**4.1. Computing information entropy**

Basic formula for computing the entropy  
 Is large.  $entropy(S) = -p1logp1 - p2logp2K - pnlogpn$   
 In the dataset we had two target classes: *No hepatitis virus* and *There is hepatitis virus*.

Entropy For *No hepatitis virus* in Class 1.  
 $Entropy(A) = -4/18 log_2 4/18 - 2/18 log_2 2/18$   
 $= 0.22x2.1 + 0.11x3.18$   
 $= 0.462 + 0.351 = 0.78$   
 Entropy For *No hepatitis virus* in Class 2  
 $Entropy(B) = -0.2/18 log_2 (2/18)$   
 $= 0.3$   
 Entropy For *No hepatitis virus* in Class 3  
 $Entropy(C) = -0.1/18 log_2 (1/18)$   
 $= 0.1$



**Figure 1 (a):** people infected with hepatitis virus type A

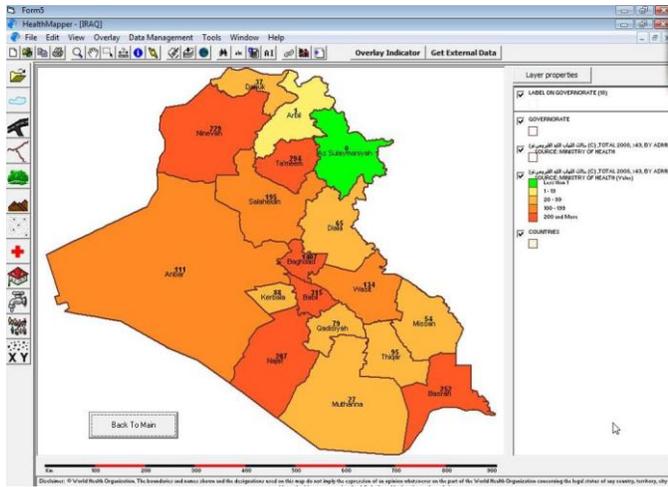


Figure 2(b): people infected with hepatitis virus type

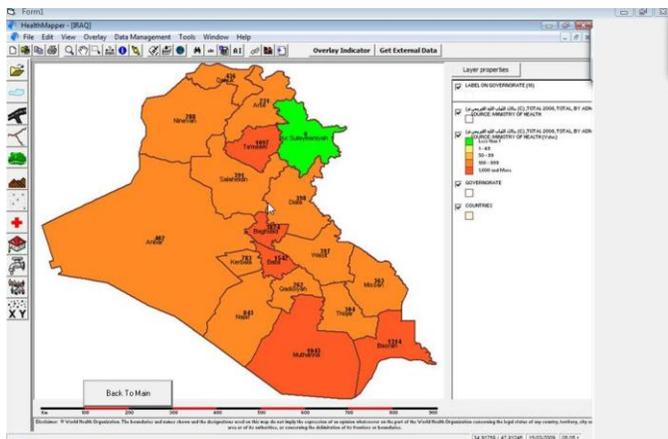


Figure 3(c): people infected with hepatitis virus type C

### 5. Conclusion

Students were mining data of course success. Understand the basics process of database. The students use GIS to limit location of the hepatitis and produced as the application in maps, which led to the acquisition of practical knowledge of data extraction techniques to develop them knowledge. A large number of hepatitis data located on maps, which led students more experience for factures how deal with real data mining and how to use it in many application which learned more important process facilities of GIS and always produces interesting results. This process has shown in learning that students can recognize virus database extraction and how they learned to solve problems in many areas of working life Students use the result of applying the algorithm to extract the data to display the number of infected hepatitis virus in the city result by using the GIS program to give them the strength in the work.

### References:

[1] Witten, I., Frank E. (2000). Data Mining. Academic Press.  
 [2] Sullivan, T. (2000) "EMC's Ruetters warns of data explosion." InfoWorld.com <http://www.infoworld.com/articles/hn/xml/00/10/04/001004hnruettgers.xml?spon sor=STORAGE>  
 [3] Shakeel, P.M., Tolba, A., Al-Makhadmeh, Zafer Al-Makhadmeh, Mustafá Musa Jaber, "Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks", Neural Computing and

Applications,2019,pp1-14.<https://doi.org/10.1007/s00521-018-03972-2>  
 [4] University of Waikato. "Weka 3 – Machine Learning Software in c++." <http://www.cs.waikato.ac.nz/ml/weka/>  
 [5] Knowledge Seeker Data Mining Tool. "Angoss Home Page." <http://www.angoss.com/>  
 [6] U. S. Forest Service. (1992). "The Eastwide Forest Inventory Database User's Manual." <http://srsfia.usfs.msstate.edu/ewman.htm>  
 [7] Whalin, P. (2000) "Data Mining Research."  
 [8] Ludwig, L., Flies, D., Wilson, A. (2000) "Data Mining Techniques Applied to the Relationship of Latitude and the Lifespan of Aspen Trees" <http://epoxy.mrs.umn.edu/~ludwigl/datamining/research.pdf>  
 [9] JOHN PICKLES (2003)"GIS for the urban environment" PRINTED IN USA CROWN5M 7/06DH PRODECT BY ESRI Ministry of health.  
 [10] Preeth, S.K.S.L., Dhanalakshmi, R., Kumar, R.,Shakeel PM.An adaptive fuzzy rule based energy efficient clustering and immune-inspired routing protocol for WSN-assisted IoT system.Journal of Ambient Intelligence and Humanized Computing.2018:1–13. <https://doi.org/10.1007/s12652-018-1154-z>  
 [11] A. Meri *et al.*, "Modelling the Utilization of Cloud Health Information Systems in the Iraqi Public Healthcare Sector," *Telemat. Informatics*, 2018.  
 [12] P. M. Shakeel, S. Baskar, V. R. S. Dhulipala, and M. M. Jaber, "Cloud based framework for diagnosis of diabetes mellitus using K-means clustering," *Heal. Inf. Sci. Syst.*, vol. 6, no. 1, p. 16, 2018.  
 [13] P. M. Shakeel, S. Baskar, V. R. S. Dhulipala, S. Mishra, and M. M. Jaber, "Maintaining security and privacy in health care system using learning based Deep-Q-Networks," *J. Med. Syst.*, vol. 42, no. 10, p. 186, 2018.  
 [14] M. A. Mohammed *et al.*, "Genetic case-based reasoning for improved mobile phone faults diagnosis," *Comput. Electr. Eng.*, 2018.  
 [15] M. Jarrar, M. S. Minai, M. Al-Bsheish, A. Meri, and M. Jaber, "Hospital nurse shift length, patient-centered care, and the perceived quality and patient safety," *International Journal of Health Planning and Management*, 2018.  
 [16] Shakeel PM. Neural Networks Based Prediction Of Wind Energy Using Pitch Angle Control. International Journal of Innovations in Scientific and Engineering Research (IJISER). 2014;1(1):33-7.  
 [17] M. T. Naseem *et al.*, "Preprocessing and signal processing techniques on genomic data sequences," *Biomed. Res.*, vol. 28, no. 22, 2017.  
 [18] M. H. Ali, M. F. Zolkipli, M. A. Mohammed, and M. M. Jaber, "Enhance of extreme learning machine-genetic algorithm hybrid based on intrusion detection system," *J. Eng. Appl. Sci.*, vol. 12, no. 16, 2017.  
 [19] M. H. Ali, M. F. Zolkipli, M. M. Jaber, and M. A. Mohammed, "Intrusion detection system based on machine learning in cloud computing," *J. Eng. Appl. Sci.*, vol. 12, no. 16, 2017.  
 [20] *INTEGRATING GIS AND SPATIAL DATA MINING TECHNIQUES FOR TARGET MARKETING OF UNIVERSITY COURSES | REQUEST PDF. AVAILABLE FROM: [https://www.researchgate.net/publication/228529975\\_INTEGRATING\\_GIS\\_AND\\_SPATIAL\\_DATA\\_MINING\\_TECHNIQUES\\_FOR\\_TARGET\\_MARKETING\\_OF\\_UNIVERSITY\\_COURSES](https://www.researchgate.net/publication/228529975_INTEGRATING_GIS_AND_SPATIAL_DATA_MINING_TECHNIQUES_FOR_TARGET_MARKETING_OF_UNIVERSITY_COURSES) [ACCESSED NOV 05 2018].*  
 [21] El-Halees, A. (2008), Mining students data to analyze learning behavior: a case study,  
 [22] Flies, D. (2001) "Designing & Implementing a Classroom Data Warehouse."MICS.  
 [23] P. Mohamed Shakeel; Tarek E. El. Tobely; Haytham Al-Feel; Gunasekaran Manogaran; S. Baskar., "Neural Network Based Brain Tumor Detection Using Wireless Infrared Imaging Sensor", IEEE Access, 2019, Page(s): 1  
 [24] L. Haoyu, L. Jianxing, N. Arunkumar, A. F. Hussein, and M. M. Jaber, "An IoMT cloud-based real time sleep apnea detection scheme by using the SpO2 estimation supported by heart rate variability," *Futur. Gener. Comput. Syst.*, 2018.  
 [25] S. K. Abd, S. A. R. Al-Haddad, F. Hashim, A. B. H. J. Abdullah, and S. Yussof, "Energy-Aware Fault Tolerant Task offloading of Mobile Cloud Computing," in *Proceedings - 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, MobileCloud 2017*, 2017.