

A Proposed Framework for Detecting and Predicting Diseases through Business Intelligence Applications

Soha Ahmed

Faculty of Commerce and Business Administration, Helwan University, Egypt
soha.ahmed2791@gmail.com

Ahmed Ibrahim El Seddawy

Business Information System Department, College of Management & Information Technology,
Arab Academy for Science, Technology & Maritime Transport
ahmed.bahgat@aast.edu

Mona Nasr

Department of Information Systems, Faculty of Computers & Information
Helwan University, Egypt
m.nasr@helwan.edu.eg

ABSTRACT

The Demand for healthcare IT and its analytics increases in the last few years. To improve quality of care (e.g., ensuring that patients receive the correct medication) which will help to improve the efficiency of clinical quality and safety, operations.

The Nature of the medical field is rich with information where there's a variety and abundance of data but untapped in a correct and effective manner to get the right knowledge. and therefore, the most serious challenge facing this area is the quality of service provided which means to make the diagnose in a proper manner at a timely manner and provide appropriate medications to patients because Poor diagnosing can lead to serious consequences which are unacceptable. And because there is a lack of effective analysis tools to discover hidden relationships and trends in data, so Health information technology has emerged as a new technology in health care sector in a short period by utilizing Business Intelligence 'BI' which is a data-driven Decision Support System. Which Was developed from 1990s to now, and gradually become one of the most important information systems applied in any sector. BI enables to deal with huge amount of data and extract useful knowledge to support decision making. Data mining 'DM' is a kind of data processing technology which can be regarded as a part of the BI system, but it can be also considered as an independent and integrated technology which can treat mass data and extract hidden relationships from it.

This introduction highlights the main importance of how to apply the business intelligence applications using data mining techniques to help medical professionals in healthcare sector rapidly diagnosing and predicting diseases of any patients not only this but also detecting the disease complications on the patient which will decrease the overall cost of expenditure that the country paid, briefly this is the central research idea which address the motivation for doing this research.

Keywords – Business Intelligence, HealthCare, Data Mining, Data-Driven Decision Support System.

Date of Submission: Dec 08, 2018

Date of Acceptance: Jan 30, 2019

I. INTRODUCTION

The introduction is divided into several sessions that address the basic concepts and define the main key terms in the research which are the Business Intelligence 'BI', Data Mining 'DM' and Electronic Health Record 'EHR'. Business Intelligence and Data mining techniques plays a major role in the medical field and the health care sector by utilizing the Patient Electronic Health Record because the Business intelligence is a broad category of methodologies, technologies, and applications for collecting, storing, analyzing, and providing access to data to help users make better and faster decisions.it also include the activities of decision support systems, query and reporting, online

analytical processing 'OLAP', statistical analysis, forecasting, and data and text mining.

First; Business intelligence solutions are used in many industries to gain insight from different data sources to make more informed decisions towards main goal of achieving efficiency and effectiveness. And is presented as a mechanism to ensure a robust and systematic approach to healthcare management with a goal of enduring impact on quality improvement and cost control. ^[2]

BI tools are technology that enables the efficiency of business operation by providing an increased

value to the enterprise information and hence the way this information is utilized. [19]

The basic characteristic has been described for BI tool for its ability to collect data from heterogeneous source, to possess advance analytical methods, and the ability to support multiusers demands. [24]

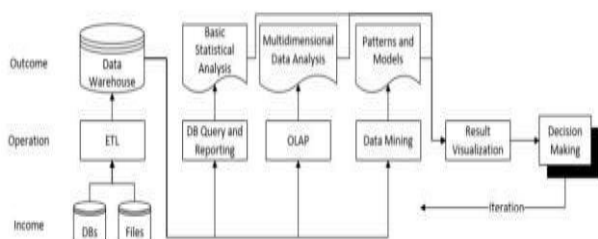


Figure 1: Traditional BI Process [25]

BI is composed of Extraction, Transformation and Loading 'ETL' system, data warehouse, database query and reporting system, online analytical processing system, data mining and data visualization system. [25] In this process, data are extracted from heterogeneous data sources, integrated and moved to data warehouse. Then, database query and reporting system, OLAP and data mining will run based on the data warehouse. After data is being analyzed, the results will be presented by visualization system.

Finally, these results will be used by organizations to support their decision making.

In our research the data will be extracted from the electronic health record as a source of patient medical record and health status and moved to the Data warehouse repositories then the Data mining play its role in finding patterns & hidden relationships, finally the results presented through the business intelligence visualization system

The Main Components of Business Intelligence

The following are the major components that constitute BI. First On-line analytical processing 'OLAP'; It refers to the way in which business users can slice and dice their way through data using sophisticated tools that allow for the navigation of dimensions such as time or hierarchies. It provides multidimensional, summarized views of business data and is used for reporting, analysis, modeling and planning for optimizing the business. [23] OLAP techniques and tools can be used to work with data warehouses or data marts designed for sophisticated enterprise intelligence systems. These systems process queries required to discover trends and analyze critical factors.

Second; Data Warehouse 'DW'; is the significant component of business intelligence and It is subject oriented, integrated. used to support the physical

propagation of data by handling numerous enterprise records for integration, cleansing, aggregation and query tasks. It can also contain the operational data which can be defined as an updateable set of integrated data used for enterprise wide tactical decision-making of a subject area. It contains live data, not snapshots, and retains minimal history. Data sources can be operational databases, historical data, external data or information from the already existing data warehouse environment. The data sources can be relational databases or any other data structure that supports the line of business applications. They also can reside on many different platforms and can contain structured information, such as tables or spreadsheets, or unstructured information, such as plaintext files or pictures and other multimedia information.

Third; Data Mining 'DM' and Knowledge Discovery 'KDD'; Is the notion of finding useful patterns in data has been given a variety of names, including data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing. The term data mining has mostly been used by statisticians, data analysts, and the management information systems communities. [6]

It has also gained popularity in the database field. The phrase knowledge discovery in databases was coined at the first 'KDD' workshop in 1989 to emphasize that knowledge is the product of a data-driven discovery. KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a step in this process. Data mining is the application of specific algorithms for extracting patterns from data. The distinction between the KDD process and the data-mining step within the process is That there's additional steps in the KDD process, such as data preparation, data selection, data cleaning. Data mining is used to describe knowledge in databases and it is a process of extracting and identifying useful information and subsequent knowledge from databases using statistical, mathematical, artificial intelligence and machine learning technique. Data mining technologies are used to discover the hidden information from the varying data sets. And it's broadly used in various real-life business applications like health care, Medicine, banking, Insurance, Fraud detection, Product retailing, Security systems etc. [12]

Data mining is a tool of business intelligence for knowledge discovery. The predictive power of data mining comes from concepts of machine learning, pattern recognition, and statistics and it automatically extracts concepts to determine interrelations and patterns of interest from large databases.

Data mining involves some complex data analysis tools to discover the previously unknown valid patterns and relationships in large data sets. The tools can be mathematical algorithms, statistical methods, learning algorithms etc.

Traditional data mining process as shown in Figure 2 compiles many previous ideas ^[25]

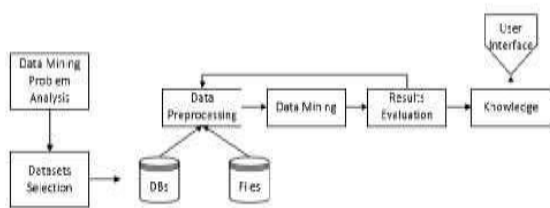


Figure 2: Traditional Data Mining Process ^[25]

In our research the Data mining takes the data from repositories in which the data are represented in a form of EHR using the tools (DM tools such as Waikato Environment for Knowledge Analysis ‘WEKA’, Konstanz Information Miner ‘KNIME’, Rapid Miner, Orange, ...etc.) to utilize the techniques (DM techniques such as classification, clustering and association rule) to get the desired after making analysis.

Data mining Techniques; the tasks of data mining can be modeled as either Predictive or Descriptive in nature. First the Predictive model; makes a prediction about values of data using known results found from different data while the Descriptive model identifies patterns or relationships in data unlike the predictive model. The Predictive model data mining tasks include classification, prediction, regression and time series analysis. Classification is probably the best classic understood technique for the prediction of all data mining techniques based on machine learning. Where it classifies each item in set of data into one or more predefined set classes or groups classification methods make use of mathematical techniques such as ^[20,21,22]

- Decision trees; is an algorithm that specifically used for decision analysis to help in reaching a specific goal and it's mainly act as a decision support tool like a model of decisions and their possible consequence
- Linear programming; is an algorithm used to achieve the best outcome and it's considered as a special case of mathematical optimization programming.
- Neural network; an algorithm which presented as a system of interconnected neurons which exchange messages between each other For example: given all records of patients who are infected with specific disease and predict who will probably infect with specific disease. Three common characteristics of classification tasks are; 1. Learning is supervised 2. The dependent variable is categorical 3. The model built can assign new data to one of a set of well-defined classes. ^[15]

For example, given classes of patients that corresponds to medical treatment responses; the form of treatment to which a new patient is most likely to respond to is identified. Unlike a classification model, the purpose of Prediction model is to determine the future outcome rather

than current behavior. Its output can be categorical or numeric value.

It includes the creation of models to predict target, or dependent variable from the set of explained or independent variables. Classification is the process of finding a function that allows the classification of data in one of several classes. Second the Descriptive model; serves to explore the properties of the data examined, not to predict new properties and includes methods such as Clustering, Summarizations, Association Rules, and Sequence analysis. Descriptive data mining is normally used to generate frequency, cross tabulation and correlation. Descriptive method can be defined to discover interesting regularities in the data, to uncover patterns and find interesting subgroups in the bulk of data. From the descriptive models are the Clustering technique which is the best used for finding groups of items that are similar. The goal is to group similar objects in the same cluster and different ones in the different clusters. Process of grouping determines groups of data that are similar, but different than other data. In this process variables are identified by which the best grouping is being realized.

For example: Make clusters for patients who have the same symptoms or take the same treatments ^[5], so it will be easy to predict the probable disease. Also, the Association Rules is another descriptive model which indicates patterns discovered based on relationships between items in the same transaction association rule also called “Relation Techniques”

It's a popular technique for market basket analysis because all possible combinations of potentially interesting product groupings can be explored. Association rule determine patterns, or associations, between elements in data sets, and are represented in the form of rules, or implications and We use it to identify patient's related diseases. The investigation of relationships between items over a period is also often referred to as Sequence Analysis.

Data mining technology can generate new business opportunities by providing these capabilities:

Automated prediction of trends and behaviors; Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data quickly. A typical example of a predictive problem is targeted diagnosing. Data mining uses data on past EHR of patients to forecast the most likely diseases could infect them and identify the populations/ patients likely to respond to a given disease.

Automated discovery of previously unknown patterns; Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of patient's historical medical records to identify the unrelated diagnosis that are often come together and the unrelated

treatments/medicines that are taken together and identifying anomalous data that could represent data entry keying errors. Business Intelligence and Data Mining could be applied on various applications and from it the clinical datasets; ^[16,18] to improve the quality and safety of healthcare delivery. we need to deploy Electronic Health Records 'EHR' with a corresponding need to apply data mining technologies to extract quality data and inference rules from the information stored in those electronic records to provide real-time decision supports to clinicians and healthcare providers. Most clinical datasets are not structured and applying BI technologies ensure quality data extractions and analysis. Clinical datasets provide an excellent environment in which combined analyses of both structured and unstructured datasets can prove fruitful and useful.

Applying BI technologies on clinical datasets has many benefits in improving the quality and safety of healthcare delivery. Where BI enables organizations to deal with huge amount of data and extract useful knowledge to support decision making. ^[1,3]

So, the Electronic health record 'EHR'; According to the Health Information Management

Systems Society's (HIMSS) definition of EHRs; "The Electronic Health Record is a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports. The EHR automates and streamlines the clinician's workflow. The EHR has the ability to generate a complete record of a clinical patient encounter, as well as supporting other care-related activities directly or indirectly via interface including evidence-based decision support, quality management, and outcomes reporting." (Electronic Health Record, n.d.) The electronic health record has been discussed since the 1970's and recommended by the Institute of Medicine in 1991. and it's represented as an electronic repository of all patient medical information includes patient's health status and health care.

The EHR Could be considered the center piece of Medical Informatics with its potential to improve patient safety, productivity and data retrieval. Although the current adoption rate is low, EHR will likely become the focal point of all patient encounters. Multiple resources that are currently standalone programs will be incorporated into the EHR. It is anticipated that EHR use will eventually be shown to improve patient outcomes like morbidity and mortality as a result of decision support tools that decrease medication errors and standardize care with embedded clinical guidelines with improved quality clinicians will be paid more through pay for performance programs. It is also important to realize that one of the outcomes of EHRs will be the production of voluminous healthcare data.

As pointed out by Steve Balmer, the CEO of Microsoft, there will be an "Explosion of

Data" as a result of automating and digitizing multiple medical processes. Adding new technologies such as electronic prescribing and regional health information organizations will produce data that heretofore has not been available. ^[11]

The system stored in a way that can be accessible from any location by any provider care about patient's health, so it referred to a computer-based patient record system. ^[14]

The components of EHR can be: patient demographics, medical history, examination and progress reports of health and illnesses, medicine and allergy lists, scheduling, retrieval, and archiving of laboratory tests, graphic image display, medication ordering including patient safety functions, clinical practice guidelines, appointment scheduling and claims and payment processing.

The Electronic Medical Record 'EMR' systems help to improve the quality level of patient care and decrease medical errors. It also may help in: integrated decision support and guidance, security, privacy, and confidentiality standards for clinical terminology, integrating the patient record with other information resources in the organization and financial benefits' ^[7]

Business Intelligence has emerged as a technology that has the potential to operationalize the repository content of EHR in supporting evidence-based practice and improving the quality of healthcare delivery, incorporating the BI technology into EHR is to improve the quality and safety of healthcare delivery. ^[4] With an increasing need to deploy Electronic Health Record and Electronic Medical Record, there is also a corresponding need to apply data mining technologies to extract quality data and inference rules from the information stored in those electronic records so as to provide real-time decision supports and evidence-based practice to clinicians and healthcare providers.

II. RESEARCH MAIN QUESTIONS

This research will be guided by a main research question, which is Formulated as follows:

How could we utilize Business Intelligence via Data Mining Techniques to detect and predict disease?

If we assume that we're focusing on diabetes disease and its complications specially the kidney failure disease and how the dialysis costs a lot of money as it treated on the states charge, so in order to answer the main question, a number of sub questions arise such as these questions are:

1.How could be predict and detect a kind of disease through Patient EHR?

2. What are the appropriate Data Mining techniques that could be used to enhance the Business Intelligence for predicting and detecting the kind of the disease?

3. What's the suitable framework that could be used to visualize the finding indicators through the BI?

The answer of this research sub-questions will address the main research objectives and also provide an answer for the general research question. In the first sub-question, we identify how we could utilize the patient electronic health record which is based on the patient medical history data in an efficient way in order to get hidden information that could help in detecting and predicting any disease. The second research sub-question addresses that there are a lot of data mining techniques which could be used for the detection and prediction, and we should select the appropriate one which will return an accurate result at the right time, so it could help to enhance the business intelligence applications. The third sub-question focuses on how the framework could help the detection and the prediction results could be visualized in an effective way through the Business Intelligence applications.

III. CASE STUDY

The motivation of addressing the research problem comes from that Diabetes is one of the major health problems in Egypt where According to DR. Hesham elHefnawy "The Dean of the National Institute of Diabetes" people who are suffering from the diabetes in Egypt are more than 7 million patients and it is expected to increase exponentially in the upcoming years so we need to make an early detection for this disease. Diabetes not only affects the patients, but it also leads to losses in the state budgets because of the costs resulting from diagnosing and treatment of the disease and this cost increases in the case of disease progression and leads to a lot of complications such as kidney failure where the dialysis itself costs a lot from the overall state's budget.

And, According to World Health Organization there are more than one million people in this world who are suffering from diabetes. Diabetes is the endocrine disease across all population and age groups and it is the seventh leading cause of death according to U.S. death certificates and, it is the major cause for heart stroke, kidney failure, non-traumatic lower-limb amputations and blindness.

One of the main goals of this research is to help in an early detection and prediction of the disease which will return benefits from many aspects. Such as will save many of people's lives. Besides doctor's experience and knowledge using business intelligence and data mining will allow a rapid detection and prediction process, save time and enhance hospital reputation.

The Diabetes is one of the fatal, metabolic and costly disease that increases blood sugar level. It is not only a disease but also responsible for occurring different kinds of diseases like heart attack, blindness, kidney diseases etc. If diabetes

goes out of control, then it increases blood glucose level more than 200mg/dL which leads to micro and macro vascular disease complications.^[13,17] The phenomenon of diabetes increased in recent years for many reasons, especially among the segments of small ages and it portends a serious danger to the health of the youth. Because of the wrong food habits and lack of exercise. And According to the International Diabetes Federation the estimated number of people with diabetes has jumped from 30 million in 1985 to 150 million in 2000 and then to 246 million in 2007, It expects this number to hit 380 million by 2025.

The aim of the research focused on the healthcare Environment where the operations grow bigger and become more complex and using huge amount of data so it's difficult to find hidden information and major challenges are how to;

- Extract useful information from huge amount of data such as from the electronic health record-diabetes patient record.
- Predict the possible diseases that could infect any patient and how to overcome this rapidly.
- Enhance patient's health care system.
- Detect the complications of a certain disease.
- Reduce the State's budget for the dialysis cases where an important reason for the kidney's failure is the diabetes.

Using Business Intelligence and Data mining tools provide framework to allow;

- Early detection of diabetes mellitus and if it affects the kidney or not
- Rapid and accurate diagnosing system
- Identify the corresponding treatments that will not affect the patient by any other disease.

IV. THE PROPOSED FRAMEWORK

The research motivation and its objective will be defined through the proposed framework where it shows all the sequence of operations that will be done, starting from entering the data which assumed to be from the Electronic health record into the steps of (Pre-Processing- Loading- Processing), after this steps the data finally will be visualized in a form that could help to make the right decision at the right time and help to achieve our objectives which briefly constituted; - Detecting & predicting diseases.

How the integration of DM techniques could help in making accurate decisions in a timely manner.

How the implementation of the previous two objectives could help in reducing the costs of treatments from the overall budget and save people's lives.

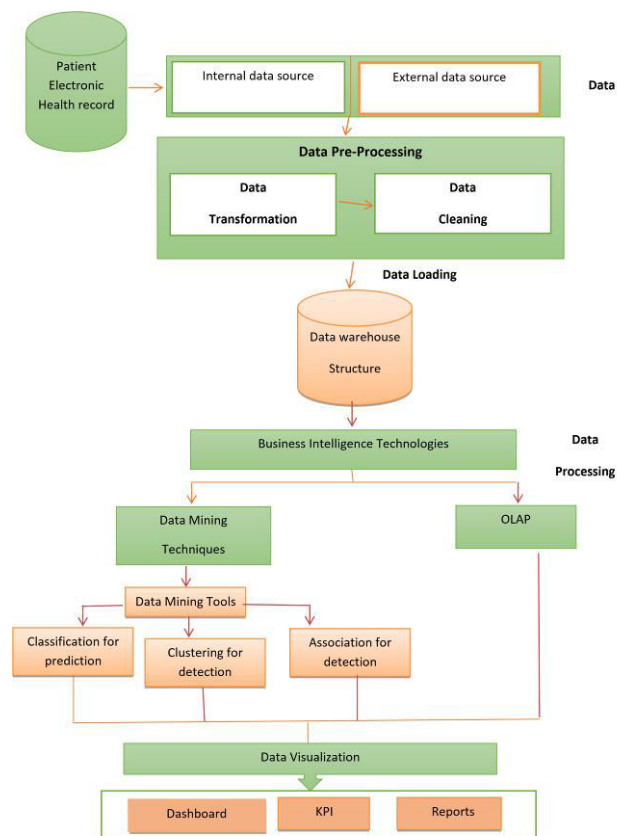


Fig 3.: The Proposed Framework

V. THE FRAMEWORK MECHANISM

The framework is divided into three phases which are input, processing and output phases. The description of the framework are as follows;

Phase1: Input

In this research, the corresponding users are the patients where the data sources mainly extracted from Patient Electronic Health Record where it come basically from either internal source such as the patient historical medical records or external source such as: laboratory tests result, radiology data etc.

Then the data enter another level which the preprocessing is where data is transformed and cleaned in a way that enable it to be correctly loaded into the data warehouse repositories.

Phase2: Processing

After the data loaded into repositories then it can be retrieved for processing and this process can be done either through OLAP “where the data have viewed in a summarized way for reporting and analysis” or through the Data Mining techniques and in our research we will focus on this technique where “hidden information discovered from the dataset” by using data mining tools we can make prediction through the DM classification

model and detection through DM clustering and association model the algorithms used in each model will be discussed later after analysis.

Phase3: Output

The output from the framework can be viewed as the way the result can be visualized to the decision makers who are the practitioners in a form of Dashboards, KPI or Reports and allow them to make the right decision at the right time.

VI. CONCLUSION

Because of the increasing rate of diseases in Egypt and this constitutes a major threat to youth and the upcoming generations.

So the research tackles the necessity of having a framework that could detect and predict any disease as well as visualize it through the business intelligence application, In other words the research attempt to employ Data mining techniques to early detect diabetes or any other disease and predict the disease’s complications and finally display this results in a form which assumed that could help practitioners to make the right decisions at the right time as well as it will save many of people’s life and a lot of state’s budget.

This research shows the utilization of Patient EHR through the Business intelligence application could provide the needed results to make the right decision at the right time.

Identify how the integration of the three rules data mining techniques could obtain the desired detection and prediction results in an efficient time manner & help practitioner to make the right decision at the right time which will save patient’s lives.

This research shows how the early detection and prediction framework can reduce morbidity & mortality, therefore the overall costs of treatments and budget of the healthcare sector will be reduced.

REFERENCES

- [1] Abd Elaziz, L., Sharaf, A., Nasr, M. (2016), Personal Integrated Electronic Health Record, International Journal of Computer Applications (IJCA), Volume 150 – No.12, September 2016, pp 44-47.
- [2] Ashrafi, N. et al (2014), The impact of Business Intelligence on Healthcare Delivery in the USA, Interdisciplinary Journal of Information Knowledge and Management, 117-130.
- [3] Anand, R et al (2013), A Data Mining Framework for Building Health Care Management System, International Journal of Engineering Research & Technology, 1639-1648.

- [4] Bonney W. (2013), Applicability of Business Intelligence in Electronic Health Record, The 2nd International Conference on Integrated Information, 257-262.
- [5] Bhuvaneswari K. (2015), A Comparative Analysis of Clustering Techniques using Genetic Algorithm, International Journal of Computer Science and Mobile Computing, 80-86.
- [6] Fayyad, U. et al (1996), From Data Mining to Knowledge discovery in Database, American Association of Artificial Intelligence Magazine, 37-54.
- [7] Garets, D., Davis, M. (2006), Electronic Medical Record vs. Electronic Health Record: Yes, There Is a Difference, A HIMSS Analytics White Paper, 1-14.
- [8] Helmi, A., Nasr, M., Farhan, M., (2015) The Pivotal Role of Geospatial Information Systems based on Hybrid Cloud Computing for the Health Sector in Egypt, International Journal of Emerging Trends & Technology in Computer Science (IJETCS), Volume 4, Issue 5(2), Page 99-103, September-October- 2015.
- [9] Helmi, A., Farhan, M., Nasr, M., (2018) A Framework for Integrating Geospatial Information Systems and Hybrid Cloud Computing, Computers & Electrical Engineering Journal, Elsevier, Volume 67, April 2018, pp. 145-158.
- [10] Helmi, A., Nasr, M., Mohamed, R. (2017), Cloud based Spatial Analysis for the Health Sector: A Case Study of Egypt, International Journal of Advanced Networking and Applications (IJANA), Volume 09, Issue 02, Sep - Oct 2017 issue, pp. 3387-3390.
- [11] Hoyt, R. (2009), Medical Informatics (Practical Guide for Healthcare Professional), Florida: Lulu.com, Electronic Health Record. (n.d.), Retrieved from HIMSS: <http://www.himss.org/library/ehr/?navItemNumber=13261>
- [12] Jensen, P., et al (2012), Mining Electronic Health Record: Towards better research applications and clinical care, Nature Reviews, 395-405.
- [13] Kavitha, S. (2012), Monitoring of Diabetes with Data Mining via CART Method, International Journal of Emerging Technology and Advanced Engineering, 157-162.
- [14] Khedr, A et al (2016), A proposed Electronic Health Record Content Structure based on Clinical organizational survey, 1-15.
- [15] Lyer, A., et al (2015), Diagnosis of Diabetes using Classification Mining Technique, International journal of Data Mining & Knowledge Management Process, 1-14.
- [16] Milovic, B. et al (2012), Prediction and Decision Making in Health Care using Data Mining, International Journal of Public Health Science, 69-78.
- [17] Rajesh, S. (2012), Application of Data Mining Methods and Techniques for Diabetes Diagnosis, International Journal of Engineering and Innovative Technology, 224-229.
- [18] Rajkumar, R. (2010), Diagnosis of heart disease using data mining algorithm, Global Journal of Computer Science & Technology, 38-43.
- [19] Stackowiak, R. a. (2007), Oracle Data warehouse and Business Intelligence Solutions, Wiley Publishing.
- [20] Soni, Y. et al (2011), Predictive Data Mining for Medical Diagnosis: An overview of heart disease Prediction, International Journal of Computer Application, 43-48.
- [21] Srinivas, R. (2010), Application of Data Mining techniques in healthcare & Prediction of heart attacks, International Journal on computer science and engineering, 250-255.
- [22] Sudhakar, K. et al (2014), Study of Heart Disease Prediction Using Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering, 1157-1160.
- [23] Sultan, T., Nasr, M., Khedr, A., Abdou, R., (2013) A Proposed Integrated Approach for BI and GIS in Health Sector to support Decision Makers (BIGIS-DSS), In the International Journal of Advanced Computer Science and Applications (IJACSA), January 2013, Volume 4, No.1, pp. 170-176.
- [24] Tvrdikova, M. (2007), Support of Decision Making by Business Intelligence Tools, 6th International Conference (p. 368), Computer information system and industrial management application.
- [25] Wang, Y. (2010), Business Intelligence and Data Mining in MBS Carbon management.