IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 10, December 2014.
www.ijiset.com
ISSN 2348 – 7968

Data Mining On Diabetics

Janani Sankari.M[1],Saravana priya.M[2]

Assistant Professor[1,2]
Department of Information Technology[1],Computer Engineering[2]
Jeppiaar Engineering College,Chennai [1] , D.Y.Patil College of Engineering,Pune[2]

ABSTRACT

Data mining is a relatively new field of research whose major objective is to acquire knowledge from large amounts of data. In medical and health care areas, due to regulations and due to the availability of computers, a large amount of data is becoming available. On the one hand, practitioners are expected to use all this data in their work but, at the same time, such a large amount of data cannot be processed by humans in a short time to make diagnosis, prognosis and treatment schedules. A major objective of this thesis is to evaluate data mining tools in medical and health care applications to develop a tool that can help make timely and accurate decisions. Two medical databases are considered, one for describing the various tools and the other as the case study. The first database is related to breast cancer and the second is related to the minimum data set for mental health (MDS-MH). The breast cancer database consists of 10 attributes and the MDS-MH dataset consists of 455 attributes.This research concentrates upon predictive analysis of diabetic treatment using a regression based data mining technique. The Oracle Data Miner (ODM) was employed as a software mining tool for predicting modes of treating diabetes. The support vector machine algorithm was used for experimental analysis. Datasets of Non Communicable Diseases (NCD) risk factors in Saudi Arabia were obtained from the World Health Organization (WHO) and used for analysis. The dataset was studied and analyzed to identify effectiveness of different treatment types for different age groups. The five age groups are consolidated into two age groups, denoted as p(y) and p(o) for the young and old age groups, respectively. Preferential orders of treatment were investigated. We conclude that drug treatment for patients in the young age group can be delayed to avoid side effects. In contrast, patients in the old age group should be prescribed drug treatment immediately, along with other treatments, because there are no other alternatives available.

Keywords: The Oracle Data Miner, Datasets of Non Communicable Diseases

1.INTRODUCTION

There was a time when data were not readily available. As data became more abundant, however, limitations in computational capabilities prevented the practical application of mathematical models. At present, not only are data available for analysis but computational resources are capable of supporting a variety of sophisticated methods. Consequently, data mining tools are now being used for clinical data. The bottleneck in data analysis is now raising the most appropriate clinical questions and using proper data and analysis techniques to obtain clinically relevant answers. In this paper, we have generated reports through the use of data mining tools on a pre-compiled dataset for non-communicable diseases in Saudi Arabia. Data mining is the process of selecting, exploring and modelling large amounts of data. This process has become an increasingly pervasive activity in all areas of medical science re-search. Data mining has resulted in the discovery of useful hid-den patterns from massive databases.

ISSN 2348 – 7968

The following six types of treatments were identified in the

2005 World Health Organization's NCD report of Ministry of

Health, Saudi Arabia and are discussed below:

(a) Drug

(b) Diet

(c) Weight reduction

(d) Smoke cessation

(e) Exercise

(f) Insulin

A. ***Drug***: Oral medications, in the form of tablets help to control blood sugar levels in patients whose bodies still produce some insulin. Drugs are usually prescribed to patients with diabetes (type 2) along with recommendations for making specific dietary changes and getting regular exercise. Several drugs are often used in combination to achieve optimal blood sugar control.

Table 1: Drug.

Sr_No	Age	N	Small_n	Percentage	SE
1	15–24	11	3	28.6	11.7
2	25–34	13	9	52.4	13.2
3	35–44	70	30	17.2	4.2
4	45–54	130	96	73.7	4.7

B***. Diet***: Patients with diabetes should maintain consistency in both food intake timings and the types of food they choose. Dietary consistency helps patients to prevent blood sugar levels from extreme highs and lows. Meal planning includes choosing nutritious foods and eating the right amount of food at the right time. Patients should consult regularly with their doctors and registered dieticians to learn how much fat, protein, and carbohydrates are needed.

Table 2: Diet

Sr_No	Age	N	Small_n	Percentage	SE
1	15–24	11	3	28.6	11.7
2	25–34	13	9	52.4	13.2
3	35–44	70	40	65.2	7.2
4	45–54	130	88	73.7	4.7

ISSN 2348 – 7968

C. Weight reduction: One of the most important remedies for diabetes is weight reduction. Weight reduction increases the body's sensitivity to insulin and helps to control blood su-gar levels.

Table 3: Weight reduction

Sr_No	Age	N	Small_n	Percentage	SE
1	15–24	11	3	28.6	11.7
2	25–34	13	9	52.4	13.2
3	35–44	70	46	65.2	7.2
4	45–54	130	96	73.7	4.7

D. ***Smoke cessation***: Smoking is one of the causes for uncontrolled diabetes . Smoking doubles the damage that diabetes causes to the body by hardening the arteries. Smoking augments the risk of diabetes.

E. ***Exercise***: Exercise is immensely important for managing diabetes. Combining diet, exercise, and drugs (when pre-scribed) will help to control weight and blood sugar levels. Exercise helps control diabetes by improving the body's use of insulin. Exercise also helps to burning excess body fat and control weight.

Table 4: Exercise

Sr_No	Age	N	Small_n	Percentage	SE
1	15–24	11	3	28.6	11.7
2	25–34	13	9	52.4	13.2
3	35–44	70	40	65.2	7.2
4	45–54	130	88	73.7	4.7

F. ***Insulin***: Many people with diabetes must take insulin to manage their disease. Diabetes is a particularly opportune disease for data mining for a number of reasons . First, many diabetic databases with historic patient information are available. Second, new knowledge about treatment patterns of diabetic scan help save money. Diabetes can also produce terrible afflictions, such as blindness, kidney failure, and heart failure. Finally, physicians need to know how to quickly identify and diagnose potential cases.

Table 5: Insulin

Sr_No	Age	N	Small_n	Percentage	SE
1	15–24	11	3	28.6	11.7
2	25–34	13	7	52.4	13.2
3	35–44	70	40	65.2	7.2

ISSN 2348 – 7968

2.DATA MINING PROCESSING BLOCKS:

A. Data selection: The first stage of the mining process is data selection from the WHO's NCD report of Saudi Arabia. In this step, the data are prepared and errors such as missing values, data inconsistencies, and wrong information are corrected.

B. Data preparation: The data preparation stage is crucial for data analysis. The Oracle Data Miner software re-quires input to be provided in a particular format. Consequently, it was deemed necessary to convert the database to Oracle Database 10g format to facilitate use with the Oracle Data Miner.

C. Data analysis: In the data analysis stage, data are analyzed to achieve the desired research objectives, for example by selecting the appropriate target values from the master table. In a data mining engine, the data mining techniques comprise a suite of algorithms such as SVM, Naıve Bayesian, etc. In this study, we used a regression technique that employed a support vector machine algorithm.

D. Result database: At this stage, the desired algorithm and associated parameters have been chosen. The Oracle Data Miner software has a specific option, such 'publish', that pro-cesses the raw data and creates a result database.

E. Knowledge evaluation and pattern prediction: This stage extracts new knowledge or patterns from the result database. An informative knowledge database is generated that facilitates pattern forecasting on the basis of prediction, probabilities, and visualization.

F. Deployment: The final stage of this process applies a previously selected model to new data to generate predictions.

3. The KDD Process:

Knowledge discovery is the process of automatically generating information formalized in a form "understandable" to humans.

Three components are required for the KDD process, which are the following:

• A goal is the outcome we need to find from analyzing the data; Example: how many people with X Y Z symptoms died with cancer?

• A database is where all the data and information about the system is located. Usually this stage is used to know the background information. This information provided will be related with the training data or examples provided which is used for the next stage. Example, what does this attribute in the database stand for?

• A set of training examples, as described earlier, the system that is created is automated, meaning the user only have to put in the database and information about what he needs to find. First the system should be trained so that it can analyze the similarities between various attributes.

ISSN 2348 – 7968

The following are the steps involved :

STEP 1:- The first step is to predefine our mission or a goal before discovering knowledge. We also have to point out from which database we can obtain the knowledge.

STEP2:- Consider a case where we have millions of data points. We have to select a subset of the database to perform the required knowledge discovery steps. Selection is the process of selecting the right data from the database on which the tools in data mining can be used to extract information, knowledge and pattern from the provided raw data.

STEP3:- Data pre processing and data cleaning. In this step we try to eliminate noise that is present in the data. Noise can be defined as some form of error within the data. Some of the tools used here can be used for filling missing values and elimination of duplicates in the database.

STEP 4:- Transformation of data in this step can be defined as decreasing the dimensionality of the data that is sent for data mining. Usually there are cases where there are a high number of attributes in the database for a particular case. With the reduction of dimensionality we increase the efficiency of the data-mining step with respect to the accuracy and time utilization.

STEP 5:- The data mining step is the major step in data KDD. This is when the cleaned and pre processed data is sent into the intelligent algorithms for classification, clustering, similarity search within the data, and so on. Here we chose the algorithms that are suitable for discovering patterns in the data.

STEP 6:- Interpretation. In this step the mined data is presented to the end user in a human-viewable format. This involves data visualization, which the user interprets and understands the discovered knowledge obtained by the algorithms.

4.DATA MINING:

Before data mining is one among the most important steps in the knowledge discovery process. It can be considered the heart of the KDD process. This is the area, which deals with the application of intelligent algorithms to get useful patterns from the data.

Some of the different methods of learning used in data mining and as follows :

• ***Classification learning:-*** The learning algorithms take a set of classified examples (training set) and use it for training the algorithms. With the trained algorithms, classification of the test data takes place based on the patterns and rules extracted from the training set. Classification can also be termed as predicting a distinct class.

• ***Numeric predication:-*** This is a variant of classification learning with the exception that instead of predicting the discrete class the outcome is a numeric value.

ISSN 2348 – 7968

• ***Association learning:-*** The association and patterns between the various attributes are extracted are from these rules are created. The rules and patterns are used predicting the categories or classification of the test data.

• ***Clustering:*** - The grouping of similar instances in to clusters takes place. The challenges or drawbacks considering this type of machine learning is that we have to first identify clusters and assign new instances to these clusters.

5. CONCLUSION:

The prevalence of diabetes is increasing among Saudi Arabian patients. The present study concludes that elderly diabetes patients should be given an assessment and a treatment plan that is suited to their needs and lifestyles. Public health awareness of simple measures such as low sugar diet, exercise, and avoiding obesity should be promoted by health care providers. In this study, predictions on the effectiveness of different treatment methods for young and old age groups were elucidated.

References:

[1] A. Vander, J. Sherman, D. Luciano, Human Physiology, McGraw-Hill,New York, 2001.
[2] S. Herrera, With the race to chart the human genome over, now the real work begins, Red HerringMagazine,April1,2001,Availableat http://www.redherring.com/mag/issue95/1380018938.html, Accessed on July 30, 2003.
[3] SNP Consortium, Single Nucleotide Polymorphisms for Biomedical Research, The SNP Consortium Ltd., Available athttp://snp.cshl.org/, Accessed on July 30, 2003
[4]Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on Diabetes and pre-diabetes in http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf