Data Mining Techniques: Contemporary Amalgam System to Predict Diabetes.

Vimalavinnarasi.A
Assistant professor, Idhayacollege of arts and science, Puducherry

Abstract: *Diabetes is a never ending disease which affects many major organs of the human body, including heart, blood vessels, nerves, eyes and kidneys. The World Health Organization (WHO) estimates that nearly 200 million people all over the world suffer from diabetes and this number is likely to be doubled by 2030. In India, there are nearly 50 million diabetics, according to the statistics of the International Diabetes Federation. To identify the diabetes mellitus the medical practitioner will diagnose the pattern consists of observable symptoms and based on the all respective test. The risk and costs may be differing according to the patient condition. In this paper, we are providing a novel approach is to act like a medical practitioner who will identify the type of diabetes and this will provide some suggestions to regulate blood sugar level and also notify the risk factor of the patient like heart attack, nerves problem will it affect the eye or the kidney. There are two different and modern approaches are used to develop this automated model. C5.0 algorithm is going to be used to classify the patient data and fuzzy inference for analyzing data. We also achieve accurate results through analyzing various data.*

I. Introduction

Diabetes is a silent disease. The biggest challenge in this disease is diagnosing the diabetes. Diabetes is a condition that develops when the body can no longer produce enough insulin to moderate blood sugar levels.Insulin produced by the pancreas lowers blood glucose. Absence or insufficient production of insulin causes diabetes.The main objectives of the research are:
1. Develop a model which is used to predict diabetes.
2. To distinguish the factors which are not consider as decision factors of diabetes
3. To identify the special effects of diabetes
4. To provide the extracted knowledge from the data
5. To recognize the undiscovered patterns which leads to diabetes

Diabetes accounts for thousand new cases of blindness in adult age 20 – 74. It also increases the risk of other serious vision conditions, such as cataracts and glaucoma. High levels of sugar in the blood start to affect the blood vessels and other areas of the body. The two types of diabetes are referred to as type 1 and type 2 and gestational. Former names for these conditions were insulin-dependent and non-insulin-dependent diabetes, or juvenile onset and adult onset diabetes.

Type 1 Diabetes

Type 1 diabetes is a chronic illness characterized by the body's inability to produce insulin due to the autoimmune destruction of the beta cells in the pancreas. Various factors may contribute to type 1 diabetes, including genetics and exposure to certain viruses. Although type 1 diabetes usually appears during childhood or adolescence, it also can begin in adults.Patients with type 1 diabetes require lifelong insulin therapy. Most require 2 or more injections of insulin daily, with doses adjusted on the basis of self-monitoring of blood glucose levels. Insulin replacement is accomplished by giving basal insulin and a pre-prandial (premeal) insulin.

Type 2 Diabetes

It affects many major organs, including the heart, blood vessels, nerves, eyes and kidneys. All these kinds of complications can be controlled easily by maintaining the sugar level. Although long-term complications of diabetes develop gradually, they can eventually be disabling or even life-threatening. Unlike people with type 1 diabetes, the bodies of people with type 2 diabetes make insulin. But either their pancreas does not make enough insulin or the body cannot use the insulin well enough. This is called insulin resistance. When there isn't enough insulin or the insulin is not used as it should be, glucose (sugar) can't get into the body's cells. When glucose builds up in the blood instead of going into cells, the body's cells are not able to function properly. Some of the potential complications of diabetes include:
1. **Heart and blood vessel disease.** Diabetes dramatically increases the risk of various cardiovascular problems, including coronary artery disease with chest pain (angina), heart attack, stroke, narrowing of arteries (atherosclerosis) and high blood pressure.

2. **Nerve damage (neuropathy).** Excess sugar can injure the walls of the tiny blood vessels (capillaries) that nourish your nerves, especially in the legs.
3. **Kidney damage (nephropathy).** The kidneys contain millions of tiny blood vessel clusters that filter waste from your blood. Diabetes can damage this delicate filtering system. Severe damage can lead to kidney failure or irreversible end-stage kidney disease, requiring dialysis or a kidney transplant.
4. **Eye damage.** Diabetes can damage the blood vessels of the retina called diabetic retinopathy, potentially leading to blindness. Diabetes also increases the risk of other serious vision conditions, such as cataracts and glaucoma.Other problems associated with the buildup of glucose in the blood include:
1. **Dehydration.** The buildup of sugar in the blood can cause an increase in urination, causing dehydration.
2. **Diabetic coma (hyperosmolar nonketotic diabetic coma).** When a person with type 2 diabetes becomes very ill or severely dehydrated and is not able to drink enough fluids to make up for the fluid losses, they may develop this life-threatening complication.

Gestational Diabetes

Gestational diabetes develops during pregnancy (gestation). Like other types of diabetes, gestational diabetes affects how your cells use sugar (glucose). Gestational diabetes causes high blood sugar that can affect your pregnancy and your baby's health.Any pregnancy complication is concerning, but there's good news. Expectant moms can help control gestational diabetes by eating healthy foods, exercising and, if necessary, taking medication. Controlling blood sugar can prevent a difficult birth and keep you and your baby healthy.In gestational diabetes, blood sugar usually returns to normal soon after delivery. But if they had gestational diabetes, they are at risk for type 2 diabetes. They will continue working with their health care team to monitor and manage your blood sugar.

II. Classification Algorithmc5.0

C5.0 algorithm is widely used as a decision tree method in machine learning. Initially we have ID3.0 algorithm. Based on ID3.0, people developed C4.5 algorithm, and finally develop C5.0 algorithm. This type of decision tree model is based on entropy and information gain. The decision tree based on entropy and information gain (ID3, C4.5, C5.0) is to build a tree by partitioning with highest information gain. For categorical variables, the partition can be done in each level and find the best variable with highest information gain. For continuous variables, the best cut point should be find to do a binary partition with highest information gain. For missing data, a separate node should be created, which is not causing changes of information gain.

Some Important Features:
- C5.0 has been designed to analyze substantial databases containing thousands to millions of records and tens to hundreds of numeric, time, date, or nominal fields. C5.0 also takes advantage of computers with up to eight cores in one or more PUs (including Intel Hyper-Threading) to speed up the analysis.
- To maximize interpretability, C5.0 classifiers are expressed as decision trees or sets of if-then rules, forms that are generally easier to understand than neural networks.
- C5.0 is available for Windows Xp/Vista/7/8 and Linux.
- C5.0 is easy to use and does not presume any special knowledge of Statistics or Machine Learning (although these don't hurt, either!)
- Rule Quest provides C source code so that classifiers constructed by /C5.0 can be embedded in your organization's own systems.
    Fuzzy Expert System

A fuzzy expert system is simply an expert system used in the field of medicine. By applying fuzzy determination mechanism the diagnosis of diabetes become simple for medical practitioners. A fuzzy expert system is a collection of fuzzy rules and membership functions thatare used to reason aboutdata. Using fuzzy expert system expert knowledge can berepresented that use vague and ambiguous terms in computer.

Applications of fuzzy set theory was reorganized in the field of medicine and, the uncertainty was found in the process of diagnosis of the disease that has more frequently focus on the applications of fuzzy set theory. The desire to better understand and teach this difficult and important process of medical diagnosis has prompted attempt to model it with the use of fuzzy set. These models will vary in the degree to which they attempt to deal with different complications of medical diagnosis such as the relative important symptoms.

III. Literature Review

Mining of Soil Data Using Unsupervised Learning Techniques: Shweta Taneja, Rashmi Arora1, Savneet Kaur2
Abstract: Data mining has emerged as one of the greatestResearch domain from the last few decades in order to extract useful information or knowledge from huge amounts of data. Discovering hidden information from the

historical data is one of the important tasks of Data Mining. Prior to data mining process, the extraction of knowledge was computed manually using various statistical techniques. But such techniques were time consuming. In our paper, we have used data mining technique in the field of soils. We have selected a subset of soils from World Soil Science Database. Our research focuses to find out the useful relationships among different types of soils. We have implemented clustering technique using WEKA tool to create clusters of the soil based on their salinity.

Efficient Implementation Of Data Mining: Improve Customer's Behavior Abdullah Al- Mudimigh1, FarrukhSaleem1, Zahid Ullah1

Abstract: Evaluating the performance of any organization isan essential part for overcoming their weaknesses. Customer isalways on prior for finding and assessing the company'sperformance. They are always respectable for everyorganization. In this paper we first examine the CustomerRelationship Management (CRM), especially customerbehavior and customer profiling. Then we describe the generaloverview of most common data mining techniques. The mainpurpose of this paper is how data mining techniques can extractrespectable knowledge from the large customer's database andhow to analyze customer behavior to improve businessperformance. Therefore, we proposed a model for CRM with theefficient implementation of data mining, for improving customerbehavior. For this, we evaluate and analyze the customerunderstanding by using rule induction process on clustered datafrom customer's database with reference to the customer query.

Data Clustering: A ReviewA.K. JAINM.N. MURTYP.J. FLYNN

Clustering is the unsupervised classification of patterns (observations, data items,or feature vectors) into groups (clusters). The clustering problem has beenaddressed in many contexts and by researchers in many disciplines; this reflects itsbroad appeal and usefulness as one of the steps in exploratory data analysis.However, clustering is a difficult problem combinatorial, and differences inassumptions and contexts in different communities has made the transfer of usefulgeneric concepts and methodologies slow to occur. This paper presents an overviewof pattern clustering methods from a statistical pattern recognition perspective,with a goal of providing useful advice and references to fundamental conceptsaccessible to the broad community of clustering practitioners. We present ataxonomy of clustering techniques, and identify cross-cutting themes and recentadvances. We also describe some important applications of clustering algorithmssuch as image segmentation, object recognition, and information retrieval.

Applications of Clustering Algorithms and SelfOrganizing Maps as Data Mining and BusinessIntelligence Tools on Real World Data Sets Lavneet Singh, ISavleen Singh and 2Parminder Kumar Dubey

Abstract: Partitioning a large set of objects intohomogeneous clusters is a fundamental operation in datamining. The k-means algorithm is best suited forimplementing this operation because of its efficiency inclustering large data sets. In this paper we present acomparative study on different clustering algorithms withrespect to k - means clustering to work on large data sets.In this paper we present a comparison among somenonhierarchical and hierarchical clustering algorithmsincluding SOM (Self-Organization Map)neural networksmethods. Data were simulated considering correlated anduncorrelated variables, non-overlapping and overlappingClusters with and without outliers. Tested withTelecommunication Users and Iris Flower data set, thecomparative algorithms had demonstrated a very goodclassification performance. Experiments on a very largetelecommunication data set set consisting of 1000 recordsand 32 categorical attributes & Iris Flower data setconsisting of 150 samples show that the SOM clusteringwith respect to k means & hierarchical clusteringalgorithm is scalable in terms of both the number ofclusters and the number of records.

Quilter Detection in Spatial Databases Using Clustering Data Mining: AmitavaKarmaker Syed M. Rahman

Abstract: Data mining refers to extracting or "mining"knowledge from large amounts of data. Thus, it playsan important role in extracting spatial patterns andfeatures. It is an essential process where intelligentmethods are applied in order to extract data patterns.In this paper, we have proposed a technique withwhich it is possible to detect whether a given data setis erroneous. Furthermore, our technique locates thepossible errors and comprehends the pattern oferrors to minimize outliers. Finally, it ensures theintegrity and correctness of large databases. We havemade use of some of the existing clusteringalgorithms (like PAM, CLARA, CLARANS) toformulate our proposed technique. The proposedoutlier detection and minimization method is simplerto implement, efficient comparing with respect toboth time and memory complexity than other existingmethods.

IV. Conclusion

From this, Contemporary amalgam system to predict Diabetes have implemented and explained.In this paper, we are providing a novel approach, act like a medical practitioner who will identify the type of diabetes and this will provide some suggestions to regulate blood sugar level and also notify the risk factor of the patient

like heart attack, nerves problem by using C5.0 algorithm and fuzzy inference. In future we also enhance our work to identify the accurate problems about the human eye and also achieve efficiency towards the system.

References

[1]   C.M Velu and K.R.Kashwan(2012) "Visual Data mining techniques for classification of diabetic patients" 2013 3rd IEEE International Advance Computing conference(IACC).

[2]   S.M. Nuwangi, C. R. Oruthotaarachchi, J.M.P.P. Tilakaratna, H. A. Caldera(2010) "Utilization of Data Mining Techniques in Knowledge Extraction for Diminution of Diabetes" 2010 Second Vaagdevi International Conference on Information Technology for Real World Problems

[3]   D. W. Patterson, (1990) "Introduction to Artificial Intelligence and Expert Systems", Prentice-Hall Inc., glewoodCliffs, USA.

[4]   B. H. Cho, H. Yu, K. W. Kim, T. H. Kim, I. Y. Kim, S. I. Kim, "Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods," *Artificial Intelligence in Medicine*, vol. 42, pp. 37-53, Jan. 2008.

[5]   I. K. Valavanis, S. G. Mougiakakou, K. A. Grimaldi, K. S. Nikita, "A multifactorial analysis of obesity as CVD risk factor: Use of neural network based methods in a nutrigenetics context," *BMC Bioinformatics*, vol. 11, pp. 453, Sep. 2010.

[6]   D. U. Campos-Delgado, M. Hernandez-Ordonez, R. Femat, and A. Gordillo-Moscoso, .Fuzzy-based controller for glucose regulation in type-1 diabetic patients by subcutaneous route,.*IEEE Trans. Biomed. Eng.*, vol.53, no.11, pp.2201.2210, Nov. 2006.

[7]   P. Magni and R. Bellazzi, .A stochastic model to assess the variability of blood glucose time series in diabetic patients self-monitoring,. *IEEE TransBiomed. Eng.*, vol. 53, no. 6, pp. 977.985, Jun. 2006.

[8]   K. Polat and S. Gunes, .An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease,.*Dig. Signal Process.*, vol. 17, no. 4, pp. 702. 710, Jul. 2007.

[9]   K. Polat, S. Gunes, and A. Arslan, .A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine,.*Expert Syst. Appl.*, vol. 34, no. 1, pp. 482.487, Jan. 2008.

[10]   X. Chang and J. H. Lilly, .Evolutionary design of a fuzzy classifier from data *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1894.1906, Aug. 2004.

[11]   L. B. Goncalves, M. M. B. R. Vellasco, M. A. C. Pacheco, and F. J. de Souza, .Inverted hierarchicalneuro-fuzzy BSP system: A novel neuro-fuzzy model for pattern classification and rule extraction in databases,. *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 36, no. 2, pp. 236.248, Mar. 2006.

[12]   H. Kahramanli and N. Allahverdi, .Design of a hybrid system for the diabetes and heart diseases,. *ExpertSyst. Appl.*, vol. 35, no. 1/2, pp. 82.89, Jul./Aug. 2008.

[13]   Chang-Shing Lee,. A Fuzzy Expert System for Diabetes Decision Support Application. IEEEtransactions on systems, man, and cybernetics.pa rt b: cybernetics, vol. 41, no. 1, Feb. 2011