Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# USING DATA MINING TO DEVELOP MODEL FOR CLASSIFYING DIABETIC PATIENT CONTROL LEVEL BASED ON HISTORICAL MEDICAL RECORDS

**TARIG MOHAMED AHMED**
Assoc. Prof., Department of MIS, Prince Sattam Bin Abdalaziz University, KSA
E-mail: Tarig_Harbi71@hotmail.com

**ABSTRACT**

Nowadays, diabetes is considered as one of the diseases which cause more deaths than any other disease in the world. To avoid the dangerous complications of the diabetes, patients should control a blood glucose level as the HbA1c (accumulative blood glucose level for 3 months) should be less than 7%. In this paper a new predicted model has been developed by using data mining techniques. The model aims to classify the diabetic patients into two classes which are: under control (HbA1c < 7%) and out of control (HbA1c > 7%). The treatments plans for 10061 diabetic patients were used to build the model. After comprehensive survey for classification techniques, three algorithms have been selected which were NaivaeBayse, Logistic and J48. By using WEKA application, the model has been implemented. Based on the results of experiment, Logistic algorithm has been selected as best one with high accuracy rate of 74.8%. To enhance the model accuracy, the nutrition system and exercise need to be added to the dataset as future work.

**Keywords**: *Diabetes, Data Mining, Classification techniques*

## 1. INTRODUCTION

Diabetes occurs when a blood glucose level is too high and the body couldn't to reduce this level. The diabetic patient doesn't produce any insulin or enough insulin to reduce the high glucose level. Diabetes grows when glucose can't enter the body's cells to be used as energy. Diabetes is an important cause of continued ill health. This disease causes death per year more than HIV-AIDS as one diabetic patient die every 10 seconds [1]. Around 347 million people around the world are suffering from diabetes. In 2010, around 3.4 million people died due to complications of the diabetes and about 80% of diabetes deaths occur in poor countries.

There are two types of diabetes disease which are: [1] Type1 Diabetes is known as insulin dependent. It is categorized by lacking insulin in body and requires insulin. The cause of type 1 is unknown and it is not preventable with current knowledge. Symptoms contain polyuria, polydipsia (thirst), hunger, loss of weight loss, eye problem and weakness [2]. Type2 Diabetes which is formerly called non-insulin-dependent or adult-onset. It is as a result of ineffective utilization of the insulin by the cells of the body. 90% of the people with diabetes are categorized type2 patients and mainly due to excess body weight and physical inactivity.

In Saudi Arabia, the number of diabetic patients is rapidly increasing according to official reports (http://www.who.int/nmh/countries/sau_en.pdf). In 2005, World Health Organization's NCD report of Ministry of Health, Saudi Arabia, identified six types of treatments for diabetes disease: Drug, Diet, Weight reduction, Smoke stop, Exercise and Insulin. With passage of time, the diabetes can make dangerous complications. Diabetes may lead to heart disease and stroke. About 50% of people with diabetes die of cardiovascular disease. One of the commonly seen complications is foot ulcer and limb amputations, which are mainly due to reduced blood flow and nerve damage (neuropathy) in the feet of diabetic patients. Damage of the small blood vessels of the eye is also common; this may lead to blindness which is due to diabetic retinopathy. Kidney failure of patients, affected by diabetes, is quite common. The overall risk of death among people affected with diabetes is double the risk of their peers without diabetes [4] .

To avoid diabetes complications, patients should control a blood glucose level. The HbA1c (accumulative blood glucose level for 3 months) is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

good measure for that. It should be less than 7%. In this paper a new predicted model has been developed by using data mining techniques. The model aims to classify the diabetic patients into two classes which are: (1) under- control patients whose the HbA1c is less than 7% or out-of-control patients whose the HbA1c is more than 7%. The treatments plans for 10061 diabetic patients were used to build the model. After comprehensive survey for classification techniques, three algorithms have been selected such as NaivaeBayse, Logistic and J48 algorithms. By using WEKA application, the model has been implemented. Based on results, Logistic algorithm has been selected as best one with high accuracy rate of 74.8%. The can be used to classify new diabetic patients either under-control or out-of-control based on their treatment plans. This paper has been distributed into several sections which provide complete detail about all the research phases, for example: The Section No. 2 presents the related works and many researchers conducted on the subject. The Section No. 3 contains the proposed model components, results as well as full results of discussion. Last Section (No. 4) concludes the research and present recommendations for future works.

## 2. RELATED WORK

Many researches have been conducted in the diabetes area by using of data mining as analytic powerful tool to extract knowledge from available massive data. In this research, some of these papers have been presented such as following: ALjumah et al developed a predictive analysis model using data mining technique for treating diabetes . By using the support vector machine algorithm in the Oracle Data Miner (ODM) tool, the predictive model was built. The datasets of risk factors in Saudi Arabia were collected From World Health Organization (WHO) to generate the model. Different types of treatments were identified for two age groups ( Old and young). As result, the model detects that drug treatment for young diabetic patients can delay the complication of diseases. In addition, there is no alternative to use drug treatment immediately by old age group [5]. Patil et al. proposed a hybrid protection model for detecting type 2 diabetic patients by using several data analysis methods. The main idea behind this model was to investigate the characteristics and measures that caused the diabetes. The model used two algorithms: Simple K-means clustering to select class labels and C4.5 to construct the classifier. Pima Indians diabetes dataset from university of California was used to build the model. As result of testing the model, 92.38% accuracy was obtained.[6] Vasudevan conducted a comprehensive statistical analysis on real dataset from National Institute of Diabetes and Digestive and Kidney Diseases. The analysis was done by using logistic regression method using SPSS 7.5 which obtained significant factors. The factors and the Iterative Dichotomiser-3 algorithm were used to build the classification model. As result of the study, the model identified the diabetes disorder risks.[7] Sanakal and Jayakumari designed a model for indicating whether the patient is diagnosed with the diabetes or not. The model used 9 attributes for diagnosing. The model was built on Fuzzy C-means clustering (FCM) as algorism and 768 cases as datasets. After implementation the model, FCM showed 94.3% accuracy and positive predictive 88.57% [8]. Aljumah et al. used a classification techniques for diabetic intervention and analysis model. The model was created by using Support Vector Machine and it was implemented using Oracle data miner. The dataset of the model was collected from world Health Organization(WHO). The result of this study mentioned that Smoking is one of highest causing for elevating diabetic rate[9]. Rahman and Farhana conducted a comparative study about data mining tools for diabetes diagnosis. Classification and clustering techniques were used. Many test were conducted to measure the performance of dealing with large dataset. After using several data mining techniques across several data mining tools, the result mentioned that The best algorithm was J48graft classifier in WEKA with 81.33% accuracy and the test spent 0.135 seconds for training. In TANAGRA tool with Naïve Bayes classifier provided 100% accuracy and the test spent 0.001 seconds. In MATLAB with ANFIS had 78.79% accuracy[10]. Sigurdardottir et al conducted a comprehensive survey related to adults with type 2 diabetes intervention. As survey result, data mining displayed that for glycosylated hemoglobin (HbA1c) level ≤7.9% the diabetes education intervention attained a minor alteration in HbA1c level, or from +0.1 to −0.7%. For initial HbA1c ≥8.0%,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

a significant drop in HbA1c level of 0.8–2.5% was found. The duration, educational content and education strength did not forecast changes in HbA1c levels.[11] Rupa Bagdi et al developed a decision support system based on OLAP and data mining. The system worked as assistant to doctors in order in diagnosing diabetes patients. OLAP analyzed the history background of the concerned patients. The system was used the classification algorithms ID3 and C4.5. The system was used to classify a diabetic patient with probability of high, low or medium.[12] Xue-Hui et al. purposed three models by using logistic regression, artificial neural networks (ANNs) and decision tree. These models aimed to predicting diabetes or prediabetes using common risk factors. The dataset was collected from Guangzhou, China. A standard questionnaire was managed to find information on demographic characteristics, history of family diabetes, anthropometric measurements and lifestyle risk factors. The models was built based on 12 input variables and one output variable from the questionnaire information. As result, the logistic regression model reached 76.13% accuracy, 79.59% sensitivity and a specificity of 72.74%. The ANN model achieved 73.23% with a sensitivity of 82.18% accuracy, 64.49% specificity. The decision tree (C5.0) reached 77.87% accuracy, 80.68% sensitivity and 75.13%. specificity. The decision tree model (C5.0) selected as best model.[13]

Gregori, Dario, et al. proposed several models that applied on medical records and demographic data for predicting outcomes (e.g. death). The dataset was related to type 2 diabetes about 3,892 outpatient patients from the San Giovanni Battista Hospital in Torino. Six model were developed: Logistic regression (LR), Generalized Additive Model (GAM), Projection pursuit Regression (PPR), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Artificial Neural Networks (ANN). The simpler models such as: LR, GAM and LDA performed better. GAM was associated with a very small misclassification rate. ANN is the model performed worse.[14]

Liu, Haifeng, et al. identified the factors that may help in a good taking decision to control blood glucose level for diabetic patients. The researchers conducted several experiments on real datasets by measuring HbA1C value to validate the results and compared with a clinical guideline. as result, the model was used to classify the best control for new patients.[15]. Butwall et al developed a random Forest classifier model by using different test parameters. The confusion matrix of the model mentioned that Random Forest Classifier performed better with 99.7%. accuracy.[16] .Karthikeyani et al conducted a survey for data mining classification algorithms for predicting diabetes. The survey were based on two types: supervised and unsupervised classifications. It involved C4.5, SVM, K-NN, PNN, BLR, MLR, CRT, CS-CRT, PLS-DA and PLS-LDA algorithms. The CS-CRT algorithm was selected as best one and Tanagra was selected as best tool [17]. Habibi et al developed a model using the decision tree algorithm J48 for the screening of diabetes type 2 patients. The data was collected on real for all people referred for diabetes screening between 2009 and 2011. As result of the model, 97,6% accuracy was obtained.[18]. Namayanja et al created a clustering model to investigate the measurements in blood glucose and doses of regular. The model was developed based on real diabetic patients information. The model aimed to identify such as the targeted care for diabetic patients. [19]

## 3. MATERIAL AND METHOD

### 3.1 Data Collection

This research has used real data from Health Facts database (Cerner Corporation, Kansas City, MO). The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals. 50 features were used to represent the diabetic patient medical record. The dataset consists of demographic information, treatment plan and measurements related to control of diabetes [20].

### 3.2 Data Preprocessing

To meet the main objective of this research, some data preprocesses have been done on the dataset. The proposed model was developed based on a classification attribute HPA1c because this attribute if of vital importance. So, all the records which were missing the value of this attribute have been removed from the datasets of the model. In addition, some attributes were not related to the research, so, they were removed from the dataset. Table 1 shows that all attributes were selected ( 28 attributes and 10062 instances ) to generate the proposed model after completion of the preprocessing task

Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 1: List of Features Descriptions*

| Feature name | Type | Description and values | % missing |
|---|---|---|---|
| Gender | Nominal | Values: male, female, and unknown/invalid | 0% |
| Age | Nominal | Grouped in 10-year intervals: 0, 10), 10, 20), …, 90, 100) | 0% |
| Glucose serum test result | Nominal | Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured | 0% |
| 24 features for medications ( represent treatment plan) | Nominal | For the generic names: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether the drug was prescribed | 0% |
| HPA1c test result | Nominal | Indicates the range of the result Values: “>7” if the result was greater than 7% “normal” if the result was less than 7%, | 0% |

The following tables (Table 2 to Table 5 ) present the frequency of some important attributes which had effected the model.

*Table 2: Gender Information*

| Type | Frequency | Percent(%) |
|---|---|---|
| Female | 5300 | 52.7 |
| Male | 4761 | 47.3 |
| Total | 10061 | 100.0 |

*Table 3: Age Information*

| Age Range | Frequency | Percent |
|---|---|---|
| [0-10) | 90 | .9 |
| [10-20) | 293 | 2.9 |
| [20-30) | 278 | 2.8 |
| [30-40) | 564 | 5.6 |
| [40-50) | 1351 | 13.4 |
| [50-60) | 2045 | 20.3 |
| [60-70) | 2023 | 20.1 |
| [70-80) | 2065 | 20.5 |
| [80-90) | 1189 | 11.8 |
| [90-100) | 163 | 1.6 |
| Total | 10061 | 100.0 |

*Table 4: Insulin usage*

| Insulin Usage | Frequency | Percent(%) |
|---|---|---|
| No | 4012 | 39.9 |
| Yes | 6049 | 60.1 |
| Total | 10061 | 100.0 |

Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 5: HPA1c Information*

| HPA1C Categories | Frequency | Percent |
|---|---|---|
| >7 | 7340 | 73.0 |
| Norm | 2721 | 27.0 |
| Total | 10061 | 100.0 |

### 3.3 Tools and techniques

After comprehensive survey, three data mining techniques were selected to develop the model. The final model was selected based on best evaluation criteria. The following sections present three experiments that were used to find out the best model for classifying good treatment plan for diabetic patients based on HPA1c

.

### 3.4 WEKA Application

WEKA (Waikato Environment for Knowledge Analysis) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. WEKA is free software available under the General Public License( GNU ). It is a collection of machine learning algorithms for data mining tasks.

### 3.5 Data Mining Techniques

Classification technique is the most important data mining technique used to extract knowledge from medical databases. It maps or classify into one of a several predefined classes, a classification model is used to produce classification rules in potential training set, then it can be used to classify future data items and develop better understanding of the characteristics of the data. There are many classification algorithms. In this research, three algorithms were selected after comprehensive survey such as following: [21]

The Naive Bayes algorithm is based on the conditional independence model of each predictor given the target class. The Bayesian principle is to assign a case to the class that has the largest posterior probability

Logistic algorithm is a classification model that combined both Logistic Regression and Decision Tree learning, it build a standard construction of tree leaves based on linear model on each leave [22]

Weka J48 is classification implementation for C4.5 algorithm. It was used by various researchers in the field of medical and health researches, The algorithm that was evaluated in the research Data classification is the C4.5 developed by Ross Quinlan (as well as other classifications algorithms are listed within this section), which build a decision tree based on training dataset, each node represent a test on an attribute, each branch present an outcome of that test which drive to a leaf node, represents a class or distribution, the topmost node is the root node in the tree. Trees are constructed into a Top-Down approach.[23]

### 3.6 Experiment(1) : NiaveBayse Algorithm

In this experiment, NiaveBayse algorithm was used and three sizes to train the model were selected 50%, 65% and 80%. Figure 1 presents the model that used 80% of dataset to train the classifier.

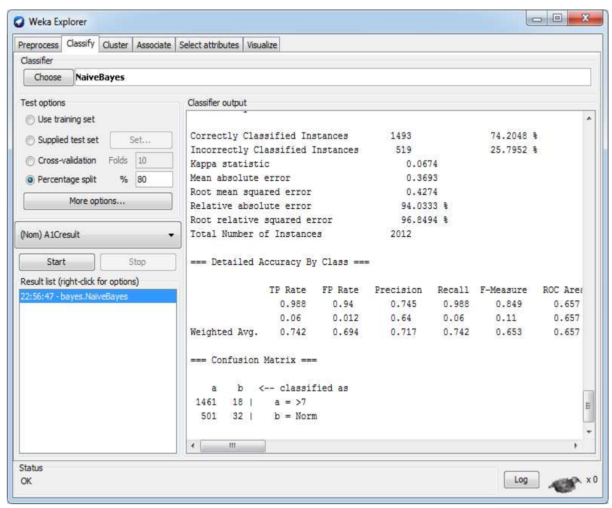


*Figure 1: Using of NaiveBayes Model*

After applying the three different sizes to train the classifier , Table 6 presents the result:

*Table 6: NaiveBayes results*

| # | Size % | Precision Avg. | Recall Avg. | F-Measure Avg. | Accuracy% |
|---|---|---|---|---|---|
| 1 | 50% | 0.683 | 0.735 | 0.646 | 73.5 |
| 2 | 65% | 0.686 | 0.736 | 0.645 | 73.6 |
| 3 | 80% | 0.717 | 0.742 | 0.653 | 74.2 |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### 3.7 Experiment(2) : Logistic Algorithm

In this experiment, Logistic algorithm was used and three sizes to train the model were selected 50%, 65% and 80%. Figure 2 presents the model that used 80% of dataset to train the classifier.

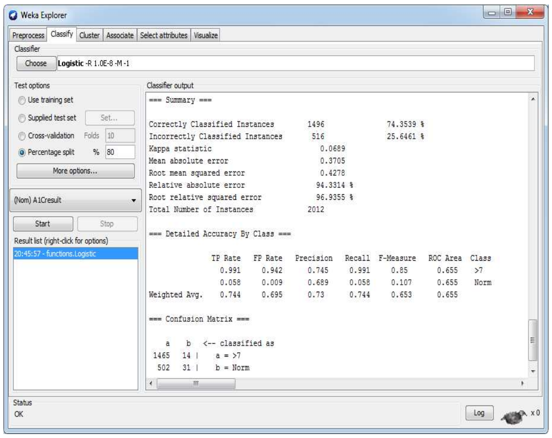


*Figure 2: Logistic Model*

After applying the three different sizes to train the classifier , Table 7 presents the result:

*Table 7: Logistic results*

| # | Size % | Precision Avg. | Recall Avg. | F-Measure Avg. | Accuracy% |
|---|---|---|---|---|---|
| 1 | 50% | 0.688 | 0.736 | 0.645 | 73.6 |
| 2 | 65% | 0.696 | 0.738 | 0.645 | 73.8 |
| 3 | 80% | 0.73 | 0.744 | 0.653 | 74.4 |

### 3.8 Experiment(3) : J48 Algorithm

In this experiment, J48 algorithm was used and three sizes to train the model were selected 50%, 65% and 80%. Figure 3 presents the model that used 80% of dataset to train the classifier.

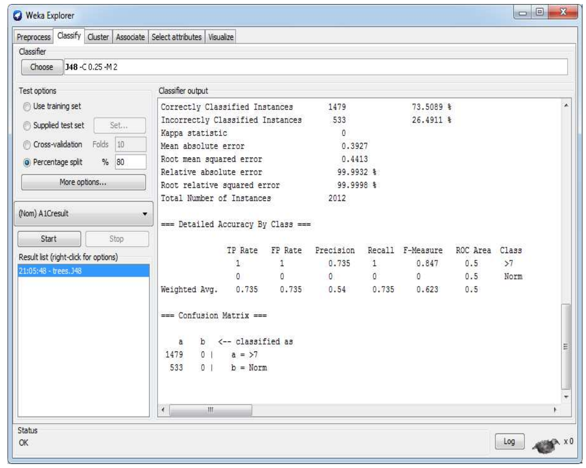


*Figure 4: J48 Model*

After applying the three different sizes to train the classifier , Table 8 presents the result:

*Table 8: the J48 results*

| # | Size % | Precision Avg. | Recall Avg. | F-Measure Avg. | Accuracy% |
|---|---|---|---|---|---|
| 1 | 50% | 0.615 | 0.734 | 0.622 | 73.4 |
| 2 | 65% | 0.539 | 0.734 | 0.622 | 73.4 |
| 3 | 80% | 0.54 | 0.735 | 0.623 | 73.5 |

### 3.9 Results Discussion

Generally, the analysis of the evaluation with the confusion matrix, which illustrate the amount of correctly and incorrectly classification classes, the correct classifications denoted by (TP) True Positive and True Negative (TN). The incorrect predicted outcomes occur when a false positive (FP) appear, that is when an outcome is predicted as positive (yes) while it is actually negative (no), furthermore, the same happened when a false negative (FN) occur if an outcome is predicted as negative but it is actually positive.[24]

Accuracy = $TP/(TP+FN)$

It is significant to use F-Measure, it produces a high result when Precision and Recall are both balanced, although Precision and Recall are valid measures, but the F-Measure was used because, Precision and Recall can be optimized at the overhead of the other.

$$Precision = TP/(FN+TP)$$

$$Recall = TP/(FP+TP)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*-Measure is given by:*
$2 * Recall * Precision / (Recall + Precision)$

It measures how good the classifier in recognizing instances of different classes is, it is a percentage of correctly classified instances in a test dataset.[25]
Based on this criteria, table 9 presents the best results which were collected by applying the three algorithms

*Table 9 three algorithms result*

| # | Algorithm name | Precision Avg. | Recall Avg. | F-Measure Avg. | Accuracy % |
|---|---|---|---|---|---|
| 1 | NiaveBayse | 0.717 | 0.742 | 0.653 | 74.2 |
| 2 | Logistic | 0.73 | 0.744 | 0.653 | 74.4 |
| 3 | J48 | 0.54 | 0.735 | 0.623 | 73.5 |

According to the above result, the proposed model was generated based on Logistic algorithm with accuracy 74.4%. To increase the model accuracy, there are additional measurements need to be added in the dataset such as the nutrition system and the exercise. Those measures have a big impact on diabetes treatment plan.

## 4. CONCLUSION AND FUTURE WORK

In this paper, a predicted model for classifying diabetic patients based on a treatment plan has been developed. The classification was performed based on HPA1c measurement. After comprehensive survey for classification techniques, three algorithms have been selected which were NaivaeBayse, Logistic and J48 algorithms. By using Logistic algorithm and 10061 medical records, the model has been build. After implementing the model, 74.6% accuracy has obtained as best result. To enhance the model's accuracy for the future work, there are additional measurements need to be added in the dataset such as the nutrition system and the exercise for the patients. Those measures have a big impact on diabetes treatment plan.

## ACKNOWLEDGEMENT

This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project # 2015/02/4140

## REFRENCES:

[1] Kaul, K., Tarr, J. M., Ahmad, S. I., Kohner, E. M., & Chibber, R. (2013). Introduction to diabetes mellitus. In *Diabetes* (pp. 1-11). Springer New York.
[2] American Diabetes Association. "Diagnosis and classification of diabetes mellitus." *Diabetes care* 31.Supplement 1 (2008): S55-S60.
[3] Stumvoll, Michael, Barry J. Goldstein, and Timon W. van Haeften. "Type 2 diabetes: principles of pathogenesis and therapy." *The Lancet* 365.9467 (2005): 1333-1346.
[4] Levin, Marvin, and Michael Pfeifer, eds. *The uncomplicated guide to diabetes complications*. American Diabetes Association, 2009.
[5] Abdullah A. Aljumah, Mohammed Gulam Ahamad, Mohammad Khubeb Siddiqui, Application of data mining: Diabetes health care in young and old patients, Journal of King Saud University - Computer and Information Sciences, Volume 25, Issue 2, July 2013, Pages 127-136
[6] Patil, Bankat M., Ramesh Chandra Joshi, and Durga Toshniwal. "Hybrid prediction model for Type-2 diabetic patients." Expert systems with applications 37.12 (2010): 8102-8108.
[7] Vasudevan, P. "ITERATIVE DICHOTOMISER-3 ALGORITHM IN DATA MINING APPLIED TO DIABETES DATABASE." Journal of Computer Science10.7 (2014): 1151.
[8] Sanakal, Ravi, and Smt T. Jayakumari. "Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine."International Journal of Computer Trends and Technology 11.2 (2014): 94-8.
[9] Aljumah, A. A., Siddiqui, M. K., & Ahamad, M. G. (2013). Application of classification based data mining technique in diabetes care. Journal of Applied Sciences, 13(3), 416-422
[10 ]Rahman, Rashedur M., and Farhana Afroz. "Comparison of various classification techniques using different data mining tools for diabetes diagnosis." Journal of Software Engineering and Applications 6.03 (2013): 85.
[11] Arun K. Sigurdardottir, Helga Jonsdottir, Rafn Benediktsson, Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis, Patient Education and Counseling, Volume 67, Issues 1–2, July 2007, Pages 21-31, ISSN 0738-3991

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12] Bagdi, Rupa, and Pramod Patil. "Diagnosis of Diabetes Using OLAP and Data Mining Integration." International Journal of Computer Science & Communication Networks 2.3 (2012).

[13] Xue-Hui Meng, Yi-Xiang Huang, Dong-Ping Rao, Qiu Zhang, Qing Liu, Comparison of three data mining models for predicting diabetes or prediabetes by risk factors, The Kaohsiung Journal of Medical Sciences, Volume 29, Issue 2, February 2013, Pages 93-99, ISSN 1607-551X

[14] Gregori, Dario, et al. "Using data mining techniques in monitoring diabetes care. The simpler the better?." Journal of medical systems 35.2 (2011): 277-281.

[15] Liu, Haifeng, et al. "An efficacy driven approach for medication recommendation in type 2 diabetes treatment using data mining techniques."Studies in health technology and informatics 192 (2012): 1071-1071.

[16] Butwall, M., & Kumar, S. (2015). A data mining approach for the diagnosis of diabetes mellitus using random forest classifier.*International Journal of Computer Applications, 120*(8)

[17] Karthikeyani, V., Begum, I. P., Tajudin, K., & Begam, I. S. (2012). Comparative of data mining classification algorithm (CDMCA) in diabetes disease prediction. International Journal of Computer Applications, 60(12)

[18] Habibi, S., Ahmadi, M., & Alizadeh, S. (2015). Type 2 diabetes mellitus screening and risk factors using decision tree: Results of data mining. *Global Journal of Health Science, 7*(5), 304-310

[19] Namayanja, J., & Janeja, V. P. (2012). An assessment of patient behavior over time-periods: A case study of managing type 2 diabetes through blood glucose readings and insulin doses. *Journal of Medical Systems*

[20 ]Strack, Beata, et al. "Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records." *BioMed research international* 2014 (2014).

[21] Yoosofan, Ahmad, et al. "Identifying Association Rules among Drugs in Prescription of a Single Drugstore Using Apriori Method." *Intelligent Information Management* 7.05 (2015):

[22] Poorinmohammad, Naghmeh, and Hassan Mohabatkar. "A Comparison of Different Machine Learning Algorithms for the Prediction of Anti-HIV-1 Peptides Based on Their Sequence-Related Properties." *International Journal of Peptide Research and Therapeutics* 21.1 (2015): 57-62.

[23] Bruno Fernandes Chimieski, Rubem Dutra Ribeiro Fagundes, Association and Classification Data Mining Algorithms Comparison over Medical Datasets, 2013

[24] Duarte de Araujo, Flavio Henrique, Andre Macedo Santana, and Pedro de Alcantara dos Santos Neto. "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process." *Latin America Transactions, IEEE (Revista IEEE America Latina)* 13.1 (2015): 299-306.

[25] Vieira JAP. Algorithm development for physiological signals analysis and cardiovascular disease diagnosis - a data mining approach [thesis]. Coimbra: University of Coimbra; 2011.